MUSIC WHILE YOU WORK:

THE EFFECT OF MUSIC ON TYPING PERFORMANCE AND EXPERIENCE

ANNA FELICITY BRAMWELL-DICKS
Doctor of Philosophy

University of York
Computer Science

September 2016

ABSTRACT

Over many years, and in a variety of contexts, researchers have shown that music affects human behaviour and emotion. In this thesis, I explore how music affects people when undertaking mundane work related computing tasks by performing a series of experiments investigating how various dimensions of music affect transcription typing performance and experience. Some results were inconclusive with significant interactions followed by non-significant post hoc analyses, but nevertheless interesting themes emerged.

Generally, music containing vocals compromised typing performance because it was more distracting than instrumental music. However, when played at a low volume performance was better with vocals in the music. This surprising result could be because vocals bring lower volume music to the attention of the typists so any effects caused by the rest of the music dominate, leading to overall performance improvements.

Typing performance without music was similar to performance when accompanied by ambient music, possibly because the ambient music could fade into the background. In contrast, classical and rock music were more intrusive and rhythmically challenging, and negatively affected both performance and experience.

Fast tempo 3/4 time music reduced typing accuracy when compared to almost all the other tempo and time signature manipulations. The frequency of the emphasized beats in this music was considerably higher than in the other conditions, which may explain why performance was particularly affected by this variation.

The findings from these experiments may influence experiment design in this field. This thesis shows music is a complex, multifaceted stimulus which should be considered as multidimensional experiential gestalt. The approach of reducing music to isolated dimensions for manipulation is inappropriate. Further, the work has shown that manipulating tempo alone may be insufficient as the combination of time signature and tempo affects frequency of the emphasized beats, which can be important.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Equations

List of Accompanying Material

The accompanying USB drive contains the music used in the experiments as follows:

Folder	Sub Folder	Files
Music Files	Experiment 1	AltRockWithoutVocals.wav AltRockWithVocals.wav PopRockWithoutVocals.wav PopRockWithVocals.wav
	Experiment 2	Ambient.wav Classical.wav Rock.wav
	Experiments 3 and 4	WithoutVocals.wav WithVocals.wav
	Experiment 5	3time_140bpm.wav 3time_180bpm.wav 3time_210bpm.wav 4time_140bpm.wav 4time_180bpm.wav 4time_1210bpm.wav

Music files on accompanying USB drive

For dearest Simon,

“If you're out on the road
Feeling lonely and so cold,
All you have to do is call my name
And I'll be there on the next train.

Where you lead I will follow
Anywhere that you tell me to.
If you need, if you need me to be with you
I will follow where you lead.”

Carole King & Louise Goffin
(The Gilmore Girls Theme Song)

I miss you, Simon.
Peace, love, Waga’s and heteroscedasticity, forever.

Acknowledgements

It is no exaggeration to say that I am only able to submit this thesis because my amazing friends, colleagues and family gave tremendous support me along the way. A few people deserve special mentions for putting up with me when I was particularly awful. And at times, I'm sorry to admit, I was more than a little bit awful.

First, I should thank my family who were forced to listen to me going on and on and on about music, typing and statistics for years. Thanks for being there. I am sorry I have been so dull and difficult. But, it is done so we can all breathe a big sigh of relief. Yippee!

Thanks also to the many participants who did those boring transcription typing tasks while exposed to different pieces of music. I know it was dull. But, I would not have a thesis to write without your willingness to take part, so thank you.

Most importantly, I am eternally indebted to Alistair Edwards and Helen Petrie for their supervision. There is no way I would be submitting this thesis without their support. I will forever be grateful for Helen's help in locating my mojo when it disappeared. From inspiring me to love statistics and experimental method, to making me take holiday, to finding me interim research work on some really, stonkingly cool projects when the money was running out and for sending me away to Milan, Atlanta and Australia for research, Helen really has gone far beyond expectations. I am so fortunate to have had you in my corner throughout this battle. I genuinely could not have done this without you and will never be able to fully extend my appreciation for your help and friendship.

Many others from the Department of Computer Science have been fantastic mentors, brilliant colleagues and even better friends. Especially, Rob Alexander, Katrina Attwood, Paul Cairns, Sam Devlin, Bláithín Gallagher, Marysia Koc, Louis Rose, Richard Paige, Fiona Polack, and Dave Swallow. All of whom always willing to be moaned at and ready to provide a balanced view at a moment's notice. It has not gone unnoticed, thank you.

I am also incredibly grateful to my colleagues in the Department of Theatre, Film and Television, who have had to put up with me during my final years of this process. I am extremely fortunate to have a job that I love and to work with some incredibly inspiring, interesting, entertaining and encouraging people, many of whom I now consider close friends. To the Interactive Media team, in particular, you are all super amazing, but especially Debbie Maxwell who has become a true and valued friend. Thank you all.

There are many other students from across the University who have helped me along the way, as friends, flatmates, office mates, critiques, arch-enemies, participants, and as inspirational students. Some have moved on, some are still here, and I mislaid some along the way. But to all of you, it has been a rollercoaster so thanks for making it entertaining, thanks for pushing me to be a better academic and, more importantly, a better human.

Thanks to Tom Calpin for propping me up through the final corrections hurdle. Thanks for bringing mountainous exploits into my life, for making me laugh and smile, and for distracting me during the hardest of times. I could not have got through the final beastly year without you taking care of me. So, from the bottom of my heart, thank you.

I love that the Bucks' crew are still an important part of my life over two decades after we met as awkward, dorky 12 year olds. The birthday parties, engagement parties, hen dos, weddings and baby showers have been some of my favourite times throughout this PhD. You gave me guilt free time away from work and do not care if I am awarded this degree or not, but will be proud of me if I do. I can't even begin to name you all… so to all the girls, the guys, the other halves, the hundreds of children, the pets and hangers on, you all completely rock. I am so excited to spend more time with you now freedom is mine. I promise to be a better friend now. Thank you all for being you and for putting up with me.

I am also very blessed to have a lot of wonderful close friends who take care of me. Emma Curry, Mike Dodds, David George, Chris Power, Tara Power, Simon Poulding (RIP), Jimmy Williams and Sarah Wiseman deserve extra special mentions for looking after me when I was very broken and for being my cheerleaders throughout. It has been stupidly tough, but you are all amazing and I am so incredibly lucky to have had you as a significant part of my life during this process. I am genuinely only still here right now because of you taking care of little old me. Words cannot really convey my gratitude here, but you know how it has been. All I can really say is thank you for not giving up on me. I love you all a tremendous amount. I hope you know how grateful I am for everything, and just how much you all mean to me.

And last but by no means least, the hugest of thanks must go to my gorgeous feline friends, Pearl and Ruby. They make everything better. Always.

Declaration

I, Anna Bramwell-Dicks, declare that this thesis and the work included within it is my own and has not previously been accepted for any degree other than the Doctor of Philosophy at the University of York.

Parts of this work have been published in the following books and at conferences:

Bramwell-Dicks, A., Petrie, H., Edwards, A. D. N., Power, C. (2013) "Affective Musical Interaction: Influencing Users' Behaviour and Experiences with Music" in Music and Human-Computer Interaction (2013) Eds. Holland, S., Wilkie, K., Mulholland, P., Seago, A. Springer Verlag, London

Bramwell-Dicks, A., (2015). "The Effect of Vocals and Music Volume on Transcription Typing Speed" in Proceedings of the Eighth York Doctoral Symposium on Computer Science, Eds. Paterson, C. Technical Report YCS-2012-480. Department of Computer Science, University of York. Awarded Best Academic Paper.

Bramwell-Dicks, A., Petrie, H., Edwards, A. D. N., (2016) "Can Listening to Music Make You Type Better? The Effect of Music Style, Vocals and Volume on Typing Performance" The 22nd International Conference on Auditory Display (ICAD-2016). Canberra, Australia. July 2nd to 8th 2016

Chapter 1

INTRODUCTION

> "Music itself is going to become like running water or electricity."
>
> (David Bowie, quoted in Pareles, 2002)

Music is ubiquitous. Instances of exposure to music within our day to day lives is now prevalent and pervasive. For many people, it would be rare to go a day without being exposed to some form of music. As they say, music is all around us.

Music is a service. As predicted by David Bowie in 2002, music has become a service people access rather than a product they own. The dramatic growth in popularity of online music streaming systems, such as Pandora and Spotify, has led to even more people consuming a wider variety of music than ever before. People can access as much music as they want for free using a service like Spotify if they are happy to listen to advertisements, or for just £10 a month without advertisements. Importantly, this music can also be downloaded to a device and accessed without an internet connection, so can be consumed from anywhere. From pop, classical, dance, jazz and many, many more genres of music, it has never been easier and cheaper to legally access and consume music from wherever you want.

Music affects people. There is a wealth of empirical evidence showing that exposure to music can change people's emotions and behaviour. However, despite the quantity of research in this area, and the popular belief that music can positively affect people, we only have a limited theoretical understanding of why music affects people. As a result, it is not really possible to systematically predict how music affects people, meaning it is hard to intentionally exploit instances of exposure to music to improve performance or experiences. Beyond the popular, yet subjective belief, that music does have a positive effect on people the theoretical basis for understanding how music affects people is inadequate to allow us to exploit music confidently to change people's behaviour. Many people seem to have strong opinions about how music affects them as individuals. But, are there any general theories for how music affects people? If so, what do they say? And are they suitable? And, if not, how can we approach developing well-evidenced theories that can be exploiting for positive effect?

Music is a commodity. In suggesting that music will become like running water and electricity, Bowie also implies music will become a product people consume and intentionally use for its affective properties. People use running water to drink and wash clothes, but how will they use music in their day to day lives? Indeed, are people already using music in their everyday lives with specific objectives that go beyond its entertainment value?

The main objective of the research presented in this thesis is to explore whether music can be used to improve performance in and experience of mundane, work related computing tasks. This introduction chapter will define some key concepts, including what *everyday* means with regard to exposure to music and the differences between *hearing* and *listening* and between *background* and *accompanying* music. Two studies employing the Experience Sampling Method (ESM)(Larson & Csikszentmihalyi, 2014), which is described in Section 1.1.2, are reviewed to help to comprehend the scale of exposure to music in people's day to day lives as well as to explore the reasons why people are so frequently exposed to music. This is followed by a discussion of a survey which investigated how often and why people listen to music while working in office environments. Having established that people frequently listen to music, both in leisure and while working, and also that people often report a belief that music positively affects them, this chapter moves onto considering whether people outside of academia are interested in understanding how music affects them at work by considering the content of, and the response to, a couple of recent articles from the popular media. Finally, some preliminary questions for the research presented in this thesis are defined. A description of the structure of the rest of the chapters in the thesis is then provided, which includes a brief explanation of the rationale for the experiments and summaries of the key results.

1.1 The Role of Music in Everyday Life

People frequently listen to music as an accompaniment to a wide variety of leisure and work related activities. For example, activities such as driving, shopping, exercising, writing, computer programming, analysing data, reading and many, many others, are all often undertaken while exposed to music. Of course, music is often an integral part of the activity itself, rather than just something that acts as an accompaniment to an activity. For example, when playing a computer game, the music and other associated audio components, such as sound effects, are integral to the gaming experience. But, the research presented in this thesis focuses on the use of music as an accompaniment to some other activity as determined by an individual's active choice to play music while doing something else.

The reasons why people listen to music as an accompaniment to other activities are as complex and varied as the contexts, situations, and activities within which the listening activity takes place. However, one reason why people choose to listen to music whilst doing something else is the perception that their enjoyment of the activity is enhanced with the presence of music. Aside from enjoyment what, if any, effects do music have on an individual's activity performance, behaviour, emotions and experience? And, is it possible to develop theories for why music affects people in different ways, leading eventually to an understanding of how we can exploit this affective property of music more systematically? My research aims to address these questions with a focus on a work context, particularly involving mundane tasks, to see if music can be used to positively affect performance at work when using a computer.

The idea that music affects people is not a new one. There are even examples of music affecting people described in historic literary sources, including the Bible:

> "And whenever the harmful spirit from God was upon Saul, David took the lyre and played it with his hand. So Saul was refreshed and was well, and the harmful spirit departed from him."
>
> (1 Samuel 16:23, English Standard Version)

The question of how music affects people's behaviour and experiences has been of interest to researchers from a wide variety of domains for many years. A review of empirical studies in this area is included in Chapter 2. Although researchers have long been interested in the effect of music on people's behaviour and emotions in a variety of contexts, it is only since the early 21st century that there has been a particular focus on the role of music in everyday life. To understand the landscape in terms of how and why people are exposed to music in their everyday lives researchers have performed a number of studies using the ESM.

1.1.1 KEY CONCEPTS AND TERMINOLOGY

Before reviewing two of these ESM studies exploring how music features in everyday life, some key concepts and terminology need to be defined. The first concept that needs to be explored is what it means for an instance of exposure to music to be characterised as an everyday experience.

1.1.1.1 Everyday Exposure to Music

Early research looking at the role of music in everyday life did not explicitly define the term *everyday* with regard to exposure to music. Instead, this research seemed to adopt the dictionary definition of everyday as an adjective, meaning "to be met with every day; common, ordinary" (Everyday, 2016). This implicit definition seems reasonable and suggests that to categorise an instance of exposure to music as an everyday experience it does not literally need to be an experience that occurs every day, but rather one that is commonplace and mundane. However, with no agreed definition of what it means for a musical experience to be classified as an everyday experience there are some discrepancies and confusion in the literature. For example, in the introduction to the book *Music and Mind in Everyday Life* (Clarke, Dibben, & Pitts, 2010) there is an in depth discussion of the role of music at Lady Diana Princess of Wales' funeral. Although the discussion of the use of music in the book's introduction is interesting and potentially relevant and provides some interesting insights into the use and role of music within our lives, it does not seem reasonable to describe the funeral of an infamous Royal as an *everyday* experience.

Sloboda presented a ten item proposition to explicitly define what it means for research to be focused on everyday musical experiences (Sloboda, 2010). The proposition contains four items about the process of performing research involving everyday exposure to music and six items focused on the categorisation of instances of exposure to music as everyday experiences. The latter items are of most interest to my research as they can be used to gain a clearer understanding of what it means for an instance of exposure to music to constitute an everyday musical experience.

Sloboda identified the following six aspects as being important in defining an instance of exposure to music as an everyday experience:

- **"Frequency of occurrence":** an everyday music experience is one that could, potentially, occur every day in the current time period and is appropriate to the people that are exposed to the music. For example, a piano teacher will have frequent exposure to live piano music but for many others hearing live piano music would be an occasional rather than common experience. However, in the 1900s exposure to live piano music is likely to have been more frequent as pianos were often in people's homes and meeting places. That said, for people in non-Western societies, where the piano was not a typical instrument in people's homes, exposure to live piano music would not have had a high frequency of occurrence, so could not be considered an everyday experience. These examples highlight the importance of understanding the context of the experience and the people involved before classifying instances of exposure to music as everyday musical experiences.
- **"Ordinariness versus specialness of the context or experience":** an everyday music experience is one that is neither surprising nor unexpected. For exposure to music be considered an everyday experience it should be an ordinary, typical, mundane part of the experience and not something special. The context within which the person is exposed to music should also be ordinary, e.g. exposure to music should occur while going about their everyday lives, rather than as part of a special event such as attending a wedding.
- **"Location of occurrence":** an everyday music experience will happen in an everyday, typical location. For example, hearing music at a supermarket is likely to be an everyday experience, but hearing music at a concert venue is, for most people, an uncommon, infrequent, special experience and therefore, not everyday.
- **"Circumstance of exposure: the role of choice":** although not true of all everyday music experiences, many everyday music experiences are likely to occur in situations where the person did not choose to be exposed to music. Instead, the music was a feature of the environment the person happened to be in at the time, for example, exposure to music when visiting a pub with friends.
- **"Nature of transmission":** nowadays most everyday music experiences will involve pre-recorded music played over a sound system rather than live music. There are however some exceptions, for example if someone was to hear a busker playing music on the street.
- **"Centrality of music to the experience, and the salience of the context":** many instances of exposure to music that are considered everyday occur as an

> accompaniment to some other activity, rather than having music as the focus of an activity. For example, exposure to music while exercising at the gym is a more typical everyday experience than going to a concert to hear your favourite musician.
>
> (Sloboda, 2010)

Sloboda's proposition is useful for defining some abstract characteristics of everyday music experiences. But, it does not help us to understand how frequently people are exposed to music or the contexts within which these listening activities take place. The proposition implies there are different types of everyday music experience but does not go as far as to define these types. The majority of the research investigating exposure to music in everyday life has focused on experiences involving hearing music. But, there are other types of music experience which can be classified as everyday. Clarke et al. (2010) proposed three different types of everyday music experience:

1. **"making music":** playing, improvising and composing music,
2. **"using music":** listening to or hearing music,
3. **"acquiring music":** learning about music, including the process of learning how to play instruments and getting exposure to music.

(Clarke et al., 2010)

Using Clarke et al.'s definitions, the research presented in this thesis aims to *use* music rather than make music or acquire musical skills and knowledge.

1.1.1.2 Hearing and Listening

The definitions of the three types of everyday music experience highlight an important distinction between listening to and hearing music. To *hear* is a verb defined as "to perceive, or have the sensation of, sound; to possess or exercise the faculty of audition, of which the specific organ is the ear" (Hear, 2016) whereas the verb *listen* is defined as "to hear attentively; to give ear to; to pay attention to (a person speaking or what is said)" (Listen, 2016). These definitions indicate that people can both hear or listen to music depending on whether their attention is focused on the sounds or not.

The distinction between hearing and listening to music often depends on whether the person was involved in the choice to be exposed to the music and if they were involved in a simultaneous activity or just focused on the music. For example, if I was to stream some music at home to help me relax after a long day at work, I am choosing to listen to the music because I perceive there to be experiential benefits. However, if music is played in a supermarket while I am doing my weekly shop then this is likely to involve me passively hearing rather than actively listening to the music. These examples present a vast oversimplification of auditory perception as there is often frequent fluctuation between hearing and listening. For example, while at home listening to music I may choose to start cleaning which takes the focus of my attention meaning I am hearing rather than listening to the music. Or, I may choose to actively listen (and maybe even sing along) to music while in the supermarket. The distinction between hearing and listening as passive versus active attention activities provides a useful frame of reference for discussion and is one that has been used elsewhere (e.g. in Clarke et al., 2010). But, the fluidity between concepts should be noted and carefully considered.

In this thesis, when the distinction between listening and hearing is considered important and is clear the term *active* is used as a qualifier for listen, while *passive* is used to qualify hearing. When it is unclear if the people are hearing or listening, the terms listening, and listener, are used by default.

1.1.1.3 Background or Accompanying Music

In much of the literature investigating the effect of music on people, the term *background music* is used to describe music that used to accompany some other activity. The use of the word background seems to imply that this is music that is heard, rather than listened to, as for music to be in the background of another activity suggests that the attention is focused on the other activity. However, there does not seem to be any evidence from these studies that the music is indeed in the background of the person's attention and thus it may well move to the foreground of their attention at times. So, in this thesis I prefer to use the term *accompanying music* rather than *background music* so that the music is not portrayed as being behind the other activity, which inadvertently suggests that the music is being heard rather than listened to, and perhaps implies that the music is a lesser part of the experience. It is entirely possible that, in many of these situations, the music is part of the foreground of the individual's attention and therefore, the term *background music* seems to be a misnomer and should be avoided unless there is clear evidence that the music should be placed behind the other activity.

1.1.2 Experience Sampling Method Studies

Having defined some key terms, we can now explore how music is used in everyday life and whether people are hearing or listening to music that accompanies their day to day experiences. To understand the role of music in everyday life and to demonstrate that music is a prevalent feature of everyday life, a couple of studies employing the Experience Sampling Method (ESM) (Larson & Csikszentmihalyi, 2014) are reviewed in this section.

The ESM is a popular approach to gather data regarding real world experiences as and when they occur in, almost, real time. The ESM involves asking participants to stop at various points in the day and make notes, or to complete questionnaires, about their current experience as it is happening. Sloboda's (2010) proposition was grounded in a review of a number of studies using the ESM to explore when, where and why people are exposed to music. The precise objectives of each study using ESM vary, but all report the frequency of exposure to music over a period of time. Many of the ESM studies do not consider the different categories of musical experience defined by Clarke et al., (2010) and some seem to assume or imply that the instances of exposure to music they are interested in only involve a person hearing or listening to music rather than considering using or making music as an everyday activity.

One study used the ESM to investigate the emotions associated with everyday exposure to music (Sloboda & O'Neill, 2001). Eight adult non-musicians were paged seven times a day over a week and each time they were asked to complete a questionnaire about what they were doing, how they were feeling and whether they were currently exposed to music. The participants reported exposure to music in response to 44% of the pages showing music is a pervasive part of their everyday lives. Music was most frequently listened to while doing housework (in 22% of episodes of exposure to music), while driving, running or cycling (in 22% of episodes of exposure to music), while doing desk work (in 14% of episodes of exposure to music) and while in bed (in 14% of episodes of exposure to music). These statistics tell us that, for these eight participants at least, exposure to music while working was a frequent part of their day to day lives.

The music was an accompaniment to some other activity in 95% of the instances where the participants were exposed to music, with the music as the focus of the activity in just 5% of episodes. This result shows that focused attention on music is atypical of most everyday instances of exposure to music. Instead, it is far more common for exposure to music to involve listening to music as an accompaniment to some other activity. The objective of this particular study was to explore the emotional impact of music. In this regard, the authors report that exposure to music resulted in participants that were “more positive, more alert, and more focused in the present” (Sloboda & O'Neill, 2001).

Another study used the ESM to try to understand exposure to music in everyday life by asking “who, what, when, where and why?” (North, 2004, p. 46). This study had a particular focus on how time of day affects exposure to music and the function that music serves within people’s everyday lives. Almost 350 people took part in this large study which took place over two weeks. Most of the participants (44.8%) were university students but there were also participants described as having managerial/administrative (22.3%) and professional (15.1%) occupations.

In this study, the participants were texted once a day over 2 weeks and asked to complete a questionnaire (i.e. 14 questionnaires each during the study). The time of day that the text messages were sent varied throughout the day. Participants reported they were currently exposed to music in response to 38.6% of text messages, with a further 29.5% reporting that although they were not currently exposed to music they had been exposed to music since receiving the last text message. Again, these results clearly show the extensive presence of music in people’s daily lives.

The majority of instances of exposure to music occurred when the participant was not alone (63.4%), with 5.8% occurring in the presence of work colleagues. In terms of time of day, most instances of music exposure occurred in the evenings, with participants more likely to be exposed to music at the weekend suggesting a prevalence of instances of exposure to music within leisure hours rather than at work. Unfortunately, this study did not ask the participants to record whether they were working or at leisure during each instance and when asked to record the place of exposure the location options did not include items such as “at the office” or “in the library” so the context of these exposure instances is not well understood.

The majority of instances of exposure to music occurred at home (50.1%) with 11.6% of these instances described as “at home doing an intellectually demanding task” (North, 2004, p. 69). The main reason reported for why they chose to listen to music while performing an intellectually demanding task was that it helped the participant to concentrate or think (55.1%), demonstrating that people perceive music to have an ability to help them focus on difficult tasks, even when those tasks are being completed at home.

When participants had chosen to hear music their main reasons for doing so were that they enjoyed the music (56.4%), it helped to pass the time (40.6%) or it was habit (30.6%). In addition, in 17.2% of episodes the participants described their motivation for listening to music as helping them to concentrate or think, i.e. in almost 1/5th of the instances where people were choosing to listen to music in their everyday lives their motivations for doing so included the belief that the music helped them to concentrate or think.

These ESM studies show that music is a widespread part of our everyday lives and that the contexts and situations where people are exposed to music is also vast and varied. It follows that the reasons why people are exposed to music in their daily lives, either by choice or involuntarily are also vast and varied, but there is evidence to suggest a belief of many people that music helps them to concentrate or think. It is also important to note that listening to music was rarely the primary focus instead music was used to accompany some other activity, e.g. listening to music when driving, or hearing music in a shop.

1.1.3 Listening to Music in Offices

Although most instances of exposure to music identified by the two ESM studies described in the previous section occurred at home, there is a suggestion that people are also exposed to music while working or performing intellectually demanding tasks. To my knowledge there have not been any ESM studies looking at the frequency and motivations for why people are exposed to music in a work related context. However, an exploratory survey investigated the motivations behind why people listen to music in office workplaces (Haake, 2010). In this survey, almost 300 people who worked in offices in the UK were asked about their music listening habits while at work. On average, the respondents reported that they spent 36% of their working week listening to music. Reasons for listening to music included improving mood and helping them to relax as well as making them less bored and improving their focus by blocking out surrounding noise and helping creative flow. The most prominent function of music accounted for 30% of the variance was categorised as "inspiration" and included the following five statements, the music:

- "inspires/stimulates you"
- "helps your creative flow"

- “provides a different perspective”
- “helps you pace your work”
- “creates a suitable atmosphere”

(Haake, 2010)

The results from this study indicated that listening to music at work can be functional as well as enjoyable. People choose to listen to music while working in the belief that it will positively affect their work performance, because it helps pacing and creates a suitable atmosphere as well as blocking out surrounding noise.

1.2 Research Motivation

Given that people are frequently exposed to music in their everyday lives, and that many choose to listen to music because they enjoy it and they believe it helps them to think, it is appropriate to consider music as a resource or commodity that has the potential to be exploited. In fact, it seems that many people are already using music in the belief that it has a positive effect on them by helping them concentrate and creating a suitable atmosphere for productive working. This perception seems to be widespread and pervasive and explains why exposure to music in everyday life is frequent, and why many people choose to listen to music as accompaniment to their daily activities, including those involving work.

People believe music positively affects them. But, is there empirical evidence that underpins this belief that music affects behaviour and experience? Or is this general perception based simply on personal preference and an individual’s previous experiences? If there is evidence to back up the perception that music positively affects behaviour, are there any underpinning theories that explain why and how music affects people? The literature review in Chapter 2 includes clear empirical evidence that music does affect people’s behaviour and their experiences from a range of contexts, but also that theories explaining this affectivity of music are currently somewhat lacking.

Empirical research into the effect of music on listener behaviour and experience has been performed in numerous contexts over many years, much of which has been picked up by the popular press. There are frequent stories relating to how music affects people, e.g. a recent article from the Guardian about Marks and Spencer’s decision to stop playing music in store (Saner, 2016), but the empirical research that the quoted academics refer to in these articles is often somewhat dated. Another recent article brings the attention to listening to music in the workplace with the attention-grabbing headline "This is the kind of music you should listen to at work" (Davidson, 2016). I attempted to get access to the study this newspaper article reports, including contacting the journalist directly (see Figure 1.1) and emailing the company involved, but was unfortunately unable to access the research so can only report the results described in the newspaper article rather than the original research.

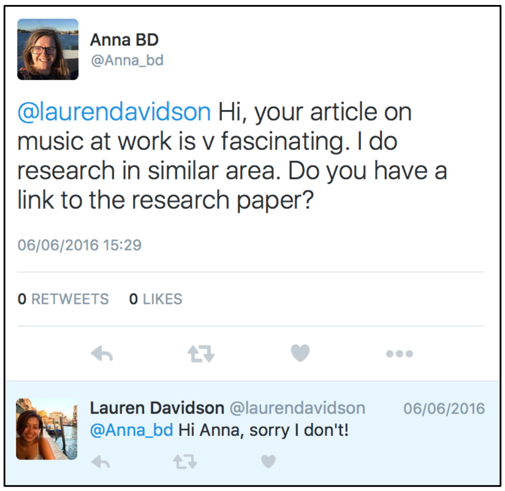


Figure 1.1: Twitter interaction with journalist, Lauren Davidson

The article states that this research study found:

- classical music was best when work involved numbers or attention to detail,
- pop music was best if the work involved data entry or working to deadlines,
- ambient music was best if the work involved solving equations, and
- dance music was best if the work involved problem solving or proof-reading.

As the research itself cannot be viewed and critiqued the strength of these claims cannot be verified. In particular, it is not possible to judge characteristics of the music that were used as stimuli or to provide a critique of the validity of the method. Although the research itself cannot be evaluated, this article provides evidence that, in the popular press at least, there is an interest in understanding how music can be exploited to improve performance at work. The result is attention grabbing headlines and generic advice based on potentially questionable research. This particular article was *liked* by 653 people on Facebook and shared by 409 (as of August 2016) indicating strong public interest in how best to exploit music to positively affect performance and experience when working.

Even more recently, an article entitled "Does music really help you concentrate?" (Burnett, 2016) was published in the Guardian. In just 12 hours it was shared 355 times directly from the Guardian and received 38 comments including people disagreeing over whether music helps or hinders their work, often including some strongly felt opinions such as those from Fortescue99 in Figure 1.2.

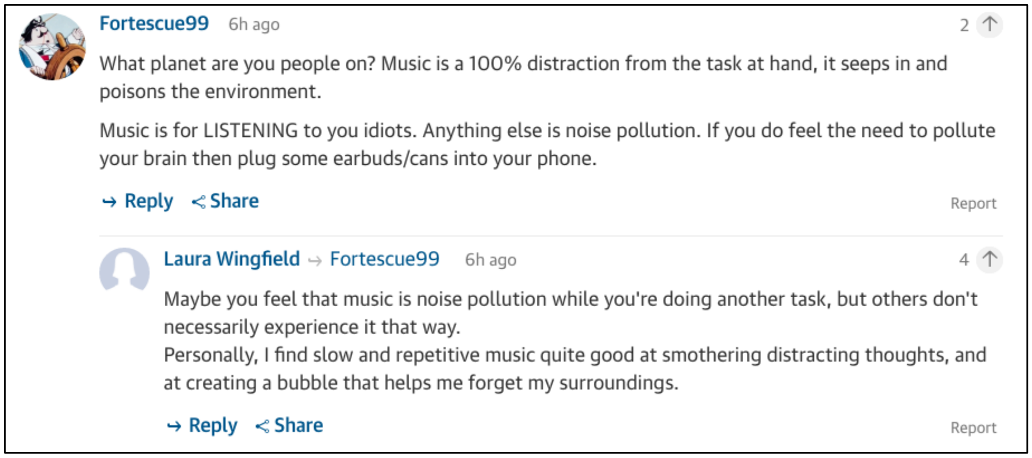


Figure 1.2: Comments from Guardian article

This newspaper article did not present unpublished academic work as fait accompli but explores various opinions about whether music can help people concentrate based on published research. The argument presented in the article is that video game music may be best for aiding concentration as this is something video game music designers focus on. The article encourages researchers to think about these questions further and, given the strength of opinions presented in the comments, this seems like an avenue of research that could have impact.

1.2.1 Aims and Objectives

The programme of research presented in this thesis aims to better understand how people are affected by music as an accompaniment to mundane, monotonous work related computing tasks. By performing empirical studies in the area, it is hoped this work will lead to a better understanding of how music affects people when working and contribute to the development of theories explaining why music can change people's behaviour.

Therefore, the main question asked in this thesis is:

- How does music affect people's performance and experience in mundane work related computing tasks?

The literature review in Chapter 2 shows much of the empirical research investigating how music affects people focuses on manipulating the tempo and loudness of the music. This thesis will try to expand the focus of research involving affecting people with music by investigating the impact of other dimensions of music on performance and experience. This leads to the additional question:

- How do dimensions of music other than tempo and loudness affect people?

These two questions formed the starting point for this research. As the experiments were completed and the data analysed a number of other potential avenues worth exploring emerged, which are highlighted in the Discussion sections of each experiment chapter.

1.3 Thesis Structure

Five quantitative experiments investigating the effect of different dimensions of music on transcription typing performance and experience are reported in this thesis. Transcription typing was chosen as this is a mundane, monotonous task involving a computer and a previous experiment had investigated the effect of music on transcription typing p

erformance (this experiment is reviewed in depth in Chapter 2, Section 2.1.5).

There are nine further chapters in this thesis, the objectives and results from which are briefly described below:

- **Chapter 2: Literature Review**
  The literature review presents a discussion of empirical research investigating the effect of music on people's behaviour and experiences from a variety of contexts. The theories explaining why music affects behaviour and experience are not well developed, so the review aims to highlight key experiments that have been performed, including discussing the methods used and the dimensions of music that were manipulated. The chapter also reviews the motivations for why researchers and industry use audio within computing interfaces. The focus of this section is on the different types of sounds used, the motivations for incorporating these sounds in interfaces and how sounds affect the users.

- **Chapter 3: Introduction to the Experiments**

The five experiments in this thesis all explore how music affects transcription typing performance and experience. This chapter provides a description of and justification for the choice of task, the methods and the analytical approaches used in the experiments, including detailed descriptions of both the stimuli and the transcription typing tasks.

- **Chapter 4: Experiment 1 - Vocals and Style**
  This first experiment explored how the presence of vocals in music affects transcription typing performance and experience across two pieces of rock music with very different characteristics. The results of the experiment suggested that vocals may reduce performance and make the tasks harder because the music was more distracting. But, there was insufficient evidence from the experiment to clearly state this with authority, possibly due to a lack of statistical power or encountering ceiling effects. One other important outcome from this experiment was that transcription typing as a task had more nuance and subtlety than expected. The data clearly showed two groups of participants, which were categorised as fast and slow typists, and the effects of music differed between the groups with the fast typists' performance reduced with vocals in the music but no evidence that the slow typists were similarly affected.

- **Chapter 5: Experiment 2 - Genre and Task Difficulty**
  The second experiment investigated how typing performance and experience were affected by instrumental ambient, classical and rock music, including a comparison to a without music condition. The effect of these four music conditions was explored with three different transcription typing tasks which were intended to have varying levels of difficulty. Two of the tasks were in English, one of these used text from a children's story and the other used a novel intended for an older audience (which were described as simple and advanced English tasks, respectively). The final task used text in Dutch, which was assumed to make the task harder for non-Dutch speaking participants. The results from this experiment indicated that task difficulty was an important manipulation with the effects due to music differing across the difficulty levels. More importantly though, the results suggest that typing with an accompaniment of ambient music is similar to typing without music playing in terms of the effect of music on performance and experience. That is, there was no evidence these two music conditions had different effects on performance or experience. In contrast, both rock and classical music reduced performance in the English tasks and made them harder. Again, the classification as a slow or fast typist was important in this experiment with the fast typists experiencing greater affects than the slow typists. The task difficulty manipulation was somewhat successful, as the Dutch tasks were harder than the

English ones. But, there were no substantial differences in difficulty ratings between the two English task conditions.

- **Chapter 6: Experiment 3 - Vocals, Volume and Task Difficulty 1**
  The experiment presented in this chapter returns to investigating the effect of vocals on transcription typing performance, this time using a classroom based methodology (rather than laboratory based) in order to increase participant numbers whilst still maintaining tight control of the experiment. The volume of music was manipulated to investigate how this dimension affects performance and experience. The objective for including this dimension was that if volume has no obvious effect then a large scale online experiment could be performed fairly easily. But, if volume does affect performance and experience careful consideration of how to mitigate for a lack of control of playback volume in an online experiment would first need to be carefully explored. Also, a task difficulty manipulation was included in the experiment because it was shown to be an interesting factor in Experiment 2, particularly with regard to the comparison of the effect of music between the advanced English and Dutch tasks. The results from this experiment suggest that the volume of the music may affect whether or not the presence of vocals in the music reduces performance in the English tasks. But, these results were again somewhat inconclusive. Although the classroom based methodology used in this experiment increased the sample size, the numbers of participants that could be classified as fast typists was relatively small, which is likely to have affected the overall power of the experiment. Given that the previous two experiments indicated that the effects of music were greatest for the fast typists, more participants were needed to better explore these hypotheses.

- **Chapter 7: Experiment 4 - Vocals, Volume and Task Difficulty 2**
  The classroom based methodology used in Experiment 3 successfully increased the overall numbers of participants, but the experiment still lacked power. This was, in part, due to only being able to classify a small percentage of the participants as fast typists, for whom the effects of music seem to be strongest. To overcome this limitation of Experiment 3, a similar experiment was performed over two years using the same classroom based methodology, which doubled the number of participants. In this experiment, the hard difficulty task used the advanced English text with randomly inserted pseudowords[1] instead of Dutch language. This change to the task difficulty manipulation led to a task that was more difficult, but still used the English language constructs and so was still readable, even though it had no meaning. The results from this experiment showed that the performance of the

1 A pseudoword is a non-word that has phonological and orthographic structures resembling a real word in that language but has no meaning in the language.

fast typists who were exposed to high volume music was reduced when the music contained vocals. However, the typists who were accompanied by low volume music actually typed faster with vocals in the music and slower when the music was instrumental. This result demonstrates the importance of the volume of the music in this type of experiment, preventing a move to an online methodology to get data from a larger sample of participants without further exploration of how to control volume or mitigate for its effects.

- **Chapter 8: Experiment 5 - Tempo and Time Signature**
  Experiments 3 and 4 showed that the loudness of the music was important in determining the scale and type of effects caused by the music, so an online approach was not really suitable for further experiments at this stage of the research. So, the experiment reported in chapter 8 returns to the laboratory based methodology and focuses on trying to understand why the rock and classical music used in the Experiment 2 had similar effects despite very different musicological characteristics. The focus of this experiment was on the effect of tempo and time signature as these are two of the many ways the rock and classical music differed. The experiment explores the effect of music at three different tempos using two pieces of similar classical music, one of which was in 3/4 time and the other had a 4/4 time signature. In essence, this experiment was exploring how the frequency of the emphasized beats affects performance and experience when transcription typing. The results from this experiment showed that in terms of typing speed, the tempo and time signature manipulations did not have a significant effect. But, in terms of accuracy, performance was reduced in the 3/4 time signature condition when the music was at a fast tempo in comparison with the slow and mid tempo 3/4 pieces and the 4/4 fast tempo music.

- **Chapter 9: Preliminary discussion**
  Chapter 9 includes a preliminary discussion of the findings from the experiments presented in the thesis. The original aims of this body of work are considered in terms of whether or not they have been achieved with a discussion of potential reasons why, as well as the limitations of the work.

- **Chapter 10: Synthesis and Conclusions**
  The final chapter in the thesis aims to synthesise the results across the experiments and draw appropriate conclusions. The chapter begins by distilling the results from across all the experiments into tabular form in order to consider how the key hypotheses have been answered through the work. Then, five different theoretical, conceptual frameworks are discussed and applied to the results to explore whether each could be used to explain the results from the experiments. The chapter also

considers and highlights the key contributions from the work in this thesis. In many ways, and as with many PhDs, this programme of research has raised more questions than it has answered. So, these questions are formalised together with possible approaches for answering them in the future

Chapter 2

LITERATURE REVIEW

The primary objective of my research is to exploit music to positively affect people's behaviour when working at a computer. To that end, this literature review begins by discussing research investigating the effect of music on people's behaviour and goes on to consider how sound is currently used in technological interfaces.

2.1 HOW MUSIC AFFECTS PEOPLE

A statistical meta-analysis of the effect of music on adult listeners analysed the results from 97 studies[2] published before 2008 which included the term *background music* (Kämpfe, Sedlmeier, & Renkewitz, 2011). Of these 97 studies, 66 included a comparison of the effects due to some dimension of background music with a condition where music was not playing. There was no overall effect of background music when these studies were considered as a single set. However, Kämpfe et al. (2011) noted the effect of music seemed to be context dependent meaning the overall null effect did not adequately capture how music actually affects people. In some situations, music had clear positive effects but in others the effect of music was negative. As a result, when the meta-analysis was performed as a review of all the papers together the significant negative effects identified in some situations cancelled out the significant positive effects seen in other contexts.

2 189 articles were originally found but 92 of these could not be included in the meta-analysis as the experiments had not been reported in sufficient detail to calculate the effect size.

To explore whether the effect of music depends on context the meta-analysis was performed again, this time focusing on four different types of dependent variable (DV), as follows[3]:

- **"mundane behaviour":** for example, the speed of eating and number of steering wheel movements when driving,
- **"cognition - judgement":** for example, perception of ability to sustain concentration,
- **"cognition - achievement":** for example, numbers of correct responses,
- **"emotion":** for example, levels of nervousness or excitement.

(Kämpfe et al. 2011)

This analysis identified that music systematically affects both emotions and behaviour but could not identify any evidence of effects specifically relating to the two cognition types. Their explanation of this result was that the classification "did not lead to homogeneous subgroups of studies" (Kämpfe et al., 2011, p. 434) meaning the four category model of DV types did not accurately account for the studies included in the meta-analysis. As a result, instead of focusing on the DV, the type of task (e.g. sports, memory tasks) was considered to be a more appropriate as a moderator variable in the meta-analysis.

Using this task based classification of studies, the music had a small positive effect on sporting performance from 11 studies. The effect of music on remembering the content of advertisements presented on radio or television was small and slightly negative. However, another experiment investigated the use of multimedia mobile phone messages to deliver advertisements and found that music in these types of messages positively affected people's ability to recall the content of the advertisements, so it was suggested this particular experiment was from a different population and should be considered a separate task type. Other types of memory task which did not involve remembering the content of advertisements were grouped and resulted in a slightly negative effect due to background music on recall ability. Finally, the impact of background music on reading performance was investigated in eight studies where an overall negative effect was identified.

This task focused meta-analysis identified that the effect of music on sporting performance, as a particular type of behaviour, is positive. But, for tasks involving

3 The examples of DVs included here were taken directly from Kämpfe et al. (2011)

cognitive processes involving memory there was a persistent, slightly negative effect on performance due to the music. Separating the studies according to task rather than DV was a more suitable approach for classification for the meta-analysis but an insufficient number of studies for some task types meant the meta-analysis could only be performed for four categories – sports performance, memory of advertisement content, memory in other contexts and reading performance. These categories did not account for all of the papers included in the meta-analysis, so some of the research papers were not able to be analysed within this task focused analysis.

The meta-analysis continued by looking at the effect of different kinds of music when compared to each other, and not just by considering how the effects of music compare to a without music condition. This analysis included 71 studies where the global effect size was close to zero. Again, the four categories of DV (mundane behaviour, cognition – judgement, cognition – achievement and emotion) did not provide an appropriate classification. Instead, the analysis continued with a focus on the type of independent variable (IV) used in the experiments which were described as the "peculiarities of music" (Kämpfe et al., 2011, p. 348) including loudness, preference, genre, tempo and familiarity. In this thesis, I use the term *dimensions* rather than peculiarities to refer to specific attributes of the music stimuli.

The meta-analysis only explored effects due to the loudness and tempo of music as there were not enough studies focused on the other dimensions. In terms of loudness, nine studies were included in the meta-analysis with an overall negligible effect due to the loudness. This null result was explained due to different approaches and thresholds for manipulating loudness (e.g. one experiment's definition of loud could be vastly different to another's) as well as the wide variety of experiment contexts. For tempo, 16 studies were analysed, again with an overall null effect. In this case, the analysis continued focusing on tasks relating to motor behaviour by looking at eight studies which investigated how tempo affected the speed of behaviour. In this analysis there was a strong, positive effect of tempo on speed of behaviour, leading to the conclusion that "faster tempo consistently covaried with faster behavior" (Kämpfe et al., 2011, p. 440).

One of the most interesting aspects of this meta-review is the classification according to music peculiarities into five areas of interest, loudness, tempo, genre, preference and familiarity. Loudness and tempo are easy to define and control and seem to be the most frequently manipulated IV when investigating the effect of music on people, hence why the meta-analysis could be performed for these two dimensions. Genre can be hard to define specifically, but is quite easy to control in an experiment as the researcher either plays music from genre A or from genre B. Though, if genre is found to have an effect, it can be unclear why music from genre A has different effects to music from genre B. Identifying what it is about the differences between the genres that are the facilitator of particular effects is particularly difficult but is clearly important. Preference and familiarity are dimensions that involve the listener's relationship with the music, rather than focusing on the music itself. Preference is hard to control but can be investigated as a covariate or quasi-independent variable. Attempts can also be made to control familiarity, for example by creating new compositions and comparing the effects to those caused by famous music. But, it is hard to have tight experimental control over familiarity as some people may be more familiar with a piece of music than others, and these different levels of familiarity may be important. Although these five dimensions are the most frequently investigated, as demonstrated by their inclusion in the meta-analysis, there are other dimensions of music that have the potential to affect people, some of which are explored in the experiments reported in this thesis.

The meta-analysis demonstrated that when it comes to understanding the impact of music on people the situation is nuanced, multifaceted and depends on the context. It also seems that empirical research in this complex area is at an emergent stage. Kämpfe et al. (2011) conclude their paper by requesting that researchers improve their methods and experiment reporting, so a higher proportion of papers can be included in future meta-analyses to advance theoretical understanding of the impact of music on people. I also wonder whether their initial search based on the term *background* music and excluding research involving children may have artificially restricted the pool of papers available for inclusion in the meta-analysis. Either way, it seems clear that music does affect people. But, theories to explain how and why music affects people did not emerge from the meta-analysis.

A review of empirical research using null hypothesis significance testing (NHST) to investigate the effect of music on people is presented in the remainder of this chapter. Some of the papers discussed in the following sections were included the meta-analysis (Kämpfe et al., 2011), but others were not included either due to not using the specific term background music or because they were published after 2008. As sporting performance, reading tasks and recall of advertisements were considered through the discussion of the meta-analysis, I have not reported further exploration of these contexts in this review.

The aim of this review is to explore why and how different research fields investigate the use of music for emotional and behavioural affect, including careful consideration of the methods used. The review aims to highlight the range of contexts where the effect of music is considered to be a potentially affective medium, beyond those mentioned in the previous section. This literature review is not intended to be a systematic and comprehensive appraisal of all of the literature using NHST to investigate the effect of music as that would be beyond the scope of the thesis. Instead, a selection of studies have been included that highlight the range and variety of topics of interest, and different methods adopted. The review describes the independent and dependent variables as well as the methods used as a means of understanding how the research has been performed. As my main objective is to exploit music within work related contexts and the meta-review reported previously suggests that task is a moderator variable, this review is structured according to different task areas, beginning with leisure based domains and moving onto areas relating to mundane behaviour or work related contexts.

2.1.1 SHOPPING

One of the objectives of marketing research was to identify different tactics to increase a company's profitability. In a shopping context, this includes discovering which *atmospherics* encourage people to spend more money (Kotler, 1973). Music is an important atmospheric in retail environments which has been found to increase sales. There is a wealth of empirical research investigating the effect of exposure to music on people's behaviour and their experiences when shopping, some of which is reviewed here.

A key paper, which is often cited, investigated the effect of music tempo on shopping behaviour in a supermarket (Milliman, 1982). The study was performed over a nine week period in a supermarket. The behaviour of genuine customers was compared under three different music conditions, slow tempo music, fast tempo music or without music playing. The music condition was changed daily. The definitions of slow and fast tempo were established by asking a number of people to describe various pieces of music as either slow or fast. In this study, music with tempo of 72 beats per minute (bpm) or less was defined as slow, with fast tempo music specified as having a bpm of 94 or higher[4]. Customer behaviour was measured by looking at how much money was spent, how long it took the customers to walk between two points in the supermarket, and whether the customers could remember anything about the music that had been playing while they were shopping in the supermarket.

The tempo of music affected sales volume with higher sales in the slow tempo condition than in the fast tempo condition. The slow tempo music also resulted in customers walking at a slower pace than when the fast tempo music was playing. In terms of the customer's awareness of the music, there were no significant effects identified due to the music condition. Milliman concludes that the slower tempo music led to customers moving more slowly through the supermarket and, as a result, spending more money (Milliman, 1982).

Another study investigated the effect of genre on purchasing behaviour in a wine shop (Areni & Kim, 1993). The shop's audio system varied playing classical and top 40 music over two months. When accompanied by classical music the customers spent more money than with top 40 music playing because they purchased more expensive products. This result was explained by the classical music providing a more sophisticated atmosphere than the top 40 music which primed the customers to feel more culturally refined and led to them purchasing more expensive wine.

Another study focused on purchasing wine in a supermarket by exploring whether the nationality of music played in the wine aisle affected the types of wine that were purchased (North, Hargreaves, & McKendrick, 1999). Shoppers bought significantly more French wine when the music played in the wine aisle was French and significantly more German wine when the music was German. This study was performed in a UK supermarket so was not confounded due to the country of investigation. Further, the customers did not believe their wine purchases were influenced by the music they heard, though they did often accept that the French music made them think of France and vice versa. The authors state "the finding is consistent with the notion that music can prime related knowledge and the selection of certain products if they fit with that knowledge" (North et al., 1999, p. 274).

4 This operationalisation of the tempo variable has been used in many other experiments investigating the effect of tempo on people. Some of these experiments are reviewed in this chapter.

The vast majority of the research investigating the effect of music on shoppers is focused on purchasing behaviour in physical shops, i.e. offline. However, there has recently been some interest in the effect of music in an online shopping context as well. One paper describes two studies investigating the impact of music on shopping behaviour in an online context by considering how the tempo of music affects purchasing behaviour for different types of products (Ding & Lin, 2012). These studies found the tempo of music positively affected arousal which led to stronger purchase intentions for hedonic products. But, with utilitarian products, increases to arousal which were manipulated by changing the music's tempo, did not affect purchasing intentions. This result is said to reflect "that recreation-oriented shoppers prefer a more exciting atmosphere, whereas task-oriented shoppers prefer simpler merchandise presentations" (Ding & Lin, 2012, p. 305), a theory which had been previously explored by Kaltcheva and Weitz (2006).

The methods used in all the studies involving physical shops used data from real shoppers who purchased products using their own money, so these studies have very strong ecological validity[5]. Contrastingly, Ding and Lin's (2012) online shopping study used fictitious websites in a laboratory based methodology. The participants were not spending their own money on products they wanted to own, instead they just reporting their purchasing intentions, so this experiment had poor ecological validity and it is not clear the results would extend beyond the controlled experiment if online retailers were to begin to incorporate music in their websites.

The difference in quantity of research attention in the offline rather than online shopping domain reflects how exposure to music is typical and perhaps even expected when shopping in physical stores. The music is chosen and controlled by the retailer so understanding how profits can be increased in a manner a retailer could implement without significant protest from customers is a sensible endeavour. In an online shopping context, exposure to music is rarely controlled by the retailer. Although many people may shop while listening to music of their own choice, it is not typical for a website to force their music on their online customers. In fact, it is questionable as to whether consumers would be happy about websites adding music to the online shopping experience. There is a clear distinction between likelihoods of consumer acceptance for exposure to music within

5 Ecological validity is focused on how well or poorly research relates to the real world. Studies performed in a laboratory setting can have poor ecological validity as they can be somewhat contrived meaning the results may not generalise the real world. Alternatively, studies like Milliman's (1989) shopping research has strong ecological validity because the data was collected from genuine shoppers spending their own money on products they wanted to purchase.

a physical shop and one that is accessed online. What is acceptable with music in an offline environment is not the same as online, which goes to explain why music is atypical of an online shopping experience and why research attention has not been focused in this domain.

2.1.2 Dining

The marketing research investigating the effect of music on customers is not limited to purchasing products from shops. The effect of exposure to different types of music has also been explored in a few studies involving dining behaviour. One study investigated the effect of music style on behaviour of customers in a student cafeteria environment over four days (North & Hargreaves, 1998). The effects of contemporary British pop, classical, and easy listening music were compared to a without music condition by playing music from the different genres on different days. The sales figures from the cafeteria were collected and compared to the same day in the previous week and the week following the test. Participants completed questionnaires asking them to determine, to the nearest penny, the maximum amount they were willing to spend on 20 food items. The questionnaire also asked for opinions as to whether the cafeteria met certain characteristics, e.g. by asking was the cafeteria fun? Participants who dined while exposed to the music were also asked questions about the music, including whether it contained the same characteristic adjectives as the cafeteria. The idea was that should the cafeteria be described as happy the music would also have a characteristic match as happy music.

The patrons exposed to classical music were willing to spend significantly more money than those dining without music or while hearing easy listening music. However, there was no difference between classical music and contemporary British pop music in terms of the maximum amount of money people were willing to spend. Again, this is reflective of the idea that classical music is more sophisticated and leads to an increased willingness to spend more money over easy listening music but does not explain why there was no difference between classical and British pop music. There was also clear evidence the music's genre affected the participants' perception of the cafeteria, for example, music that was described as fun was played on days the cafeteria was described as fun.

Two studies have looked at how music tempo affected how long diners spent in a restaurant (Caldwell & Hibbert, 2002; and Milliman, 1986). In both experiments, slow and fast tempo music was played in restaurants over a number of days, both using the operationalised definitions of tempo from Milliman (1982). Both studies found the patrons spent significantly more time dining in the restaurant when the tempo of the music was slow. In the experiment reported by Caldwell and Hibbert (2002) both food and drinks sales were significantly higher in the slow tempo condition. In Milliman's (1986) study, drink sales were higher with slow tempo music but there was no effect of tempo on food sales. However, when the gross margin was considered by combining the sales figures and profit margins from food and drink, the effect of tempo was significant with higher gross margin with slow tempo music.

As with the offline shopping examples, these studies had strong ecological validity as were performed in real cafeterias and restaurants with participants spending their own money on food they would consume. Though, unlike shopping there is no obvious online equivalent of the dining experience, except perhaps for purchasing take away food through a website, but this would be subject to similar concerns to the online shopping context regarding the appropriateness of forcing the customers to listen to music chosen by the establishment.

2.1.3 Gambling

Music is an important feature of the real world gambling experience as, similarly to shopping, music is used as an atmospheric in casinos. However, unlike the research investigating the effect of music in shopping and dining contexts, I have not been able to identify any empirical research using NHST that take place in a real, genuine offline gambling environment.

One empirical study investigated the effect of music tempo and loudness on slot machine gambling when compared to typical ambient casino sounds using a laboratory based methodology (Noseworthy & Finlay, 2009). In this research, the music led to improvements to estimates of duration of play over the ambient casino sounds, particularly when accompanied by slow tempo, high volume music.

Although there have been no studies performed in actual real world casinos, there has been some research attention investigating the impact of music on gambling behaviour using laboratory based methodologies focused on online gambling. For example, the effect of background music tempo on gambling behaviour when playing online roulette has been investigated in two studies (Spenwyn, Barrett, & Griffiths, 2009). The risk of the bet (i.e. the amount of money spent) and the speed at which bets were placed were recorded for no music, slow and fast tempo music conditions (Dixon, Trigg, & Griffiths, 2007). The definitions of slow and fast tempo were again taken from Milliman's (1982) supermarket study. The music's tempo had no effect on risk taking behaviour in these experiments. However, the speed at which people placed bets was significantly higher in the fast tempo condition. A similar study by Spenwyn et al. (2009) concluded similar results. Here the authors speculated the relationship between tempo and speed of bets is due to the increased arousal felt by participants in the fast tempo music condition. The authors also propose that online gambling websites should have an option to turn off music as players are more likely to become addicted with fast tempo music accompaniment as the time for contemplation between bets is reduced, even if their overall risk taking behaviour was not affected. Nevertheless, they also acknowledge that some websites or casinos may wish to profiteer by ensuring the tempo of any music playing is fast, thus encouraging faster betting with less time for someone to consider the consequences of placing a bet.

Given the limited empirical attention given to the impact of music on gambling behaviour in real world casinos and arcades it is interesting that the virtual world equivalent has received considerably more empirical attention. This contrasts with the shopping situation where there has been much research in a physical context with minimal attention in the corresponding online setting. Perhaps this is due to the perception that as online gambling is an entertainment activity somewhat reminiscent of playing computer games it is more acceptable for a gambling website to play music. Alternatively, the aims of these studies may be to apply the results from the laboratory style study performed using the online gambling experience to the offline alternative.

The results from the online gambling research show that even in laboratory based gambling context music has a substantial impact on the players' behaviour. The participants in these studies did not place bets with their own money therefore conceded no financial risk, nor could they win any real money from the gambling activity. Despite this lack of potential benefit or reward clear outcomes were identified through significant effects. Although the ecological validity of studies conducted in this manner is reduced, it seems that significant outcomes can still be achieved, verifying the acceptability of the method.

2.1.4 VIDEO GAMES

Developers of video games exploit music, sound effects and speech within games to create the optimum gaming experience. Therefore, one might expect a large quantity of empirical research investigating the effect of game audio on the player. So, it is somewhat surprising that there are relatively few empirical studies focusing on how audio affects the gaming experience (Hébert, Béland, Dionne-Fournelle, Crête, & Lupien, 2005; and Lipscomb & Zehnder, 2004).

Researchers have begun considering the impact that music has on gamers' levels of immersion. A study by Sanders and Cairns (2010) identified that music preference, i.e. whether the gamer enjoyed the music or not, significantly impacts gamers' immersion levels. The aim of the original study was to use music as a means of manipulating immersion in an investigation of the relationship between immersion and time perception. However, during the first run of the experiment participants in the music condition were significantly less immersed than those in the without music condition. The initial choice of music was described as "a fast paced piece [that was] also quite jarring and discordant" (Sanders & Cairns, 2010, p. 163) which the participants did not enjoy. As such, a different choice of music, validated by 10 people as being enjoyable to listen to, was used in a repeat of the initial experiment. In this instance, the music condition significantly increased immersion when compared to the condition where there was no music playing. The unintended outcome from this study, which initially focused on the relationship between time perception and immersion, was that enjoyment of the music is a key factor influencing players' immersion. From a musical perspective, they conclude that built-in game music that people enjoy can have a positive effect on immersion. Conversely, if the game includes music the players do not enjoy immersion is reduced.

Another computer gaming study took an objective measurement approach when investigating the physiological stress response due to built-in game music (Hébert et al., 2005). Fifty-two male participants played Quake III Arena for 10 minutes in either a without music or with music condition where the built-in game music was used, which was described as being in a pop-techno style. In both the music and without music conditions the volume of sound effects and speech was set to zero to ensure the only difference between conditions was the presence of the game music itself. Saliva samples were taken from the participants before the experimental game-play and at three intervals post-game playing, immediately after completion of the game, after a 15 minute rest period, and after 30 minutes rest. The levels of cortisol in the saliva were measured with the authors hypothesizing that cortisol levels 15 minutes after playing the game would be significantly higher in the music condition as this is “the moment around which cortisol reaches its maximum value after the beginning of a stressor” (Hébert et al., 2005, p. 2373). Their hypothesis was validated with a statistically significant difference in cortisol levels 15 minutes after game play, with an increase in the music condition when compared to the without music condition. This result demonstrates that the games’ music had a physiological effect on the players.

An earlier quantitative empirical study compared gamers’ experiences in three conditions, the game with music, the game without music and with music only (Lipscomb & Zehnder, 2004). Although the title of this paper “Immersion in the Virtual Environment: The Effect of a Musical Score on the Video Gaming Experience” suggests that the researchers investigated gaming immersion, the measurement scales employed in this study seem to focus on experiential aspects that are arguably not factors of games’ immersion. The participants reported verbal attribute magnitude estimation scales by rating the following adjectives: “active, annoying, bright, busy, cold, colorful, dangerous, exciting, fast, gentle, good, high, intense, labored, loud, masculine, pleasant, powerful, relaxed, simple and strange” (Lipscomb & Zehnder, 2004, p. 339) on scales from not to adjective, for example from not annoying to annoying. Most of these adjectives do not appear to relate closely to immersion, neither do all of them relate to how the participant feels. This discrepancy in intention is because the scale was developed in relation to properties of music played by wind instruments, rather than the listeners’ experiences when hearing the music. Essentially, these adjectives are descriptive of the musical stimuli rather than the person’s responses to listening to the musical stimuli. Given my reservations with regard the measurement technique, and a potential methodological issue (the participants in the music only condition listened to the excerpts for shorter durations than those in the gaming conditions, and there appear to have been no time limits placed on the gamers, so it is not clear if time spent playing the game will have impacted on the verbal scale responses) the results and their conclusions are not presented in detail here. Though the authors do reach the conclusion that the “relationship between a specially-composed orchestral musical score and involvement with an RPG video game is highly complex and varied” (Lipscomb & Zehnder, 2004, p. 339-340). This conclusion is vague and is reflective of the quality of the validity of the experiment.

Researchers have also investigated the impact that music has on performance in driving simulation games. However, in this particular gaming genre, the driving simulation is often used as a means of investigating what would happen in real-world driving situations. Obviously, it is not possible for ethical and safety reasons to purposefully reduce people’s driving performance in real cars on real roads, so simulations are often used as an alternative. In this context, the studies do not necessarily investigate the affectivity of music in order to establish a means of improving the gaming experience; the focus is more on the ability of drivers to be safe in the non-virtual world. This is one of the occasions where, the distinction between human-computer, human-system, or human-machine interface is distorted.

However, there have been a couple of studies in a driving game environment where the researchers are not attempting to recreate a real-world driving experience, rather they are using the driving game for entertainment. For example, in a study by Cassidy and MacDonald (2010) the effect of music preference on driving game performance and enjoyment was investigated. In situations where the participants self-selected the music to accompany the driving game they enjoyed the experience more, whilst performance and experience diminished when the experimenter selected the music.

2.1.5 Typing and Word Processing

In 1931, the effect of jazz and dirge music on transcription typing performance was compared to a without music condition (Jensen, 1931). Fifty students who had received 37 weeks of typing tuition were exposed to music in their typing classes on three consecutive days. Each day the participants completed three 5 minute transcription typing tasks under the three different music conditions. The music was played using “a heavy steel needle . . . in playing all the records, the tempo was kept normal and no changes were made in volume control” (Jensen, 1931, p. 458). There was an increase in errors with jazz music when compared to the without music and dirge music conditions, but there was no corresponding effect on typing speed. The dirge music reduced the speed of typing when compared to the without music and jazz music conditions, but there was no corresponding effect on error count. Jensen concludes this leaves “no doubt as to the seriousness of the influence of jazz music on typing, so far as errors are concerned” (Jensen, 1931, p. 460).

A more recent study by Jiang and Sengupta (2011) looked at the effect of music on transcription typing tasks under the banner of word processing. The participants brought their own music to listen to while completing the transcription typing tasks. The software used in the tasks “allowed participants to return and correct wrong letters in one word, but once the space key was pressed for that word, participants could not return and correct the word” (Jiang & Sengupta, 2011, p. 3262) which is somewhat unusual as this restriction does not really match how people typically type and correct their errors. In the experiment, a lot of data was collected regarding the physical process of typing, including measuring typing force and attaching electrodes to the participant’s finger to collect electromyography (EMG) data from the left hand extensor digitorum muscle. Music was found to significantly reduce accuracy and the authors conclude this is due to the “distracting influence of music” (Jiang & Sengupta, 2011, p. 3264).

2.1.6 Software Engineering

One study looked at whether listening to music prior to or during software programming tasks reduced state anxiety and affected the quality of the code produced (Lesiuk, 2000). The objective of this work was to understand if music could be used to improve the mood of stressed software programmers by reducing their anxiety.

In this study, undergraduate students undertook computer programming tasks within their programming practical classes in different experimental conditions. The control group did not hear music, the primer group heard 11 minutes of classical music prior to completing the programming tasks while the periodic group heard classical music both prior to and during completion of the programming tasks. Two programming tasks were used in the experiment, the first focused on correcting syntax errors while the other required the students to locate logic errors. It is not clear from the paper how the music was played to the students, but the assumption is that the music was played over a speaker system rather than using headphones. Anxiety was measured using the State Anxiety Inventory Form (STAI) at the start of each practical class, and then for the primer and periodic groups the STAI was completed after the 11 minutes of classical music was played prior to commencing the programming task, and then again after completion of the tasks.

So, the control group completed 2 STAIs, while the primer and periodic groups completed 3 STAI forms.

Initially all the students had similar levels of state anxiety, but after 11 minutes of listening to music the students in the primer and periodic conditions had significantly lower levels of anxiety than the control group. On completion of the programming tasks, there was no longer a difference in state anxiety between the control and primer groups, but the difference between the control and periodic groups was significant. These results show that hearing classical music prior to completing the programming task had the effect of reducing anxiety and that by continuing to hear this music throughout the programming task the students' anxiety was kept lower.

In terms of the quality of the code produced, there were no significant differences in either the syntax or logic tasks, though the authors note that the mean scores of quality were higher in the periodic music condition. Although there was no improvement in the quality of the programming output from the students, the improvements to their anxiety levels is interesting and suggests music has the potential to positively influence the mood of software programmers while at work.

A follow up study was performed in four Canadian software development companies over a five week period (Lesiuk, 2005). During two of the weeks the participants were instructed to listen to whatever music they wanted to whenever they wanted to. They could either listen to their own music or choose from a set of 65 CDs which provided music from a variety of genres. Then, in another week, the participants were instructed that they should not listen to music at all and had to sign a document each day confirming that they had not listened to music that day. In the final week, the participants could choose if they listened to music or not. State positive affect and quality of work were lowest in the without music week and the time spent on the work was highest. Lesiuk (2005) concluded that music listening can evoke positive feelings in work environments which also improves performance in creative tasks.

2.2 How Sound is Used in Computing

The previous section explored the effect of music on people's behaviour and emotions from a variety of contexts, including those relating to technology. There are some clear examples of research investigating the affectivity of music in technological contexts, which seems to be a growing field. Historically, audio has been used in computing with the objective of communicating messages to the user. In this section, a review of research regarding how sound is used in computing to communicate messages is presented.

Interactive technologies frequently exploit sounds as part of the user interactions, but music is not often included in the interface to directly affect the user's behaviour or experience, although research attention in this area is growing (e.g. see review of research on affecting software developers with music, or in areas where the computing technology is an integral part of the activity itself, like with online gambling or video games). Given most of the attention has been focused on using sounds, rather than music, it is interesting to explore the motivations behind incorporating sounds in interfaces by reviewing the literature regarding the use of audio within interactive technologies.

My analysis of the auditory interaction literature identified that the overriding primary motivation for incorporating audio within an interactive technology is to communicate information to the user. Information is encoded in the form of an *auditory message* which the user hears, decodes and then responds to in some way. These auditory messages can be created using speech or non-speech audio (NSA).

To use speech based audio within interactive systems the messages must first be created either using a speech synthesizer or by recording samples of speech. Text-to-speech synthesis can be realised using a database of pre-recorded sounds that represent whole words or parts of words. Alternatively, for pure speech synthesis a model of the vocal tract can be developed, which includes other characteristics of human speech.

The term non-speech audio can be refined into categories based on the types of sounds used, as follows:

- **sound effects:** such as auditory icons (Gaver, 1986),
- **pitched tones:** either individual tones, simple tones or as collections of tones (e.g. Earcons, see Blattner, Sumikawa, & Greenberg, 1989),
- **generated music:** typically generated in real time using MIDI but can include combining audio samples on the fly,
- **pre-recorded music:** either pre-recorded pieces of music or samples (e.g. musicons, see McGee-Lennon, Wolters, McLachlan, Brewster, & Hall, 2011) or created using MIDI in advance.

It can be hard to differentiate between pitched tones and generated music. "At what point does a collection of pitched tones become music?" is an interesting philosophical question that is, arguably, outside the scope of this thesis. But for consistency within this chapter, collections of pitched tones are only categorised as music if either the researchers describe the sounds in that manner, or they have considered the artistic merit of the sounds in their selection. For instance, Francioni et al. use music in their parallel programming debugger interface as they intentionally chose their sounds "to create tension on the send that is resolved with the following receive melody" (Francioni, Albright, & Jackson, 1991). It is clear that some considerable attention has been given to the sound design in this instance, so the collections of short melodies are designated as music rather than pitched tones. Alternatively, earcons are specifically designed so as not to be musical, so earcons are labelled as pitched tones.

The information that these auditory interfaces attempt to convey to users varies widely. After analysis a range of auditory interaction literature I identified four key motivations for using sounds to communicate information. These motivations involve using sound to:

- aid accessibility,
- improve usability,
- provide status information,
- support understanding

The auditory interaction literature is reviewed in the following sections focusing on the types of sounds used and the motivations for their inclusion in the interface. Again, there is a vast amount of literature from the field of auditory interaction, so this review is presented as an introduction to the types of sounds used and the effect of their inclusion in the interface on the users, rather than as a systematic review of all the auditory interaction literature. As my research focused on using music, a particular emphasis has been placed on literature that used generated or recorded music in the interface.

2.2.1 Aid Accessibility

One of the original objectives for exploiting the auditory modality in human computer interaction (HCI) was to help make computers accessible to users with visual impairments. Screen readers are an early example of technology that was designed to assist users with visual disabilities to interact with computing interfaces through the auditory channel. When used in combination with a text based interface, these auditory screen readers spoke the same text (albeit in a synthesized voice) that sighted users were able to read from the screen. Thus, users with visual impairments were provided with the same information that was conveyed to sighted users.

The development and adoption of graphical user interfaces (GUIs) as the dominant means of interacting with systems put users with visual impairments at a significant disadvantage to sighted users. This disadvantage is due to GUIs including elements that screen readers struggled to convey through speech. For example, GUIs regularly exploit object positioning, colour and visual icons to communicate information to users; techniques that are difficult to depict through speech. Interacting with these GUIs, which rely on graphics to communicate information, was very difficult, if not impossible for visually impaired users. The auditory channel has been exploited using non-speech audio to help overcome the difficulties involved in representing graphical interfaces using speech.

There has been a wealth of research in the area of auditory assistive technologies since the 1980s. Auditory interaction systems are designed as an alternative to interfaces that employ visual or spatial interaction modes. The aim of these auditory interaction systems is to assist non-visual users to interact with a particular technology or device. Auditory interaction systems tend to rely on a combination of both speech and non-speech audio. For a visually impaired user to successfully interact with an auditory interaction system, information must be conveyed in a precise and clear manner, which explains why these systems tend to exploit speech audio as well as non-speech audio, though the speech is often employed redundantly. In other words, the information available via the speech aspect of the interface is also available from the non-speech audio. But, if the user cannot translate or understand the messages portrayed through the non-speech audio then the user is able to access the information via the associated speech based audio. Two examples of systems which use sounds to convey information to visually impaired users as an alternative system are SoundTrack (Edwards, 1989) and ACCESS (Morley, Petrie, O'Neill, & McNally, 1999).

Other systems represent visual information audibly, instead of simply providing an alternative to the visual aspects of the interaction. The aim of these systems is to present complex visual information in an alternate manner so that non-visual users can understand it. Given the complexity of the information portrayed by these systems, and the speed with which it ought to be communicated, speech is not necessarily an appropriate means for this type of message. However, different types of non-speech audio have been used in a wide range of systems. For example, MIDI generated music was used in AUDIOGRAPH (Alty & Rigas, 1998) to communicate diagrams to users while sound graph used sine waves (i.e. pitched tones) to depict line graphs (Mansur, Blattner, & Joy, 1985). Although these systems have been designed with an accessibility focus, other systems have been developed which attempt to portray complex visual information using sounds as well. The aim of these systems is not to assist non-visual users, rather to provide a complementary representation that can aid understanding of the domain by anyone. See section 2.2.4 for more on non-accessibility focused systems using audio to describe complex visual information.

A number of auditory navigation aids have also been developed to assist visually impaired users to negotiate specific environments. These devices often employ speech based audio when precise information is required, e.g. in the Ping! interface the user first engages with a telephone based interface (Landau, Wiener, Naghshineh, & Giusti, 2005). As well the speech based audio, these systems typically include at least one type of non-speech audio. In the case of Ping! the non-speech audio is a personalised auditory beacon that the user selects via the telephone interface and then moves towards to reach their destination. While, in the Accessible Aquarium system (Pendse, Pate, & Walker, 2008) a number of different sound types were trialled, including a MIDI version of Johann Strauss' The Blue Danube with each fish represented by a different instrument. Again, auditory navigation systems are not always created specifically to support visually impaired users. Other non-accessibility focused auditory navigation systems are described in section 2.2.3.

2.2.2 Improve Usability

The development of the WIMP (window, icon, menu, pointing device) paradigm to interact with GUIs led to another objective for using the sounds to provide information, that of improving effectiveness and efficiency when using the interface. A cornerstone of HCI for many years has been the idea that improving usability, in terms of effectiveness and efficiency in interaction, will lead to users that are satisfied. As such, auditory messages have been investigated as a means of improving usability aspects of GUIs.

The two original formats for non-speech audio sounds were the auditory icon (Gaver, 1986), a sound effect, and earcons (Blattner et al., 1989), a collection of pitched tones that form a motive. The aim of including these sounds in GUIs was to improve the effectiveness of users when interacting with GUIs. In other words, the aim was to improve the usability of the interface.

One advantage of using sound effects in GUIs to improve usability is that the meaning associated with the sound can often be representative of the visual effect, whether this is an icon or an action, for example, moving a file to the recycling bin will trigger the sound of paper thrown into a rubbish bin. The sound employed relates closely to the users' intended action. In this instance, the sound improves usability by informing the user that the file has successfully been moved to the recycling bin. The user does not need to go and check the recycling bin to see if the item now resides there, as they understand that the rubbish bin sound means the item was deleted. There are many examples of auditory icons incorporated in modern day operating systems and commercial interfaces. It is worth noting, however, that many GUI interactions do not have a sensible sound effect that can be associated with them. So, some sounds employed as auditory icons can be relatively abstract, requiring learning before the user can successfully interpret the messages the system wishes to convey.

The other method for improving usability of GUIs with non-speech audio is to use earcons which are a pattern of pitched tones (up to 4 or 5 separate notes) that collectively form an auditory motive (Blattner et al., 1989). These sounds are a complete abstraction from the action or item they represent so the user must learn their meaning. Earcons have been the focus of investigations with GUIs in both desktop computer and mobile devices, although commercial products do not typically adopt this method of auditory message communication for usability.

In terms of improving usability, neither speech based audio nor music based techniques seem to have been explored. Instead, the types of sounds used in systems are either sound effects or earcons. This design is probably because to improve usability of systems the sounds need to provide immediate feedback to the users. To be successful, this process needs to be quick, so the sounds' duration must be short. Therefore, given the longevity associated with speech and music these are both formats that are inappropriate for use in improving aspects of usability.

In desktop computing interfaces visual elements of the interface such as icons, buttons and scrollbars, can be sonified to improve the user's effectiveness and efficiency. For example, Brewster investigated sonically enhanced buttons using earcons (Brewster, 1997) with significantly improved performance and lower workload reported.

Sounds are also an important feature of mobile devices, particularly those with touch screens as with minimal screen real estate available on mobile devices, auditory components in the interaction can dramatically improve the usability, in terms of efficiency and effectiveness. Sound effects are incorporated within most commercial touch screen mobile phones, and the effect of different auditory types on accuracy and user errors has been investigated (Brewster, 2002).

2.2.3 Provide Status Information

Sounds are also regularly investigated within auditory interaction as a method of providing information regarding the status of an artefact or environment. Sounds are often employed in situations that require monitoring, as they are difficult to avoid or ignore. While the presence of a particular sound will convey information to the user, the absence of a sound can also communicate important information.

Sounds can be used for event notification and monitoring systems. In these contexts, sounds are used to notify the user that a particular event has occurred, or something has changed in the ongoing behaviour or activity within the environment. The effectiveness and suitability of different types of non-speech audio have been researched numerous contexts, for example in email monitoring systems (Hudson & Smith, 1996), care homes (Kanai, Nakada, Hanbat, & Kunifuji, 2008), non-urgent medical systems (Sanderson et al., 2008) and in the home (McGee-Lennon et al., 2011).

Alarms typically make use of the auditory modality as well, with sounds used to alert the user and/or people in the vicinity, that an exceptional event has occurred. The purpose of an alarm is to alert a user to the occurrence of a significant event, normally something that requires immediate attention. The information conveyed by the sound and the ability of the user to quickly interpret that information is of vital importance with alarms. Alarms can be considered an exceptional form of event notification. One interesting aspect of research into the suitability of sounds for alarms is that substantial importance is attributed to the affectivity of the sounds. For example, research has focused on trying to understand how sounds portray urgency to the user and how they influence the users' resultant behaviour (e.g. in Arrabito, Mondor, & Kent, 2004; Burt, Bartolome, Burdette, & Comstock, 1995; and Edworthy, Loxley, & Dennis, 1991). So, a key focus of much of the research is the alarms domain has been on the affectivity of the sounds. Alarms are designed predominantly to convey information to the user, the affective element of the design is to ensure that the message being conveyed contains the appropriate information.

Navigation systems also often exploit sounds to help users to monitor their position in a particular environment, while notifying them when they have travelled in the wrong direction (e.g. by providing an event notification). Audible navigations systems have received a considerable amount of research attention, and not just with human participants, e.g. the effect of sounds on navigation capabilities of rats was investigated (Rossier, Haeberli, & Schenk, 2000). In research focused on human users, music has been used to support navigation in gpsTunes (Strachan, Eslambolchilar, Murray-Smith, Hughes, & O'Modhrain, 2005), as well as spatial audio (Holland, Morse, & Gedenryd, 2002) including use of animal sounds (Stahl, 2007).

All non-speech audio types, including music, have been used in navigation aids and to monitor events. Part of the explanation for this is that monitoring an activity tends to be continuous in nature, taking place over a prolonged period; so, it is appropriate to use music (along with other non-speech audio types) as a means of conveying the information. However, these systems rarely incorporate speech based audio. The lack of speech audio, particularly with the navigation based systems, contrasts with the accessibility focused navigation aids that often use speech audio as well as non-speech. Music is not an appropriate method of alerting users to the occurrence of an exceptional event, i.e. an alarm, which is again probably due to the transient nature of music and the urgency of information communication required in alarms.

2.2.4 Support Understanding

The final motivation I identified for using sounds in technological interfaces to communicate information was to complement the visual representation of a complex domain to support understanding of that domain. Presenting complex information audibly as well as visually can help users to gain enhanced understanding of difficult topics.

In computer programming sounds have been used to provide information to help programmers understand different aspects of their code, such as the existence of repeating loops, or the outcomes from conditional statements. Generated MIDI music has been used to help novice programmers to understand their code and help with debugging, (e.g. in Francioni et al., 1991; and Vickers & Alty, 1996). Sounds have also been used to help people to understand complex algorithms through auditory representations of how the algorithm, for example, sorts data, (e.g. in Brown & Hershberger, 1992).

2.3 Conclusion

This literature review has focused on two areas that are important when considering how music can be exploited to improve performance at work related computing tasks. The first section of the review focused on research investigating how music affects people's behaviour across a variety of contexts. Although not an exhaustive review of the literature relating to the use of music in all contexts and situations, a few key areas were considered including shopping, gambling and software engineering.

In this empirical work using NHST, the manipulation of tempo is a frequent occurrence but the definitions of slow and fast are sometimes inconsistent. Many studies adopted Milliman's operationalisation of slow and fast tempo in terms of bpm, but in other studies (e.g. in Thompson, Schellenberg, & Letnic, 2012) the values used to define slow and fast tempo music are different. Volume aka loudness has been investigated, again with varying definitions of the threshold used to designate low or high volume music. Experiments have also frequently involved self-selection of music by the participants. The research has been performed both using laboratory style methodologies and manipulations in real world environments, though the latter is typically constrained to setting where exposure to music in that context is typical.

Next, a review of literature regarding the use of sound in computing interfaces was presented focusing on four primary motivations for using sound to communicate information to users via auditory messages. Sounds are used to communicate information alone and also redundantly to provide information that is also available to the user through another sense, typically vision. This model for the use of sound to communicate information as a means of changing behaviour can lead to increased performance at work, and in fact that is one motivation that is often reported for exploring the use of sound in technological interfaces.

There has previously been some research into the annoyance factor of sounds in interfaces and the effect of being annoyed by a sound has on performance. But, when it comes to sounds in computing research the primary aim has been to convey information via sound which then affects the user behaviour or experience. The exception to this, is when music is used in an interface. Music, as has already been discussed, has been used as an affective medium within a vast range of domains and with a wide variety of motivations, including those involving technology.

Chapter 3

Introduction to the Experiments

The review in Chapter 2 showed music can affect people's behaviour and experiences across a variety of circumstances and in accompaniment to numerous activities. There have only been a few examples of empirical research investigating this affectivity within a work based setting, and even fewer examples involving computing related tasks. Exposure to music while working at a computer to improve performance has been somewhat neglected by the research community to date. Given the prevalence of computing activities in the workplace and the ubiquitous nature of music, the seemingly minimal scientific attention applied to exploiting the affectivity of music in relation to improving performance and experience with work related computing tasks is perhaps surprising. Of course, it might be that listening to music is often considered a leisure activity, so the premise does not align well with being used to positively affect performance in work related computing tasks. It is possible that listening to music is often perceived as a leisure rather than work activity which has led to a situation where a medium that could, potentially improve productivity is not being utilised. The experiments reported in this thesis aim to extend investigations into using music as a positive influence in work related computing tasks.

The concept of working at a computer potentially has a vast scope, so it was necessary to focus my research on a simple, clearly defined work related computing task. So, in this programme of research I performed five experiments which all investigate the effect of music on transcription typing performance and experience. The important questions “what does typing performance and experience mean? And, how are they measured?” are addressed in Section 3.3.1.2. Typing on a keyboard is one of the fundamental mechanisms used to interact with a computer. Many everyday computing tasks involve users entering text using a keyboard. If a user’s typing performance can be improved with music, it follows that their overall work performance can also be improved.

The typing tasks used in my experiments all involved the participant copying text presented to them visually on a website into an editable text field on the same website. The task is, therefore, a transcription typing, rather than free typing, task. It is unlikely that transcription typing forms part of many people’s work related computing tasks, but the task is representative of mundane, monotonous computing tasks that many people undertake at work. The interaction method, i.e. typing, under investigation is universal even though the precise nature of the task itself is somewhat contrived. This precisely defined task helps to ensure strong construct validity[6] as the experiments measure typing performance rather than typing and thinking performance.

The experiments described in this thesis expand on Jensen’s (1931) previous research which found that jazz and dirge music affected typing performance in different ways. With dirge music accompanying typing, the speed of keystrokes[7] was significantly lower than in both the without music and jazz music conditions, though no corresponding effect on error count was identified. Contrastingly, with jazz music, the error count was significantly higher than with dirge or without music, but there was no evidence that the keystroke speed was also affected. Despite Jensen’s research identifying these significant effects it is not clear why jazz music led to more errors, but dirge music resulted in lower typing speeds and the paper does not include any attempts to explain why the effects were different. Clearly there was something about the jazz music that made the typists more error-prone, as well as a particular quality of the dirge music that resulted in slower typing - but Jensen does not present a discourse in these terms. Presumably there was also some motivation behind performing this particular experiment to compare the effects of jazz and dirge music to a without music condition in the domain of transcription typing, but unfortunately a justification for the initial hypothesis was not provided. Nevertheless, significant effects were identified and the conclusion that jazz music is detrimental was

6 Construct validity means the dependent variable being measured and analysed will provide an answer to the research question being asked and not something different.
7 The definition of “keystrokes” adopted in this study is not clear. See section 3.2 , page 6 for further discussion.

resoundingly stated, so it is interesting to take this research further to consider whether music still reduces transcription typing performance, or if it can improve performance, and to explore potential reasons why music has these effects.

This chapter provides an introduction to the five experiments reported in this thesis. First, by considering how typing technology has advanced and how the role of typing in people's everyday lives has developed since the 1930s and by defining some key concepts and terminology used in this thesis. Key aspects of the experiment's methods are described, focusing on features used in more than one of the experiments. Although many aspects of the method are similar, two different methodologies were used in the experiments. The first was a standard laboratory based approach, and another used a more unusual classroom based methodology where the students ran the experiment in order to learn how to perform empirical research. This classroom based methodology introduced some important ethical and design considerations which are discussed. Finally, the quantitative and qualitative techniques used to analyse the data are described and justified.

3.1 The Evolution of Typing

It has been at least 80 years since Jensen (1931) would have collected the data in his experiment. During this period there have been considerable changes to the technology used in typewriting[8]. Unfortunately, the model and brand of typewriter used by participants in the experiment was not described. However, given the study was performed in the early 1930s it is likely that the participants used a mechanical typewriter to type directly onto paper with a QWERTY layout, such as that shown in Figure 3.1.

By 1931 typewriters included a shift key so that two values were assigned to each key, allowing both uppercase and lowercase characters as well as other special symbols and punctuation. Further, a backspace key would have been available to erase any mistakes. In terms of moving the current typing position on the paper, there were neither arrow keys nor, obviously, a mouse.

8 The terms "typing" and "typewriting" are synonymous in meaning. Typing tends to be the more modern term, while typewriting is more traditional.

Figure 3.1: Example of a mechanical typewriting interface from the early 1900s[9]

With the introduction of the personal computer, typewriting technology changed dramatically. Instead of typing directly onto a piece of paper people began typing onto a screen based representation of a piece of paper. Figure 3.2 is a photo of the Kensington keyboard that was used in all the experiments reported in this thesis, which contrasts starkly with the technology available in 1931, shown in Figure 3.1.

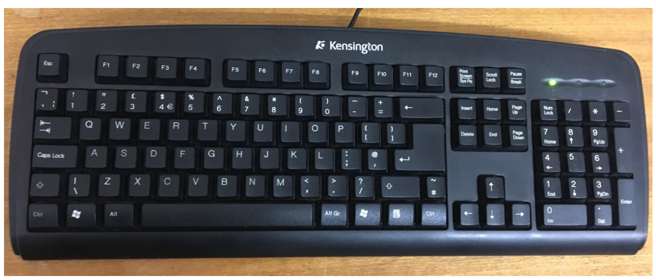


Figure 3.2: Photograph of keyboard used in all experiments

Although modern keyboards are drastically different to 1930s typewriters, the QWERTY layout of keys persisted with a few additions and changes to functionality. For example, when using a modern keyboard interface, the act of pressing and holding the backspace key erases multiple characters, meaning the user only needs to hit the backspace key once to repeatedly perform the deletion action. Instead, using a 1930s typewriter to achieve multiple deletions typists would need to repeatedly hit the backspace key.

Another area where keyboard technology has evolved is with the addition of the concept of positioning using a cursor. Typists can now move the cursor to adjust their current position in the text so that additional symbols can be inserted by pushing the already typed text along to make space for new characters. Arrow keys have been added to the interface, so typists can move their cursor location and with some modern keyboards a

9 Image taken from http://filthyplaten.blogspot.co.uk/p/5-most-wanted.html

delete key can be used to delete the character located after the cursor's position rather than the one before it. Further, with Graphical User Interfaces (GUIs) a mouse or trackpad device can be used to change the cursor position without needing to interact with the keyboard at all.

One final important difference between mechanical and modern, electronic typing interfaces, is that software now often includes checking functionality that can identify, highlight and automatically correct typing errors. This software may have resulted in negative consequences for typing performance, particularly in terms of accuracy, as users can now expect the software to automatically correct some of their mistakes and to visually highlight mistyped words or incorrect grammar on screen. Some typing mistakes will not be identified by the checking functionality, but many errors are highlighted and are automatically corrected without the user having to do anything to trigger the correction. This functionally may have led to typists who are less focused on typing accuracy than speed as they can rely on the software to correct or highlight many types of errors. In the 1930s, typists had to be aware of and correct all their mistakes as they were typing, which may mean they focused more on accuracy.

Aside from the changes to hardware and software, there have also been considerable changes to how typewriting as an activity has been perceived over the past 80 years. In the 1930s typing was a considered to be a skill that required training and typically only people at school, in college or with an explicit need would be taught how to type. Whereas now typing is a more organic skill that tends to evolve through regular typing activity rather than being taught formally through classes at college or school. Children are now frequently exposed to the QWERTY keyboard interface at school, or perhaps by using their parent's (or even their own) devices. Typing is no longer considered a skill that needs to be formally taught, as instead, in the developed world, most people start learning to type through exposure to the interface in an informal manner from a young age.

In Jensen's (1931) experiments, there were 50 participants (12 male, 38 female) who were students attending the Training School of the Central State Teachers College in Michigan. The participants received "thirty-seven consecutive weeks of typing instruction in high school" (Jensen, 1931, p. 458) and as such, can be considered trained, skilled typists. These skilled typists achieved an average typing speed of 33.64 ± 0.92 words per minute (WPM) under normal typing conditions (i.e. without accompanying music). By today's standards this speed would be considered very slow typing as typically, administrative positions now require a minimum of 60 WPM (Ober, n.d.).

My research aims to extend the work conducted by Jensen, albeit in a modern-day context, so it is important to remember the differences in settings, both technologically and societally, when comparing the results. The participants that took part in the experiments in this thesis are unlikely to be trained typists like the participants from Jensen's experiment. But, that does not necessarily mean these participants are unskilled typists, just that they are unlikely to have had much, if any, formal training. The variability in previous typing experience of the participants in these experiments is larger which may potentially affect the validity of any direct comparisons with Jensen's work.

3.2 Key Concepts and Definitions

To ensure that any comparisons between the results from Jensen's (1931) study and the work in this thesis are legitimate it is important to carefully analyse the paper and to understand and define all terminology used. Unfortunately, Jensen did not provide clear descriptions of many of the terms used in the paper which makes it hard to accurately interpret the results. Nevertheless, the following sections provide an analysis of terms used in the paper together with explanations of the key terms used in this thesis. Providing clear definitions should ensure my research achieves strong levels of scientific rigour and allow it to be carefully scrutinized.

In this thesis, the term *character* is a noun used to describe the symbols produced on screen by the act of pressing a key or keys on a keyboard, while the term *keypress* is a verb which refers specifically to the action of pressing a key. Some characters, particularly those for punctuation and capitalisation of letters can only be achieved with multiple, typically two, concurrent keypresses. There are also some keypresses and combinations of keypresses which trigger functions rather than produce characters, such as pressing the arrow or backspace keys.

In the experiments reported in this thesis, key logging software was used to record all keypresses made by the participants[10]. It is highly unlikely that Jensen would have been

10 An analysis of keypresses per minute (KPM) was performed in all experiments, but this measure did not add any interesting or important insights beyond those identified through the analysis of characters per minute so these analyses have not been included in the thesis.

able to track keypresses as 1930s typewriting technology would not have supported key logging functionality. Instead, it is probable that Jensen looked at the typing output produced by each participant on paper and counted each character. Therefore, it seems likely that Jensen's *keystroke* measure is the same as my character measure.

In this thesis, the term used to describe the text that was provided to participants as material to copy from is *presented text*. While the term *transcribed text* is used to describe the text that was produced by the participant by copying the presented text. The terms *input text* and *input string* are used to refer to the exact sequence of characters and function keys that were entered by a participant. These input strings will typically differ from the transcribed text as the typist will correct some of their typing errors. In this context, *output text* or *output string* are used as a synonym to describe the transcribed text when compared to the input text.

The participant's input string will include characters and keypresses that do not appear in the output text. For example, the typist may enter the combination of characters and function keys shown in Figure 3.3 as an input string, which includes function keypresses that are shown using square brackets. The corresponding output (or transcribed) text is shown in Figure 3.4.

The quick brownfox[backspace][backspace][backspace] fox jumped over the lazy dog.

Figure 3.3: Input string example

The quick brown fox jumped over the lazy dog.

Figure 3.4: Output text which corresponds to the input string in Figure 3.3

3.3 METHOD

This method section forms the basis for all five experiments presented in this thesis. The basic setup is consistent between the experiments. The independent and dependent variables and the design differed between experiments, but there were also important areas of similarity. This section focuses on describing the aspects of the method that are alike across the five experiments.

Three experiments were performed using a standard laboratory based methodology, with the other two using a more unusual classroom based methodology. Section 3.4, page 99, describes the classroom based methodology in detail, including a discussion of ethical and pedagogical concerns when using teaching as a means to collect research data. Before formally defining the classroom methodology, one key difference between the laboratory and classroom approaches needs to be discussed to justify some of the decisions taken in the design of these experiments. This key difference in methodology which affected the experiments' methods is that in the laboratory setting I controlled the running of the experiment completely. I timed the experiment, I selected the typing tasks, I chose the right music for each task, etc. But, in the classroom setting, pairs of students worked together with one acting as participant and the other acting as experimenter. This approach meant that a large amount of data could be collected concurrently in a relatively short period of time, but I handed over control of running the data collection process to an undergraduate student, who was likely to be an inexperienced researcher. In other words, although I oversaw the sessions where the data was collected, a student had control of choosing the correct tasks and performing the experiment.

Sections 3.3.1 to 3.3.3 describe the properties of the design, equipment and materials, and procedure that were used in more than one experiment. Further details are also provided in the individual experiment chapters, as necessary.

3.3.1 Design

Many aspects of the design of each experiment vary depending on the precise aims and hypotheses for the experiment. However, there are some areas of the design that were common to more than one experiment, particularly with regard to the independent and dependent variables. These common aspects are described in the following sections.

3.3.1.1 Independent Variables

The independent variables (IVs) in the experiments either defined the music that accompanied the typing task, or the text that was presented to the participant. Table 3.1 summarises the methodology, music and task IVs used in each experiment.

Experiment	Music IVs	Task IV
Experiment 1 **Laboratory based** Chapter 4	**Vocals** (With; Without) **Style** (Alt Rock; Pop Rock)	None
Experiment 2 **Laboratory based** Chapter 5	**Genre** (Rock; Classical; Ambient)	**Task Difficulty** (Simple English; Advanced English; Dutch)
Experiment 3 **Classroom based** Chapter 6	**Vocals** (With; Without) **Volume** (Low; High)	**Task Difficulty** (Advanced English; Dutch)
Experiment 4 **Classroom based** Chapter 7	**Vocals** (With; Without) **Volume** (Low; High)	**Task Difficulty** (Advanced English; Advanced English with pseudowords)
Experiment 5 **Laboratory based** Chapter 8	**Time Signature** (3/4; 4/4) **Tempo** (Slow; Mid; Fast)	None

Table 3.1: Summary of the experiments reported in this thesis

Explanations of how the stimulus to control each IV manipulation was created are included in the following sections, first describing the music IVs and then the presented text IVs.

Music Independent Variables

The effect of a number of different dimensions of music were investigated as IVs in the experiments reported in this thesis, including:

- presence or absence of music,
- presence or absence of vocals,
- genre (e.g. rock, classical or ambient),
- style (e.g. pop or alternative as a subset of the rock genre),
- volume,
- tempo,
- time signature (e.g. 3/4 or 4/4).

The following paragraphs describe how the music stimuli were created and controlled to ensure high levels of construct and internal validity[11]. A description of how the music stimuli were created to retain tight control of the music IV under investigation as well as how other dimensions of the music (e.g. volume in experiments where volume was not an IV) and the listener's relationship with the music (e.g. the participant's familiarity with the music) were controlled is provided. All pieces of music are included on the accompanying USB drive, see page 21 for a full listing.

As well as the music IVs shown in Table 3.1 the three laboratory based experiments (Experiments 1, 2 and 5) varied the ***presence or absence of music*** by including a without music condition. It would be inaccurate to describe this condition as silent because it was rare for the environment to be completely noiseless. Typically, there were environmental sounds that could not be removed or reduced.

The first music IV investigated in Experiment 1 was ***vocals.*** To vary the presence or absence of vocals while keeping the rest of the music the same in every way required the creation of new music mixes as unfortunately, it is not possible to simply subtract the vocal tracks from pre-mixed recordings of music. Instead, the best way to achieve two mixes of one piece of music that are the same in all aspects except for inclusion of vocals is to take a multi-track recording[12] and mix[13] into two finished pieces of music. One version then included all the vocals tracks (typically a lead singer and backing vocals) while the other was mixed without including any of the vocal tracks.

It is fairly unusual for musicians to publish the unmixed, pre-production individual tracks for a song. However, Cambridge Music Technology provides a Multitrack Download Library known as *Mixing Secrets* (Senior, n.d.) as part of a training website for aspiring audio engineers and producers. This resource encourages musicians to upload raw, pre-mixed multi-track recordings of their music so anyone can learn how to mix and remix music without having to record their own tracks which is an expensive process. The raw, unedited, unmixed individual audio tracks are uploaded to the website by musicians for use in training and education, and not for commercial gain. The quality of recordings

11 Internal validity means changes in the dependent variable can be attributed to the manipulations of the independent variable.

12 A method of recording where separate recording tracks (typically 16, or 24) are created for each instrument/combination of instruments. E.g the keyboards are recorded using one microphone. Later (or simultaneously if strongly directional microphones are used), a recording is made of the lead vocals, which is saved to a different track.

13 Audio mixing is part of the post production phase and involves taking a multi-track recording and combining the tracks into a stereo (i.e. 2 track) output.

available on the Mixing Secrets website varies dramatically, from some very amateur sounding music to more professional quality recordings.

I downloaded the pre-production unmixed multi-track versions of two songs from the Mixing Secrets resource and created mixed using GarageBand (Apple, 2010). There were no complex audio editing or engineering processes applied to the mix, for example I did not add any audio panning or reverb. The only aspect of the audio recordings where I used my own judgement to create a suitable mix was in setting the volume levels for each of the individual tracks. The recorded volume for some tracks was reduced so that the mix sounded pleasant with no particular track or instrument dominating the overall mix, e.g. by reducing the volume of drum tracks to stop them from monopolizing the song.

I selected two rock songs for use in the experiments, which also determined the ***style*** IV. The first song, *Atrophy* (The Doppler Shift, n.d.) is described as "Epic Indie Rock" within the Alt Rock resource section. In this thesis the Atrophy piece is referred to as *alt rock*. The second piece of music, *Big Dummy Shake* (Moosmusic, n.d.) is described as "Indie Pop/Rock" style on the website and referred to as *pop rock* in this thesis. The genre of both of these pieces is rock, but stylistically they are quite different.

Along with fundamental properties of the audio track, such as the presence or absence of vocals, there are other dimensions of music that could, potentially have affected the participants' performance and experience. Before moving onto describe how the music was chosen or created to control the other music IVs in Table 3.1, one of the possible confounding variables should be addressed – that of the participants' ***familiarity*** with the music.

How familiar a piece of music is may affect how the music influences someone's behaviour. Some people may know a song very well which means the music distracts them, or perhaps it means they can block it out more easily. Other people may have never heard that piece of music before. It is important to try to control how familiar the music was to the participants as it is unclear how familiarity would affect performance[14].

One approach to control familiarity is to pick obscure music that is unlikely to have been heard by any of the participants. The music available from the Mixing Secrets resource

---

14 It is also worth noting that some of the pieces of music Jensen used in his experiments are quite popular/famous compositions. For example, the "Méditation" from the Thaïs opera by Jules Massenet is a well-known piece of recognisable music, perhaps not by name but by melody. It is not clear if the music was familiar to the participants in Jensen's study, but there is certainly the prospect that they may have heard the pieces before.

meets this unfamiliarity criterion as the artists are typically not famous and their music is not well known. This approach aims to ensure the music stimuli were new to the participants, though, unless original music compositions were developed for this research this is something that I cannot be certain of. Some of the participants may have been aspiring audio engineers who were familiar with Mixing Secrets resource, so it was important to ask the participants if they had ever heard the songs before. All participants responded they had not heard either the alt or pop rock pieces before the experiment[15]. Therefore, the music was sufficiently unfamiliar to the participants that they could not recall having heard it before the experiment, though it is possible they had been exposed to the pieces previously but were unable to remember hearing it. At worst, the music was sufficiently unfamiliar to the participants that they could not recall ever having heard it before the experiment, so with either the alt rock or pop rock music familiarity is not expected to compound the experiments.

In some experiments, to manipulate the IV under investigation I did not need to create a post-production mix of a multi-track recording as fully mastered[16] recordings were suitable. To mitigate for the possible familiarity, confound these music selections came either from somewhat obscure musicians, or were less popular works from famous musicians. For example, the effect of ***genre*** was manipulated in Experiment 2 using pieces of ambient, classical and rock music as stimuli. The ambient and classical pieces were both fully mastered recordings, while the rock music was the alt rock without vocals version of *Atrophy* (The Doppler Shift, n.d.) described earlier.

The ambient music was *Continuity Part I* by Liquid Mind (2001, track 3), a somewhat obscure musician whose music is unlikely to be well-known outside of people who frequently listen to ambient music. The classical music was the *Allegro Assai* movement from Mozart's Piano Concerto No. 20 (2005, track 11), which is a movement that is not notable for being very well-known, although Mozart is obviously an infamous composer. It is again possible that the participants were keen listeners of ambient or classical music and may have known these pieces. So, after each transcription task I asked the participants if they had heard the piece of music before. If the participant responded "no" the first time that piece of music was played in the session it is clear the piece of music was

15 This experiment involved playing the same piece of music, or a version of the music that had been slightly altered, at multiple times during the experiment. Interestingly, although most participants did notice they had already been exposed to a piece of music during the experiment, many participants did not identify that they had heard the piece of music earlier.

16 Audio mastering is used in post-production to take a mix and create the final version ready for distribution.

unfamiliar to them and not a well-loved piece of music[17]. At worst, the participant may have heard the music before, but they had not recalled it, so the music can be described as unfamiliar to them. None of the participants, in any of the experiments, reported that they had heard any of the pieces of music before the experiment.

The manner used to control playback ***volume*** depended on whether the experiment was being performed using the laboratory or classroom based methodology. Volume was not manipulated as an IV in any of the laboratory experiments, but it was important to set the same volume parameter for all participants otherwise this dimension of the music could have been a confounding variable. In the laboratory experiments, the music was played using Audacity (Open Source, 2012) running on my MacBook Pro laptop connected to a pair of external laptop speakers. So, to standardise the volume between participants, the volume control on both the speakers and my laptop were set to maximum. Then, to set the music to play at a comfortable level for the participants, the volume control parameter in Audacity was set to either 50% or 75%, depending on the music used as the stimuli in the experiment.

In the two classroom based experiments volume was manipulated as an IV. The music playback was controlled entirely through a bespoke website I had developed to collect the data (for more about the website see the software section 3.3.2.2, page 90). The participants had been provided with pairs of inexpensive headphones to use in the experiment. The experimenters were told to set the volume on the computer to 100%, and the media player that was embedded in the website set the music's playback volume to either 50% or 100% depending on whether the participants had been assigned to the low or high volume group.

Music tempo was manipulated using Audacity (Open Source, 2012), which allows music to be sped up or slowed down by a specified factor, without changing the frequency response (i.e. the pitch) of the recording. This is important as typically, without the intelligent processing algorithms used by Audacity, the process of speeding up a piece of music results in an overall increase to the pitch. Audacity allows the speed of audio recordings to be manipulated while preserving the original frequency response, and

17 Again, in this experiment the participants were exposed to the same piece of music more than once, and again some of the participants did not noticed the music was repeated during the session.

therefore without changing the pitch of the audio. As a result, multiple versions of a piece of music can be created which have different tempos but are identical in all other respects.

The ***time signature*** manipulation was achieved by using music from a 3/4 and 4/4 movement from one symphony. Each movement's time signature was established by inspecting the published score of the work. As the two pieces are from a single symphony, performed by the same orchestra, with the same conductor, and from a recording made by the same producers and engineers, the style of the music, the interpretation by the musicians and conductor, and the recording are consistent between the two pieces of music. The tempo of both movements was manipulated in Audacity (Open Source, 2012) to improve similarity between the 3/4 and 4/4 pieces across as many dimensions of music as possible. This means that although the two pieces of music may be different in characteristics and dimensions other than the time signature, they are also quite similar.

If these two movements with different time signatures affected transcription typing performance and experience differently the time signature manipulation is one possible explanation of the effect. Though, as there are other differences between the movements care will need to be taken in the discussion not to attribute significant differences specifically to the time signature. Rather than saying "the time signature had this effect on transcription typing" the discussion would need to be framed as "the two movements of music had different effects on transcription typing, which may be attributed to the time signature as this was a key dimension where the two movements differed".

All of the experiments in this thesis involved 4.5 minute transcription typing tasks. Some of the pieces of music chosen were initially shorter than 4.5 minutes in duration. To make sure music was playing throughout the whole task, the pieces of music were lengthened in Audacity (Open Source, 2012) by copying the waveform of the music to the end of the original recording.

Further specifics about the music used in each experiment, are included within the materials section for that individual experiment.

Presented Text IVs

This section describes the different text passages that were used as presented text in the experiments reported in this thesis. The experiments all used some form of repeated measures design where the participants completed multiple typing tasks. So, several passages of text were needed to avoid learning and practice effects caused by the participants copying the same material repeatedly.

As well as needing multiple examples of the text, four different sources were used for the texts, which were categorised as follows:

- simple English,
- advanced English,
- advanced English with pseudowords,
- Dutch.

More information about the presented text sources, including short samples and an explanation of how they were created, is provided in the following paragraphs.

The intention behind using four different sources of text was to vary the difficulty of the typing tasks, as shown in Figure 3.5.

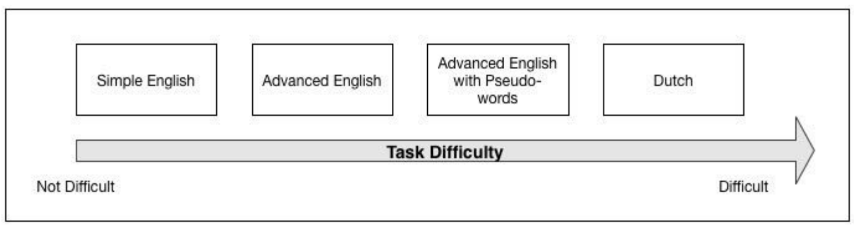


Figure 3.5: Planned impact of text type on perceived task difficulty[18]

Task difficulty was included as a text IV in Experiments 2, 3 and 4 so the presented text passages came from more than one type of text. The presented text passages used in Experiments 1 and 5 were from the advanced English source only as task difficulty was not including as a text IV in these experiments.

Two texts using English language were used in the experiments, both from novels that were no longer under any copyright restrictions. The excerpts used as presented text came from the first few paragraphs of different chapters from the novels.

The advanced English text was The Outlaw of Torn (Burroughs, 2008) which is a novel intended to be read by adults or older children (see Figure 3.6 for an excerpt). The simple English text was from The Adventures of Grandfather Frog (Burgess, 2004) which is a children's book (an example is included in Figure 3.7).

18 Although the planned difficulty was determined as in Figure 3.5. In reality, the perceived difficulty of the texts actually varied differently as the simple English and advanced English texts were not actually perceived by the participants in Experiment 2 to have different difficulties.

> Here is a story that has lain dormant for seven hundred years. At first it was suppressed by one of the Plantagenet kings of England. Later it was forgotten. I happened to dig it up by accident. The accident being the relationship of my wife's cousin to a certain Father Superior in a very ancient monastery in Europe.

Figure 3.6: Example of the advanced English text

> Longlegs the Blue Heron felt decidedly out of sorts. It was a beautiful morning, too beautiful for any one to be feeling that way. Indeed, it was the same beautiful morning in which Grandfather Frog had caught so many foolish green flies.

Figure 3.7: Example of the simple English text

A Flesch Reading score (Flesch, n.d.) was assigned to all the presented text passages written in English. A higher score on the scale indicates the text is easier to read. The mean score for the simple English passages was 78.85, which is at the top of the 7th grade reading level and described as fairly easy. The mean score for the advanced English passages was 64.25, which is at the midpoint of the 8th and 9th grade reading level.

A statistical comparison of the scores of all of the passages was performed using a Mann-Whitney *U* test. The results of this analysis are included in Appendix C. The simple English text's Flesch reading scores were significantly lower than the advanced English text scores. That is, from a reading perspective the text from the children's book was easier to read than the advanced English text. Of course, just because a text source is easier to read does not mean that it is easier to type, which is a hypothesis that is investigated in Experiment 2 (Chapter 5).

Another way to increase the difficulty of the typing task is to change the language of the transcription task to one that is foreign to the participant. Clearly, it is going to be quite difficult to achieve fast, accurate, high levels of typing performance when transcribing in a language that is not your native one. Dutch was chosen for the foreign language text for a few reasons. First, the Dutch alphabet is similar to English. The source text was modified to remove any characters with accents or unusual punctuation. For example, instances of 'ö' were changed to simply 'o'. Also, Dutch is not a language that is taught in schools in the UK, so it is unlikely that the participants would be familiar with it. However, all participants were asked if they could speak Dutch and if so they would be assigned to a different set of transcription tasks, if possible, or their data would have been discarded from the sample. Though, none of the participants in the experiments which used the Dutch language text reported they were able to speak Dutch.

The Dutch language text used in these experiments was Op Eigen Wieken (Alcott, 2007) for which a direction translation is On Your Own Wicks. The text itself is a Dutch translation of Good Wives by Louisa May Alcott (see Figure 3.8 for an example).

> Jo ging niet weer terug naar tante March, want de oude dame had zoo′n voorliefde voor Amy opgevat, dat zij haar tot blijven wist te bewegen, door het aanbod van teekenlessen van een der beste meesters; en voor zulk een onverhoopt geluk zou Amy wel een veel harder meesteres willen dienen.

Figure 3.8: Example of the Dutch text

Transcription typing in Dutch is going to be more difficult than typing in English for non-Dutch speakers, and represents a mundane, monotonous task that would require high levels of concentration to do well. However, it is possible that the task is actually too difficult, so a task was needed that fell somewhere between typing in English and typing in Dutch on the difficulty scale.

To achieve this, the material for the advanced English with pseudowords type of text was creating. Pseudowords are “a string of letters that is pronounceable and conforms to the English orthographic pattern and is pronounceable, but has no meaning” (Hughett, 2006). So, including pseudowords in a passage of English text should make the transcription tasks more difficult than typing in English but still fundamentally English-like and readable.

A source of pseudowords was identified which contains a master list of 14,000 probable English pseudowords generated using a “Markov chain method”. The paper which reports the precise method of generating the pseudowords does not seem to have been formally published anywhere, though a reference was provided by the researcher[19]. Each pseudoword had a rating from 3 to 9 indicating the “order of the Markov chain used”. The meaning here is somewhat unclear without access to the original seemingly unpublished work. However, inspection of the pseudowords themselves suggests lower rated

19 The software which generated the list of 14,000 probable pseudowords is available from Pereleman School of Medicine at the University of Pennsylvania’s Department of Psychiatry.

pseudowords are generally shorter, with fewer syllables and are perhaps easier to read. Table 3.2 presents some examples of the pseudowords with their associated rating.

Rating	Example Pseudowords		
3	exsess	tebosis	lectic
4	eskine	thredite	letable
5	espit	tilbited	lindrochemic
6	estoppeur	tinctionist	lovesomene
7	ethanolysin	tranquilnesses	lymphatisation
8	eudiaphorous	transsepulcher	leucosyenites
9	excretitiousness	tribofluoresceine	logogrammaticism

Table 3.2: Example pseudowords with rating

Two different advanced English with pseudowords passages were needed in Experiment 5, so a Python script was created which randomly selected two subsets of 500 words from the 10,000 pseudowords with a rating between 3 and 7. These two subsets formed the basis for inserting a randomly chosen pseudoword every 2 to 4 words in the first few paragraphs of two different chapters of the advanced English text (from different chapters to those used in other typing tasks). This approach to creating the texts ensured that every few words there was a pseudoword, but each pseudoword was only used once. Figure 3.9 includes an example from the advanced English with pseudowords presented text, the pseudowords have been highlighted in bold, though this was formatting was not included in the text that was presented to the participants.

> As **acclimatic** De **pressnes** Vac drew **scientious** his sword from **glossia** the heart **cuting emint** of the Lady Maud **resens**, he winced **plic**, for, **minism** merciless though **trally** he was, he **metaphy** had **crystally** shrunk from **overprop** this cruel **cani** task. Too far **unized** he had gone, **seculate** however, **intered** to **prevely** back down **prous** now, and, **plut** had he **prevely** left the **nalle** Lady Maud **sesque** alive.

Figure 3.9: Example of the advanced English with pseudowords text

3.3.1.2 Dependent Variables

The primary aim of the experiments in this thesis was to investigate the impact of different dimensions of music on performance and experience when transcription typing. Quantitative data was collected as a means to objectively measure each participant's typing performance. Experiential data was also collected, using quantitative 7-point Likert items[20] for ratings and, in the laboratory based experiments, using a qualitative open ended question to further explain why participants found particular pieces of music more distracting than others.

This section defines the DVs and explains how they were measured and calculated. The techniques for analysing the data are described in Section 3.5, page 103.

Objective Measures of Typing Performance

Typing speed and accuracy were measured as the objective typing performance DVs. Each of these DVs are explained further in the following paragraphs.

The standard measure for typing speed is words per minute (WPM) where a *word* is defined as five characters. A character is the symbol that is output in response to a single, or combination of keypresses. For example, it takes a single keypress to output the ‘c’ character and two keypresses to output the ‘C’ character. In addition to letters and numbers, punctuation is included in the character count. For example, if the user enters a ‘ ’ (i.e. by pressing the spacebar) this is also classified as a character. Similarly, if the user types the carriage return key (i.e. ‘\n’ or return) this is also counted as a character.

In these experiments, rather than reporting WPM, characters per minute (CPM) is used. The analysis is not affected by using CPM rather than WPM, but the scaling factor makes the differences in speed easier to identify visually when comparing a distribution’s descriptive statistics such as the mean. For example, if the difference in CPM is 1 character, this is easier to identify than the equivalent 0.2 words. Equation 1 shows how CPM was calculated.

$$characters\ per\ minute = \frac{number\ of\ characters\ in\ transcribed\ text}{length\ of\ typing\ task\ (in\ minutes)}$$

Equation 1: Formula to calculate characters per minute

Typing accuracy was measured by calculating the error rate percentage. The first important concept to define is what constitutes an error in the transcribed text. There are numerous approaches to defining and counting errors in a string of text, so it is somewhat unfortunate that Jensen does not provide a clear definition of what constitutes an error. In my experiments, the Levenshtein Distance (LSD) algorithm was used to establish the

20 The term “Likert scale” is often used incorrectly to describe Likert items. In these experiments, a single question was used to establish perceived task difficulty, so this is a single Likert item. If more than four Likert items were grouped to get a score for a particular concept, then it would be correct to use the term Likert Scale. But, in the experiments reported in this thesis Likert items were used rather than a consolidated Likert Scale which has implications for the analysis procedure, described in section 3.5.3, page 40.

number of errors in the transcribed text. The LSD algorithm compares the transcribed text to the presented text and establishes the Levenshtein Minimum String Distance (MSD), which is, in essence, the number of moves required to transform one string into another using additions, deletions and substitutions. Figure 3.10 shows examples of transcribed text that all contain one error as counted using the Levenshtein MSD:

Addition Error	Presented
	Transcribed
	Error
		`The quick brown fox jumped over the lazy dog`
		`Th quick brown fox jumped over the lazy dog`
		`  ^`

Deletion Error	Presented
	Transcribed
	Error
		`The quick brown fox jumped over the lazy dog`
		`They quick brown fox jumped over the lazy dog`
		`   ^`

Substitution Error	Presented
	Transcribed
	Error
		`The quick brown fox jumped over the lazy dog`
		`Tha quick brown fox jumped over the lazy dog`
		`  ^`

Figure 3.10: Erroneous transcription examples

Using the Levenshtein MSD as a means of counting errors in typing is a recognised technique used in numerous research papers, (e.g. in both Kane, Wobbrock, Harniss, & Johnson, 2008; and Morimoto & Amir, 2010).

For the experiment designs I used, the participants typed for 4.5 minutes in each condition. Therefore, the total number of characters entered by each participant depended upon the speed of their typing. So, simply using error count as a cumulative metric is not appropriate – if someone enters 2 incorrect characters in a 100 character output, this is much less severe than someone entering 2 incorrect characters in a 5 character output. Instead, an error rate measure is a more appropriate method as the number of errors is considered as a percentage of the total number of characters. Equation 2 provides the error rate formula used in this thesis.

$$error\ rate = \frac{number\ of\ errors\ in\ transcribed\ text}{number\ of\ characters\ in\ transcribed\ text} \times 100\%$$

Equation 2: Formula to calculate error rate percentage

Subjective Measures of Typing Experience

Typing experience was measured using 7-point Likert items relating to the perceived task difficulty (all experiment conditions) and perceived music distraction (only experiment conditions with music). In the laboratory experiments, an open ended question asking the participants to describe why they found a piece of music distracting was also included after each transcription task with music.

The first question after each typing task always related to the participant's perception of how difficult the task was, using the question shown in Figure 3.11.

Please rate on a scale from 1-7 how easy you found it to complete typing task?

Extremely difficult			Neither difficult nor easy			Extremely Easy
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

Figure 3.11: Task difficulty Likert item

If they had just completed a typing task with music the participants were also asked the question in Figure 3.12 to establish if they perceived the music to be distracting.

Please rate on a scale from 1-7 how distracting you found the music?

Extremely distracting						Not at all distracting
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

Figure 3.12: Music distraction Likert item

3.3.2 EQUIPMENT AND MATERIALS

The equipment and materials used for each experiment depended somewhat on the methodology adopted. The following sections describe the experiment environments, the software and hardware and the questionnaires used in the experiments, focusing on areas of similarity.

3.3.2.1 Environment

All experiments took place in the Department of Computer Science at the University of York. Experiments performed in the laboratory took place in the Usability Lab while the classroom based experiments took place in one of the software labs.

3.3.2.2 Software

Custom built webpages were developed for the experiments. The webpages were created using HTML, CSS, JavaScript and PHP to control the interface and automatically collect and save the data on a server. The websites were optimised for the FireFox browser.

The experiments all contained multiple transcription tasks using different text passages. Each typing task (i.e. each text passage) had its own webpage, which loaded the presented text from a .txt file stored on the server into a non-editable text area field on the left hand side of the screen. The right side of the webpage contained an editable text area form field which the participant used to enter their transcription. The spellcheck attribute for the editable text field was set to false to prevent the webpage from highlighting errors in the participants' input text. The size of both the presented text and the input text areas were identical, with both using 16 point Arial font.

Figure 3.13 shows an example typing task interface used in the laboratory based experiments with one of the advanced English transcription tasks loaded into the left hand panel. An example of the interface used in the classroom based experiments is included in Appendix B.3. At the end of the typing task, the experimenter enters the participant's ID and clicks on the submit button which saves the transcribed text output as a .txt file on the server with a unique filename using the participant's ID number.

In the laboratory setting, the experiments used a corpus of individual, unlinked web pages. As I was running the laboratory experiments, I controlled loading each webpage and made sure the participants completed the correct transcription typing tasks in the right order to achieve the required counterbalancing. These webpages did not control playback of the music nor time the length of each transcription task as I regulated those aspects of the procedure. More details on the laboratory procedure are provided in Section 3.3.3.1, page 93.

In the classroom based experiments, the websites were more complex because they needed to include some important interface setup instructions, as well as controlling the music playback, monitoring the task length, and guiding the experimenter to choose the correct transcription typing task. So, instead of individual unlinked webpages, a larger website was created comprising a number of linked pages. Appendix B.3 contains screenshots from all pages of the classroom based website.

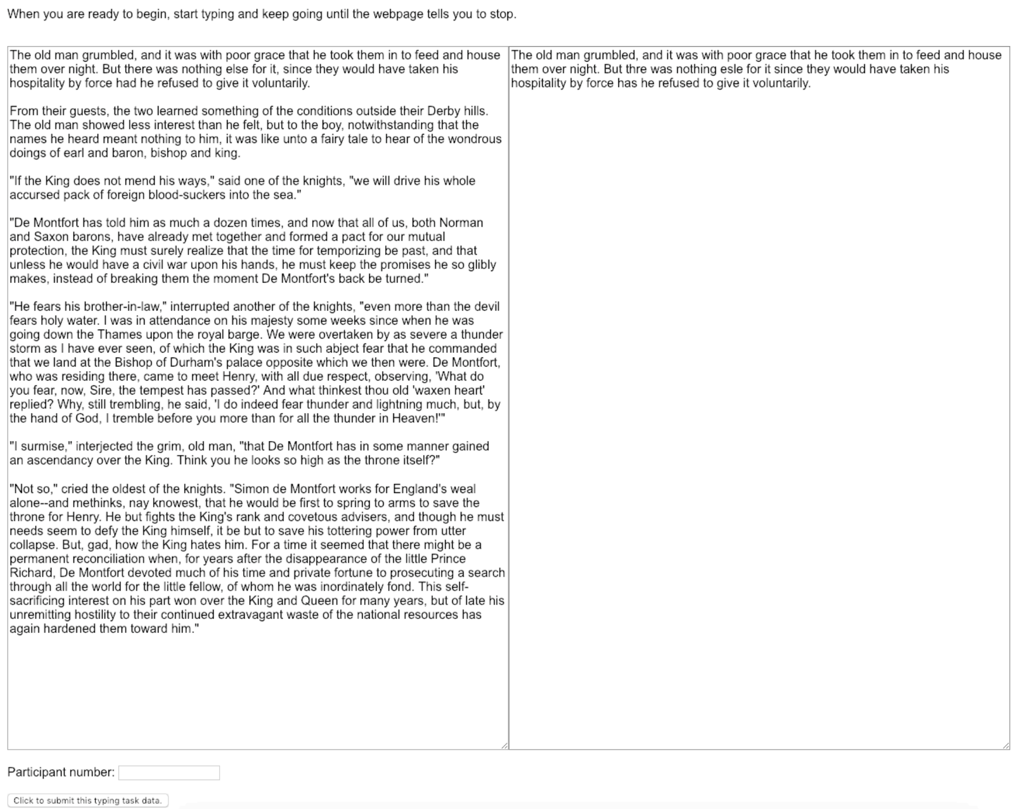


Figure 3.13: Example typing task interface used in laboratory based experiments

The first page of the website used in the classroom based experiments included instructions to setup the browser and make sure all of the students in the class performed the experiment using the same interface. These instructions included ensuring the screen was maximised, the volume of the computer had been set to maximum and that the FireFox functionality which sends the browser back a page when the backspace key is pressed outside of a text box had been disabled. Figure 3.14 shows a screenshot of the checkboxes used to ensure the setup had been performed correctly. The full setup instructions can be found in Appendix B.3.1. The red outline was used to highlight the importance of this section of the instructions. All fields were required, so the form could not be submitted, and the experiment could not be started unless all checkboxes had been marked. The participant ID field included some JavaScript client side processing which prevented the form submitting unless this field contained a maximum of two characters which had to be numbers.

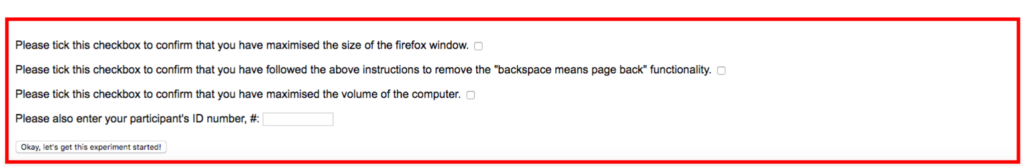


Figure 3.14: Setup check boxes for classroom based methodology

After successfully submitting this form a practice task page was loaded into the browser. The individual transcription typing task webpages for the practice and all real tasks were visually the same as those used in the laboratory experiments. However, the back end of the task webpages had more functionality. The webpages used in the classroom experiments had two important pieces of functionality triggered when the participant entered their first keypress in the editable text area, 1) the music automatically began playing and 2) a timer started. After 4.5 minutes (or 20 seconds if a practice task), a JavaScript alert box then popped up on the screen telling the participant they had finished that task and the music was stopped. Then, the experimenter entered the participant ID to submit the participant's data.

After data was submitted from any task, a page loaded that contained some options to guide the experimenter in choosing the next transcription typing task. Figure 3.15 contains a screenshot of one of the options pages.

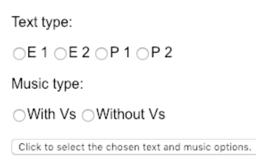


Figure 3.15: Example task selection page used in the classroom based methodology

Radio buttons were used to ensure the experimenter only selected one option and the form could not be submitted without a selection of both a text and music type. The text type varied depending on the experiment, but this example includes E1 and E2 which refer to the two English texts and P1 and P2 which were the two English with pseudowords passages. Both classroom experiments included a vocals manipulation, so the music type option presented was always With Vs and Without Vs which refers to the with and without vocals condition, without explicitly indicating to the participant that the music manipulation was the presence or absence of vocals.

After the participants had completed their final transcription task, a thank you page was displayed which reminded the experimenter to provide the participant with the demographics questionnaire to complete.

3.3.2.3 Hardware

The hardware used in the experiments again depended upon the methodology used in the experiment. The computers used by the participants in these experiments were all desktop machines running Windows (either Windows XP or Windows 7), with a separate keyboard and mouse. The laboratory and classroom based experiments used identical keyboards with an English UK layout.

In the laboratory based experiments the music was played through Philips SPA 2210 2.0 laptop speakers connected to my MacBook Pro. The participants in the classroom based experiments were all given pairs of inexpensive Astra Tools ATA 1144 headphones to use.

3.3.2.4 Questionnaires and Forms

After each typing task, the participants completed short paper based questionnaires (examples in Appendix A.2, A.3 and B.4). All participants also completed an informed consent form (examples in Appendix A.1 and B.1) and questionnaires regarding demographics and about their music and computing habits (see Appendix A.4 and B.5).

In the classroom based experiments, the students acting as experimenters were given step by step instructions for how to perform the experiment. An example of the experimenter instructions document is included in Appendix B.2.

3.3.3 Procedure

The experiment procedure is again similar across all the experiments, but there were some important difference due to the two methodologies. So, for clarity the procedure described in this section presents both methodologies separately.

3.3.3.1 Laboratory Procedure

The procedure used in the laboratory experiments is described in the following sections. The procedure is separated into four phases, the initial setup, taking informed consent, performing the typing tasks and concluding the experiment.

Initial Setup

The initial setup phase involved preparing the participant's computer which was used to collect the data and my MacBook Pro which was used to play the music.

On my MacBook Pro the setup involved loading all the music files into Audacity (Open Source, 2012) and setting the volume in Audacity to either 50% or 75% depending on the experiment. The speakers were then connected to the MacBook Pro but positioned either side of the monitor of the participant's desktop computer. The volume on the speakers and the MacBook Pro were both set to 100%.

The setup of the participant's computer just involved preparing Morae to record the session.

Taking Informed Consent

The experiment session began with me explaining to the participant that I am investigating the impact of different types of music on performance and experience when transcription typing. I told the participants that they were going to be asked to complete a number of typing tasks with and without different pieces of music playing. I told the participants what was being recorded during the session. I explained to the participants that if they wanted to withdraw from the experiment at any point they could do so without prejudice, that all data would be treated anonymously and the only people who would see the data in its raw format were myself and my supervisors. Finally, I explained they would receive a £10 Amazon or Marks and Spencer's gift voucher. Assuming the participants were happy, I asked them to sign an informed consent form (an example form is provided in Appendix A.1).

Performing the Typing Tasks

The typing tasks were all hosted online on a bespoke website created especially for these experiments. Each experiment was administered using an interface similar to that shown in Figure 3.13 previously.

I asked the participants to make themselves comfortable by adjusting the chair, position of the keyboard, etc. Before the first task, I told the participants they should type as naturally as possible without prioritising either speed or accuracy and that whatever approach they used, they should aim to be consistent across all of the tasks. I also asked the participants not to use the mouse at any point.

The first typing task in the experiment was a 20 second practice typing task with music. I loaded the practice task into FireFox and prepared one of the music selections. I asked the participant if they were ready and told them to start typing when they heard music. I started playback of the music from my MacBook Pro and started a timer when the participant made their first keypress. After 20 seconds, I asked the participant to stop typing and stopped the music playing. I told the participants that in a real task, they would now be asked to complete a short questionnaire but as this was just the practice it was not required. I scrolled the webpage down and entered the task description (e.g. in this case, I entered *practice* and the participant's ID number) and then clicked on the submit button to send the data to the server.

I asked the participants if the volume of the music was okay (i.e. that it was not too loud). Assuming the volume of the music was acceptable, the first of the real typing tasks was loaded into the web browser. I explained that the real experiments would be longer than 20 seconds, but they would not be typing for long enough to copy all of the presented text.

I opened the first music selection in Audacity on my MacBook Pro (if their first task was a music condition) and told the participant to start typing when they heard the music, or to just start when they were ready if it was a without music condition. Again, I started playback of the music and began a timer. After typing for 4.5 minutes, I asked the participant to stop typing and stopped the music playing. I gave the participant a paper based questionnaire about the task to complete, see Appendix A.2 and A.3 for example forms. While they were filling in the questionnaire I submitted their data using the website with the task name and participant ID number and prepared the next typing task.

This procedure described in the previous paragraph was repeated for all conditions in the experiment.

Concluding the Experiments

After all of the typing tasks were completed, the participants completed two questionnaires, one to collect demographic data and another about their use of music in their everyday activities (examples of these questionnaires are included in Appendix A.3 and A.4). After completing the questionnaires, I debriefed the participants by explaining the experiment hypotheses and asked if they had any questions. Finally, the participants were thanked for their time and asked to initial the areas of the informed consent form that confirmed they had been adequately debriefed.

3.3.3.2 Classroom Procedure

The classroom based procedure had six phases, beginning the practical class, taking informed consent, the initial set up, performing the typing tasks, concluding the experiment and debriefing the class. These phases are described in the following sections.

Beginning the Practical Class

When the students entered the classroom the module leader introduced the session and explained that the students were going to be involved in two experiments during the session, one as an experimenter and one as a participant. The module leader introduced the researchers who had designed the two experiments to the class. The students were asked to arrange themselves into pairs and decide who was going to be the participant in the first experiment, and who would be participant in the second.

My experiment was the first to be performed in each session. Once the students had organised themselves into pairs and decided who would be the participant in the first experiment I explained to the class that my PhD research focuses on investigating the effect of music on transcription typing and that the participants would complete a number of typing tasks with different pieces of music. I explained that, at the end of the session, I would debrief the whole class to explain the hypotheses and objectives of the experiment.

I went onto explain to the class that I would be giving the experimenters packs of paperwork in plastic sleeves which contained precise instructions for how to run the experiment in an ethical manner and which should ensure the data collected was valid. I emphasised that it was very important the experimenter followed the exact procedure, so they should read through the full set of instructions before starting the experiment. I highlighted the importance of making sure the computer's volume was set to 100% and that, if this was too loud for the participant, it could be changed but that they needed to inform me they had reduced the volume.

I explained that if the experimenter did not understand any step they should raise their hand and ask for help from myself, the module leader or one of the session teaching assistants, who had all been thoroughly briefed on the correct procedure. Finally, I explained that Amazon vouchers would be awarded to the three best typists and five participants would also be randomly selected to win a voucher. I stressed that participation was entirely voluntary, and they could withdraw at any point without prejudice, but if they did wish to withdraw they should inform someone. I also explained to the participants that if they did not want to be a participant in the typing experiment, they could swap roles with their experimenter.

Pairs of headphones, experimenter instructions and data collection packs containing the questionnaires were then distributed to the students. The experimenters were told to begin reading the instructions in the pack as soon as they received them.

Taking Informed Consent

The first step in the experimenter instructions involved taking informed consent. Figure 3.16 includes the excerpt from the instructions including the script that had to be read out to the participants.

> First, you need to briefly explain what the experiment is about to the participant and thank them for their time. So, read them the following blurb:
>
> "Thank you for taking part in this experiment, which is part of Anna BD's PhD studies, where she is looking at how background music affects users in different situations. This particular experiment is looking at the impact of background music on copy typing. You will be asked to complete a few typing tasks while listening to different pieces of music.
>
> Your participation is entirely voluntary, you can withdraw at any point during the experiment and your data will be destroyed. 5 participants will be randomly selected to win £10 Amazon voucher, while the top 3 scores will win £30, £20 and £10 Amazon vouchers respectively. Are you happy to continue?"
>
> If they say "yes" give the participant, the Informed Consent form to read and sign. If they say "no" ask a demonstrator to come over.

Figure 3.16: Excerpt from experimenter pack regarding informed consent

The experimenter read the script to the participant, who was then given the informed consent form to sign. An example informed consent form is provided in Appendix B.1.

Initial Setup

The experimenter then had to set up the computer. This involved plugging the headphones into the audio out port, setting the volume to maximum, loading the website into the browser and updating some browser settings. The experimenter had to check a number of boxes to agree that they had completed the setup process correctly before the experiment could begin. The full experimenter setup instructions web form is included in Appendix B.3.

Performing the Typing Tasks

After the experimenter submitted the form verifying they had setup the computer correctly, the website loaded a practice typing task. The instructions pack told the experimenter to inform the participant that:

- they must not use the mouse to move the cursor,
- they should try to type as fast as they can but should correct errors that the notice,
- the music would start playing as soon as they make their first keypress into the text box,
- the first task is a practice to ensure they both understand the process and the volume level is not too loud.

The experimenter then asked the participant to start typing, the website began a timer and playing music on the first keypress. After 20 seconds, the website generated a JavaScript alert box to stop the task.

After clicking on the okay button to dismiss the alert box, the experimenter entered the participants ID number and submitted their transcribed text. The experimenter explained that in the real typing tasks, they would now complete a short questionnaire about their experience of typing in that task. The experimenter was instructed to ask the participant if the volume of the music was okay, if the participant found it was too loud the experimenter was instructed to ask for help from one of the teaching assistants in the class. The teaching assistants had been told it was okay to reduce the volume of the computer if necessary, but to record the ID number of any participant who had requested a change of volume, so their data could be discarded. None of the participants requested to lower the volume of music in either of the two classroom based experiments.

The website then loaded a page which contained options for the experimenter to select the next passage of text for the transcription typing task and the music version that was to be played. The website then loaded the appropriate webpage containing the correct music choice and text passage and asked the participant to begin the typing task when they were ready. Again, the music began playing automatically, but in the real tasks the timer was set to 4.5 minutes. At the end of the task, a JavaScript alert popped up to stop the task. The experimenter asked the participant to complete a single page questionnaire containing Likert items and the experimenter submitted the transcribed text through the website.

The process described in the previous paragraph was then repeated until all experiment conditions had been completed.

Concluding the Experiment

After the participants had completed all the transcription typing tasks they were given a demographics questionnaire to complete on paper. The experimenter entered the participant's written data into a spreadsheet so that they could get experience of coding data and emailed the data to me[21].

Debriefing the Class

When the participants had completed all the tasks and questionnaires I debriefed the whole class together by explaining the objectives of the experiment as well as describing and justifying the independent and dependent variables.

3.4 The Classroom Methodology

The two different experiment methodologies that were used in my research have already been introduced briefly. The classroom based methodology used in two experiments reported in this thesis requires further explanation as it is a non-standard approach for performing experiments with some important ethical implications that need to be addressed.

As part of their undergraduate degree, students in the Computer Science department at the University of York take a module called *Human Aspects of Computer Science* where they are taught quantitative empirical methods, including the concept of experiment validity and statistical analysis methods. During one of their practical classes the students take part in two experiments, by acting as a participant in one and as experimenter in another. The experiments used in the classes are real rather than contrived experiments, including two of the experiments that are reported in this thesis.

21 Although the experimenters sent me their transcription of the participant's data, for analysis I transcribed all the data myself to ensure it was coded correctly.

This approach to teaching experimental method and analysis material has pedagogic value for the students as it provides them with experience of actually being a participant in an authentic research project as well as control in running a real experiment to collect genuine data. From a researcher's perspective, experiments can be performed with a large number of participants over a small amount of time. However, a number of important ethical considerations needed to be considered carefully before collecting the data in this manner. There were also some important implications for the design and delivery of these two experiments as I had to give over most of the control for running the experiment either to the students or to the website used to collect the data.

3.4.1 Ethical Considerations

The informed consent form used in the classroom experiments was an extension of the standard form used in laboratory based experiments. Additional aspects included a Frequently Asked Questions (FAQs) section which addressed some of the subtleties and complexities regarding running experiments in an ethical manner when it forms part of their education in an assessed module. An example of the informed consent form used in the classroom based experiments is included in Appendix B.1.

It was vitally important that informed consent was carefully managed in these classroom based experiments, beyond ensuring the consent form was clear and included a FAQs section. The experimenters were provided with a pack of materials that provided step by step instructions for exactly how to run the experiment. These instructions also included a script which the experimenter had to read out verbatim to the participant to take informed consent (see the Procedure described in section 3.3.3.2).

Aside from informed consent, there are further ethical considerations to be taken when performing experiments using a classroom based methodology which forms part of the students' education experience. Although the experiment itself is not assessed, the experiment is being used as a learning mechanism within a practical class as part of an assessed module which has some important ethical implications. Firstly, the students' learning is more important than the research outcomes and results, from an ethical perspective. Secondly, at the time of running the experiments in the Department of Computer Science, at the University of York, students' attendance at practical classes was compulsory. So, it was important to emphasise to the students that although their attendance at the class was compulsory, it was not a requirement for them to act as a participant in this (or any) of the experiments taking place in the class if they did not want to. However, as it is an important pedagogic exercise that forms part of the Intended Learning Outcomes (ILOs) for the module, the students were advised they should take part in one of the experiments on offer during the session as their education experience would be heightened and more valuable as they are taking part in real research.

If a student was adamant that they did not want to be a participant in the typing experiment, then they could request to be a participant for the other experiment or sit out the experiment activities entirely. However, this did not happen in either of the two experiments which are reported in this thesis.

3.4.2 The Pilot Process

Before running the first classroom based experiment a thorough pilot process was performed with students working in pairs. The students were given drafts of the experimenter instructions and asked to identify any issues with clarity and any aspect that was unclear. I observed the process carefully to detect any mistakes being made by the student acting as the experimenter. The data from the students in the pilot studies was recorded and checked for problems but discarded without analysis.

Three pairs of students took part in the pilot process using different versions of the website and instructions. These students were second and third year undergraduates in Computer Science and would not be taking part in the experiment. One student acted as experimenter, with the other as the participant. I told the experimenter to ask me questions and clarify meaning wherever they felt they needed to. After they had followed the procedure, the participant and experimenter explained any areas where they felt the instructions or website could be improved. I then analysed the file names and the submitted text carefully to see if there were any potential problems with how the data was being stored on the server.

During the piloting process a number of issues were identified with the instructions and website which led to changes to the bespoke websites that had been created for these experiments. The next sections provide some examples of the problems that were encountered and how they were resolved to ensure that the final experiment ran smoothly and in such a way that the data can be considered safe. After the third pilot study I felt that the final website and experimenter instructions had a good balance between ensuring the experiment runs smoothly and not providing the experimenter with too much information so as to overwhelm them.

3.4.2.1 Problems Caused by Alert Notifications

The website was developed so that each transcription typing task included a timer function which began when the participant entered their first keypress on that webpage. After 4.5 minutes, the website generated a JavaScript alert to stop the typing task. This was found to be the easiest means of ensuring the typing task only lasted for 4.5 minutes. If the students had been trusted to time the experiment themselves and stop the participants at the right time, there was clear potential for inconsistency in the duration of teach task. However, a problem was introduced when the alert box was generated at the same moment the participant was beginning a new paragraph. If this occurred the participant would hit the 'Enter' key twice to add space between the paragraphs and this double Enter press action would dismiss the alert message and automatically submit the form. This initially led to the transcribed text being saved on the server with a file name that did not include the participant ID and, as a result was not associated with a particular participant. To overcome this problem the form field for the participant ID was given a required attribute which prevented a form from being submitted unless a participant ID had been entered.

3.4.2.2 Incorrect Entry of Participant ID

By making the participant ID a required field it ensured that all data was submitted and stored on the server with an associated participant ID. However, this did not eliminate all problems as there were some examples identified in the pilot task where an experimenter entered the wrong participant ID or entered the ID in an inconsistent format, e.g. by entering "P1" rather than just "1". At best these small irregularities by the experimenter led to inconsistencies in the file names which required additional work to interpret in order to run analysis Python scripts on all the transcribed text files as a single batch of data. At worst, incorrect participant IDs could have led to the data overwriting another participant's file. To alleviate this potential problem, the participant ID that was entered by each experimenter was taken only from the first page of the website when the experimenter verified they had set up the computer according to the instructions. The participant ID number was stored as a PHP session variable from this page which could then be used by the other web pages to send the transcription data to the server.

This approach reduced the requirement to have a separate form field for the participant ID on each page, and in one pilot test this field was removed so the participant clicked on okay in response to the alert pop up and the data was immediately submitted, and the next page was loaded. This approach however led to some confusion with the participants, especially if they had been looking at the keyboard rather than the screen and not noticed the alert. So, the participant ID field was kept as a required field on each webpage, but the contents entered by the experimenter on the individual task pages was not actually used.

3.4.2.3 Ambiguous or Unclear Instructions

The instructions included the order which each participant should complete the transcription typing tasks in which was important as it included some complex counterbalancing to alleviate fatigue, practice and learning effects due to the repeated measures design. The original instructions included the table in Figure 3.17, a. The experimenter should have been aware that *With* and *Without* referred to vocals if they had thoroughly read and carefully processed the aims of the experiment, but in the pilot one of the experimenters thought this instruction meant *With Music* and *Without Music*. To overcome this potential source of confusion, the instructions were updated to specify *With Vs* and *Without Vs* (Figure 3.17, b).

Order	Music	Text
1st	With	English 1
2nd	Without	English 2
3rd	With	Dutch 1
4th	Without	Dutch 2

(a)

Order	Music	Text
1st	With Vs	English 1
2nd	Without Vs	English 2
3rd	With Vs	Dutch 1
4th	Without Vs	Dutch 2

(b)

Figure 3.17: Task order

This change provides sufficient information to clarify the condition was with and without vocals but was still sufficiently ambiguous for the participant to be unaware that the manipulation in the experiment was vocals. If the instructions had included the terms vocals, vocs or vox, the participant may have been primed to the precise experiment manipulation, but by using Vs the experimenter is clear what condition is needed without explicitly informing the participant.

3.5 Quantitative Analysis Procedures

This section formalises the procedures used to analyse the quantitative data collected in the experiments, particularly in terms of ensuring assumption requirements for the statistical tests were met. The design of all of the experiments in this thesis called for analysis using either repeated measures or mixed analysis of variance (ANOVA).

An alpha level of 0.050 was used in all statistical tests, apart from post hoc tests where a Bonferroni adjustment was used to compensate for possible over testing. The results sections of each experiment describe results where p is less than 0.1 as "trends towards significance" on the understanding that these non-significant results may have been significant if the experiments had higher power, so are potentially interesting. Though, care has been taken not to overstate any non-significant results.

Before explaining the analysis procedures, some terminology used in the results sections of each chapter needs to be clarified. When referring to the *result* of an analysis or statistical test I am referring to the statistic calculated by the software package, SPSS. In referring to the *outcome* or *outcomes* from an experiment analysis I am referring to the interpretation that can be drawn from the result. For example, $p < 0.0005$ is a result, but the outcome is a highly significant effect. Alternatively, $p = 0.070$ is a result, with a non-significant outcome indicating a trend towards significance.

3.5.1 ANOVA Assumptions

The standard ANOVA computes an F statistic based on the following assumptions about the distribution:

- the DV is a continuous variable,
- the sampling distribution mean is normally distributed on each level of the IV,
- the DV is free from outliers on each level of the within participant factors.

The experiments in this thesis typically use a repeated measures design meaning an additional assumption is required of an:

- independence of errors.

Some of the analyses have mixed designs with one or more between groups factors. The mixed design ANOVA also has the assumption of:

- homogeneity of variance across the different groups.

To be sure that the resultant *F* statistic is accurate, all assumptions appropriate to the design, must be achieved. In some circumstances, the outcome from an ANOVA can be reasoned to be safe in spite of seemingly meaningful assumption violations. It is typically better to understand how the outcomes from the analysis react to assumption violations, rather than completely dismissing the analysis method each time an assumption is violated. The key question to ask is not “does the data violate the assumptions for ANOVA?” but instead “what is the effect of violating this assumption in this case?” (Glass, Peckham, & Sanders, 1972). Taking this approach to the analysis allows me to use ANOVAs in situations where assumptions are violated in some manner which is particularly important when there is no suitable alternative analysis procedure as there is no equivalent non-parametric test that can accommodate the design.

The following sections take each of the ANOVA assumptions above and presents the procedure used to establish conformance, as well as explaining the implications of violations.

3.5.1.1 The DV is a Continuous Variable

The two DVs that were measured to establish typing performance, CPM and error rate percentage, meet the assumption that the DV is a continuous variable. However, the subjective ratings for typing experience were obtained using Likert items not Likert scales resulting in discrete ordinal data rather than the continuous interval data that is required for an ANOVA to be an appropriate test. So, ANOVA is not suitable with the subjective ratings of task difficulty and music distraction. The analysis procedure used with the Likert item data is explained in Section 3.5.2, page 110.

3.5.1.2 Normality of Distribution of Sampling Means

To verify the normality of distribution of sampling means, the data needs to be visualised and tested. If the distributions of the original data are normal, then it follows that the distributions of the sampling means will also conform to normality (Howell, 2008, p. 303-304). So, it is important to verify the normality of the original data distributions. A few different methods were used to assess normality of distribution, which are described in the following paragraphs.

The first method was to plot histograms of each level of a within participants factor and perform a visual inspection of the plot[22]. Shapiro-Wilk's tests were performed on the typing performance DVs as well. The results from all Shapiro-Wilk's tests are included in the Appendices of each experiment. The Shapiro-Wilk's test assesses the normality of distribution by rejecting the null hypothesis when the distribution does not adhere to normality. Therefore, a result from this test with $p > 0.05$ allows me to confidently describe the distribution of data, and therefore also of the sampling means, as normal.

For distributions that were significantly non-normal according to the Shapiro-Wilk's test, the skewness[23] and kurtosis[24] of the sample was inspected. Figures 3.18 and 3.19 provide a visual representation of the guidelines for describing skewness and kurtosis parameters of a sample based on information provided in Bulmer (1979) and Tabachnick and Fidell (2007).

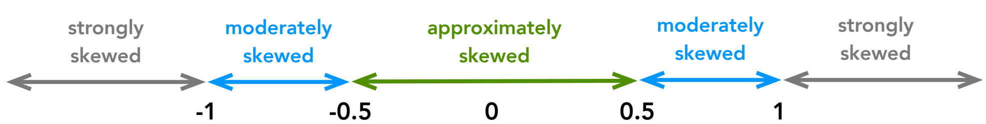


Figure 3.18: Skewness descriptors

22 A possible bimodal distribution of CPM was identified in Experiment 1, with two overlapping normal distributions rather than a single distribution. This led to classification of the participants as slow and fast typists based on a threshold level determined by inspection of histograms and scatterplots of the data. In essence, this classification added a between participants' typing speed quasi-IV to the design of the experiment. The typing speed classification turned out to be an important factor in the analyses. So, typing speed classification, as determined by inspection of each participant's CPM, was used in the experiments whenever suitable. The results sections in each experiment describe in detail how the threshold level for assigning participants to the slow or fast typists group was determined.

23 Skewness values describe the symmetry of the distribution. A normal distribution is fully symmetric with skewness of zero.

24 Kurtosis values describe the peakedness of the distribution. A distribution with a strong peak and thin tails is described as leptokurtic. A distribution that is flat with thick tails is described as platykurtic.

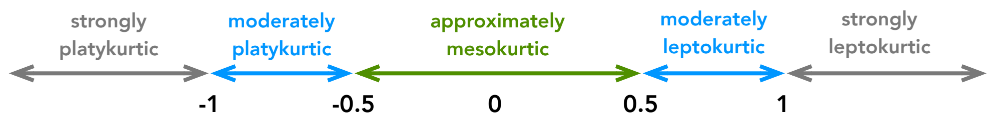


Figure 3.19: Kurtosis descriptors

However, these values of skewness and kurtosis only describe the data distribution from the sample, rather than the anticipated skewness and kurtosis of the wider population. *Z* transformations of skewness and kurtosis were calculated so the values are representative of the population and not just the sample.

Equations 3 and 4 were used to calculate z_s for skewness and z_k for kurtosis. An $\alpha = 0.01$ level is sufficient for normality which translates to $-2.58 < z < 2.58$ for both skewness and kurtosis (Tabachnick & Fidell, 2007, p. 46).

$$z_s = \frac{skewness}{standard\ error\ for\ skewness}$$

Equation 3: Skewness z-score equation

$$z_k = \frac{kurtosis}{standard\ error\ for\ kurtosis}$$

Equation 4: Kurtosis z-score equation

If the sample size is large enough then the Central Limit Theorem ensures that even if the raw values are not normally distributed, the sampling distribution of means will retain the characteristic of normality (Tabachnick & Fidell, 2007, p. 87). A violation of skewness has minimal effect on the resultant *F* statistic, but leptokurtic distributions with small samples can inflate the *F* statistic and so can platykurtic distributions (Glass et al., 1972, p. 273). If the *z* score for kurtosis indicates either a platykurtic distribution, or a leptokurtic distribution with a small sample size then there is a high probability that the sampling distribution of means is not normal. This type of violation leads to an increase to the ANOVA's power meaning the *F* statistic becomes unsafe and ANOVA is an unwise analysis method to use. In this instance, it is better to either transform the data (Section 3.5.2, page 110, describes my approach to data transformation) or, if available, to use a non-parametric alternative to the ANOVA test.

In the experiments reported in this thesis, the sample sizes can be relatively modest, so it is probably insufficient to rely on the Central Limit Theorem and accept distributions if they indicate non-normality. Instead, the general rule is that, assuming equal sample sizes ANOVAs are robust to non-normality of the sampling distribution provided that the degrees of freedom for error are greater than 20 (Tabachnick & Fidell, 2007) which aligns with the view presented by Glass (1972, p. 273). In cases where the degrees of freedom for error were less than 20, the *z* score for kurtosis were inspected. If the *z* score indicated a leptokurtic distribution, a transformation was applied, otherwise the analysis proceeded.

3.5.1.3 The Data is Free from Outliers

Outliers are data points that are unusually extreme, either much higher or lower than the majority of the data. When outliers are identified in the data, they need to be considered carefully as their impact on the *F* statistic can be substantial, particularly with small sample sizes. The mean of the data can be considered unsafe if there are outliers as it either becomes too high or too low to be representative of the population. If the mean value is incorrect, this can lead to either statistically significant outcomes that are incorrect because the *F* statistic is inflated (a Type I error), or a non-significant outcome when, without the outlying data point, the analysis would have resulted in significance (a Type II error caused by an *F* statistic that is lower than it would have been without the outlier in the data).

Visual inspection of the histograms of each distribution usually provides a good indication of potential outliers as these data points are typically somewhat disconnected from the rest of the data. A heuristic to determine if a data point is an outlier that is widely used, is to identify if the point is more than 1.5 times the interquartile range above the third quartile, or below the first quartile.

Removing outliers from the data without good justification is poor practice. In other words, unless there is a good reason to believe that the data point is not a real part of the population it should be retained in the data. Although the statistical analysis becomes more robust if outliers are excluded, the outlier itself is often actually a genuine data point. Excluding outliers may, on the face of it, improve the analysis as the outcome seems more secure. However, removing that data point has an effect on the generalisability of the outcomes. It is almost as if one is saying "we got this outcome, but only when this particular extreme case is not included".

In some of the analyses reported in this thesis, there are at least one outlying value in the data. The procedure for dealing with outliers that I used was to perform the analysis both with and without the data from the participants identified as outliers in at least one condition included and assess the impact of their inclusion on the outcomes of the analysis. If the outliers were shown to have no effect on the analysis outcomes, the reported analysis included the original data from all of the participants. If, however, the analysis outcomes were affected by inclusion of the outlier participants (i.e. the result was a significant *p* value when the outliers were included, but not significant when they were excluded and vice versa) a winzorizing (otherwise known as or outlier-capping), approach was taken where the extreme values are replaced with the highest (or lowest) non-outlier value (Everitt & Howell, 2005). By capping the data in this way, the participants' data is retained in the analysis, but the outliers do not distort the outcomes from the statistical test. The general principle I used was to try to retain outlier participants within the data unless there was a strong reason to remove them.

3.5.1.4 Independence of Errors

The assumption of independence of errors required in a repeated measures ANOVA can be conceptualised as needing the variances of the differences between all combinations of the within participants factor to be equal. This concept is known as sphericity and was analysed using Mauchly's test of sphericity. Where violations to sphericity were observed, a Greenhouse-Geisser correction was applied (Howell, 2008, p. 490-491).

3.5.1.5 Homogeneity of Variance

When a between groups factor is included in the experiment design there is a further requirement of homogeneity of variance across the different groups of each factor. Levene's test was used to assess the sample for homogeneity of variance.

ANOVA is robust to violation of the assumption of homogeneity of variance if the sample size is large and the group sizes are similar (Tabachnick & Fidell, 2007, p. 88). However, the sample sizes the experiments reported in this thesis can be somewhat modest, and sometimes unequal due to the speed classification described earlier, so it is insufficient to simply ignore violations of the assumption of homogeneity of variance as they may affect the *F* statistic, and the outcomes from the analysis.

The procedure used following violations of homogeneity of variance (otherwise known as heteroscedasticity) depended on the seriousness of the violation. If the heteroscedasticity is only moderately severe[25], the *F* statistic calculated by the ANOVA is only slightly affected as long as there was an equal number of samples in each group (i.e. the design was balanced). But, if there is moderate heterogeneity of variance and the sample sizes are unbalanced then the impact on the *F* statistic is more serious. If the larger variance occurs in the groups with the larger sample size the chance of a Type I error is smaller than the nominal level. In this situation, while the heterogeneity of variance should be acknowledged, the resultant *F* statistic is fairly robust. On the other hand, if the larger variance is observed in the groups with smaller sample sizes, the ANOVA is too liberal a test and the results should be subject to scrutiny as the *F* statistic is likely to be unsafe. In this situation, the chances of a Type I error, i.e. incorrectly identifying a significant outcome when one should not have been observed, are increased (Tabachnick & Fidell, 2007, p. 88).

3.5.2 Data Transformation

Distributions of continuous data which failed to conform to normality were transformed. Likert item data was not transformed, instead the analysis procedure described in Section 3.5.2 (page 110) was followed.

Typically, the error rate data had a strong positive skew with many participants achieving low error rates and a few with high error rates, so this data needed to be transformed before inferential analysis using ANOVAs. A square root transformation was sufficient to improve normality Experiment 1, but in the other experiments a logarithmic transformation was needed.

A logarithmic transformation cannot be applied if the data incudes a value of zero, because $\log_{10}(0) = -\infty$, which is an extreme outlier. Some participants achieved no errors, so using Equation 2 their error rate would be zero and transformed error rate would be $-\infty$. To overcome this problem the equation to calculate the original error rate was modified when a logarithmic transformation was needed so that a value of zero error rate could not be obtained. Instead of using Equation 2 to calculate the error rate percentage, the number of

25 Assessed by examining the p value obtained from Levene's test. If $p < 0.030$ then the assumption violation is considered to be substantial.

errors and characters were first increased by 1, as shown in Equation 5. This approach ensures consistency across all participants, while also preventing any zero error rates, meaning a logarithmic transformation can be applied safely.

$$error\ rate = \frac{number\ of\ errors\ in\ transcribed\ text + 1}{number\ of\ characters\ in\ transcribed\ text + 1} \times\ 100\%$$

Equation 5: Modified formula to calculate error rate percentage

3.5.3 Analysing Likert Item Data

Likert items need to be considered carefully before inferential statistical analysis to ensure the results are valid. Typically, single Likert items (rather than the summative Likert scales) violate the assumptions required for ANOVA to be considered a suitable analysis method. The Likert item response format collects ordinal rather than interval type data meaning that the distribution of responses is discrete rather than continuous. Also, the distribution of data collected using a single Likert item is unlikely to be normally distributed as the question will often elicit a response that either clusters around the centre value, or one or both of the extremes. If the data distribution is non-normal, the sampling distribution of means will also be non-normal. Although ANOVA is robust to non-normality of the sampling distribution of means, the combination of this assumption violation plus the non-continuous nature of the responses leads to some concern about the appropriateness of the ANOVA procedure with the Likert item data collected in these experiments.

Given these concerns regarding acceptance of the underlying assumptions for an ANOVA to be appropriate for the analysis of Likert items, another procedure for performing the analysis of the quantitative, subjective data was needed.

For repeated measures designs, the Friedman test is a suitable non-parametric alternative test to the one-way ANOVA which can be used with ordinal rather than interval data. Unfortunately, the design of my experiments tended to include multi-level repeated measures or mixed designs for which there is no alternative non-parametric test available. So, the approach I took to analyse Likert item data in all the experiments was to begin by performing either a repeated measures or mixed ANOVA despite the assumption violations. Then, if significant effects or interactions were obtained, a non-parametric test was performed. For example, if a 2-level between groups factor had a significant effect on the task difficulty rating in a mixed ANOVA, two separate Friedman or Wilcoxon signed rank tests were performed. Alternatively, if the 2-level between groups factor was non-significant, then the data from the two groups was combined and analysed together as if it had been collected from a single group.

Only significant effects, or those that showed a trend towards significance (i.e. $p < 0.1$) from an ANOVA were followed up and interpretations were only made based on the results from non-parametric tests. This approach conservatively deals with the problem of analysing Likert items when the experiment design can only be accommodated by parametric tests.

3.5.4 Post Hoc Tests

The repeated measures and mixed ANOVAs reported in this thesis led to numerous significant two-way and three-way interactions. These interactions were broken down using a simple effects analysis, by identifying simple interactions and simple main effects based on the most appropriate conceptualisation indicated by the experiment hypotheses.

For IVs with more than two levels, post hoc pairwise comparisons were performed to identify the cause of significant omnibus and main effects, simple or otherwise. Some of the post hoc pairwise comparisons were significant at the Bonferroni adjusted level, while others were significant only at the non-adjusted level. Pairwise comparisons with $p < 0.1$ are reported in the thesis, but stronger value is placed on comparisons which were significant at the Bonferroni adjusted level.

3.6 Qualitative Analysis Procedure

In the laboratory based experiments the participants were asked an open ended question to explain why they found particular pieces of music distracting. These comments were transcribed, and a lightweight content analysis procedure followed. Categories emerged through analysis of the comments, and the frequency with which comments under each category were reported by participants was recorded. The results sections of these experiments include summary tables which contain the comment type category, the frequency of occurrence and example quotes. These qualitative comments were used to gather further insights into whether the music IV manipulation was recognised as being the cause of distraction, or if some other dimension of the music was the perceived cause. The qualitative analysis was used to guide the aims and objectives of the next experiments.

Chapter 4

VOCALS AND MUSIC STYLE

Experiment 1

The aim of this first experiment was to investigate how the presence of vocals in music affects transcription typing performance and experience, as well as establishing if the method would be suitable for future experiments. The key paper which forms the foundation and starting point for my research investigated how music affects the performance of skilled typists by comparing speed and error count in without music, jazz and dirge music conditions (Jensen, 1931). The results from Jensen's experiment showed that the jazz and dirge music had different effects on typing performance with the jazz music accompaniment leading to an increase in errors, and the dirge music reducing typing speed. It is possible that the tempo of the music was an important dimension that caused the effects observed through Jensen's experiment, but this is conjecture as there is no evidence to support this claim. While the tempo of the jazz and dirge music selections used did differ greatly, with the jazz music faster than the dirge, the differences in effect on performance were not attributed specifically to tempo, instead they were explained through the broader genre dimension. Instead of hypothesizing that music from genre A has a different effect on performance to music from genre B which is again different to a without music condition, this first experiment reported in this thesis isolated the presence of vocals as a dimension of music to understand how and why vocals in music may affect typing performance. The experiment also looks at whether any effects caused by vocals are the same across music from different styles, though from within the same wider genre of rock music.

Previous research in the typing literature looked at the effect of concurrent verbal stimuli on transcription typing performance. For example, a number of experiments were performed with a single, trained typist, investigating how different presentation methods for verbal stimuli affect transcription typing performance (Shaffer, 1975). In these experiments, the typist was able to maintain high levels of performance when transcribing material that was presented aurally through headphones in one ear, while simultaneously repeating words aloud that were heard through the other ear. The typist also performed well when copying text presented visually whilst reciting nursery rhymes out loud. These results suggest that, for this skilled typist, typing performance was maintained in the presence of verbal stimuli meaning that it did not conflict with the transcription typing task. But do these results extend to being able to perform transcription typing while hearing music containing vocals? Would less skilled typists be affected in a similar way? And, given the experiment was performed with a single typist, would other trained typists be able to maintain performance while exposed to conflicting verbal stimuli?

The typing literature reviewed indicates the performance of skilled typists was not affected by the presence of conflicting verbal stimuli, but there is a strong logical argument which can be constructed that contradicts the evidence. I have not taken the evidence from Shaffer's study as absolute because the study reports experiments with a single, skilled participant and I have been unable to find any follow up work with a larger sample, or less experienced typists. Instead, I argue that as transcription typing requires someone to visually attend to a verbal source by including an aurally presented verbal stimulus their cognitive load will be increased and in turn transcription typing performance will be impaired. Therefore, I hypothesize that transcription typing while listening to music containing vocals will be more difficult than with instrumental music as the vocals will be distracting. As a result, I expect that typing performance will be reduced when exposed to music that contains vocal tracks.

The primary aim of the research in this thesis is to improve performance with music in work related computing tasks by better understanding how music affects the user. However, the hypothesis behind this experiment is that vocals are a dimension that will reduce typing performance. Therefore, this particular experiment may seem to contradict the main objectives of this research. However, by understanding if particular dimensions of music have negative effects, I can begin to identify which dimensions should be incorporated within an ideal music accompaniment to improve typing performance, and hence performance at mundane work related computing tasks.

This experiment also compares the effect of style on typing performance and experience by looking at the impact of two different pieces of music from different styles from within the broader rock genre. One piece was in an alternative rock (alt rock) style, with another in a more pop rock style. The two pieces of music used within the experiment are quite different across a number of dimensions, but one important dimension in which they differ is that of tempo. If the style of music is shown to have an overall effect on typing performance and experience, it might be that the tempo of the music is a key dimension in causing this effect, which can be investigated later.

4.1 Method

4.1.1 Aims

The aims of the experiment were as follows:

1. to investigate whether transcription typing when accompanied by music affects typing performance and experience,
2. to investigate whether the presence of vocals in a piece of music reduces typing performance and experience,
3. to establish if any effects occur across two different styles of music,
4. to confirm the method is suitable for use in future experiments.

4.1.2 Participants

The participants were chosen through an opportunity sample, recruited using email lists, adverts on university boards, and via word of mouth. Twenty-eight participants (22 male, 6 female), aged between 18 and 44, took part in the experiment. The majority of the participants (53%) were in the 18 - 24 age group. All of the participants were native English speakers who did not have a known hearing problem and were not dyslexic.

All the participants were involved in academia at the University of York. Fifteen of the participants were PhD Students (11 from Computer Science and 4 from Social Science subjects) and 11 were undergraduate students from the Department of Computer Science. Two of the participants were professional researchers from Humanities departments. All participants were compensated with a £10 Amazon or Marks and Spencer's voucher.

4.1.3 Design

The experiment used a fully factorial repeated measures design with each participant completing all five experiment conditions by completing transcription typing tasks:

- without music,
- with alt rock music that contained vocals,
- with alt rock music that did not contain vocals,
- with pop rock music that contained vocals,
- with pop rock music that did not contain vocals.

Two different experiment design paradigms were used. The first paradigm focuses on the impact of the pieces of music as a whole by comparing a without music condition to the four music conditions, meaning there were five levels of the music independent variable (IV). Figure 4.1 shows the arrangement of IVs for experiment paradigm 1.

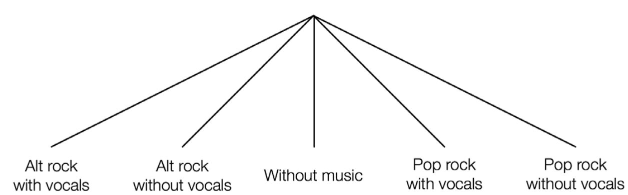


Figure 4.1: Structure of IVs for experiment paradigm 1

The second experiment design paradigm compares the effect of vocals (2 levels: with and without vocals), and the style of music (2 levels: alt rock or pop rock). Figure 4.2 shows the arrangement of IVs for experiment paradigm 2.

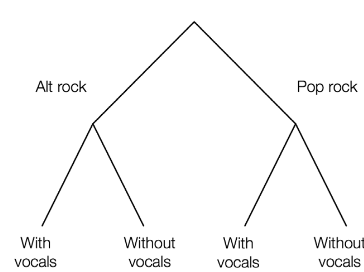


Figure 4.2: Structure of IVs for experiment paradigm 2

To avoid fatigue and practice effects the order of presentation of tasks was counterbalanced.

The dependent variables (DVs) measured in this experiment were typing performance in characters per minute (CPM), and error rate percentage and typing experience ratings of perceived task difficulty and music distraction (see Chapter 3, Section 3.3.1.2 for more on how the DVs were measured).

4.1.4 Equipment and Materials

The experiment was performed using the laboratory based methodology and took place in a quiet usability lab in the Department of Computer Science at the University of York.

The participants used a desktop PC running Windows XP with FireFox 10. The music was controlled using Audacity (Open Source, 2012) running on my MacBook Pro, which was connected to a pair of Philips SPA 2210 2.0 laptop speakers.

The alt rock music was *Atrophy* (The Doppler Shift, n.d.). *Big Dummy Shake* (Moosmusic, n.d.) was the pop rock style music. Both pieces of music were mixed down into two different versions, one with vocals and the other without vocals, as described in (see the Music Independent Variables section, in Chapter 3, Section 3.3.1.1).

In this experiment, all transcription tasks used passages from the advanced English text (see the Presented Text IVs section in Chapter 3, Section 3.3.1.1).

The participants completed short questionnaires to collect subjective ratings of task difficulty, music distraction and how much the participant liked the music (see Appendix A.2 and A.3). The questionnaires also included an open ended question asking why the music was distracting, and a question to establish if they had heard that particular piece of music before. The participants completed a demographics questionnaire. All questionnaires were completed on paper and are included in Appendix A.4.

4.1.5 PROCEDURE

The procedure followed exactly that presented in Chapter 3, Section 3.3.3.1. The participants completed five typing tasks, which lasted 4.5 minutes each, meaning they spent 22.5 minutes typing during the experiment.

4.2 RESULTS

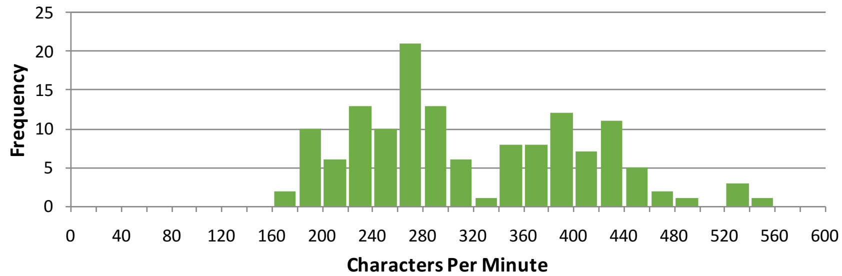
The results section begins with an analysis of the objective measures of typing speed and accuracy. Then, an analysis of the subjective metrics relating to typing experience is presented. The results section ends with a discussion of the qualitative comments made by participants explaining why they found particular pieces of music distracting.

4.2.1 TYPING PERFORMANCE

Typing speed was measured using CPM and typing accuracy by calculating the error rate percentage. The data was explored by carefully investigating the descriptive statistics before inferential analysis of the impact of the various IVs on typing performance is presented.

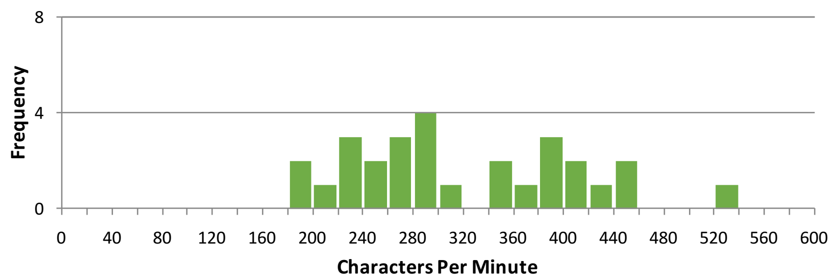
4.2.1.1 Characters Per Minute

The first DV investigated in this experiment was typing speed, measured in CPM. A consolidated histogram of typing speed across all five experiment conditions was created (Figure 4.3) as well as histograms for each of the transcription typing tasks separately (Figures 4.4 to 4.6). All the histograms are presented with the number of data points (*N* for the total sample or *n* if a subset of the total sample), the mean (*M*) and the standard deviation (*SD*).



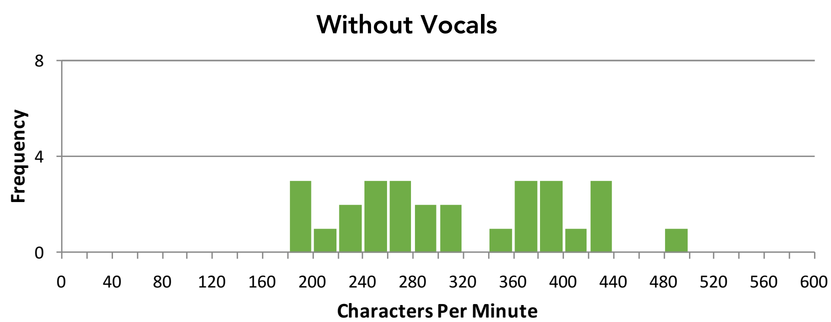
$N = 140$, $M = 317.67$, $SD = 91.62$

Figure 4.3: Histogram of CPM for all conditions

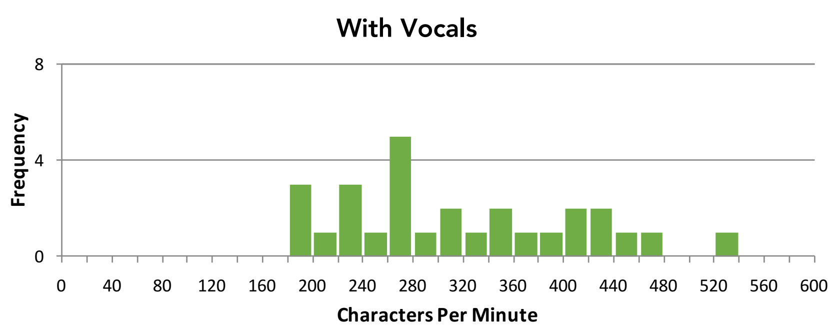


$n = 28$, $M = 321.13$, $SD = 90.43$

Figure 4.4: Histogram of CPM for without vocals condition

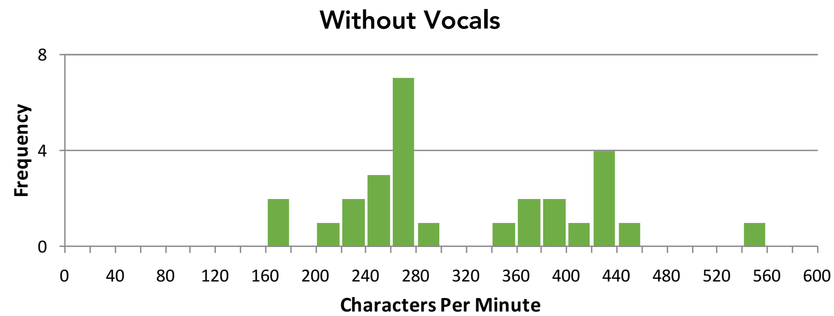


(a) $n = 28$, $M = 316.36$, $SD = 85.68$

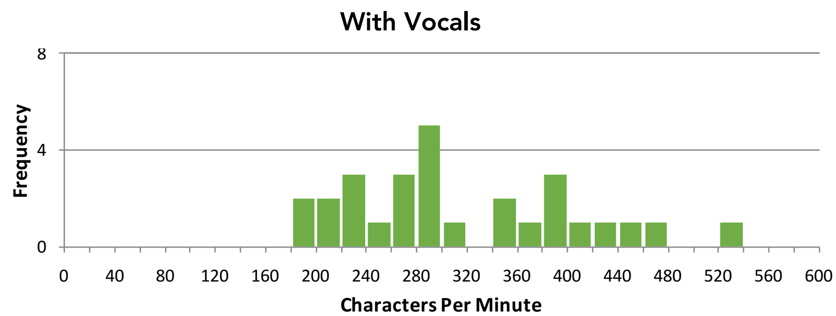


(b) $n = 28$, $M = 318.46$, $SD = 93.72$

Figure 4.5: Histograms of CPM for alt rock music



(a) n = 28, M = 319.09, SD = 96.78

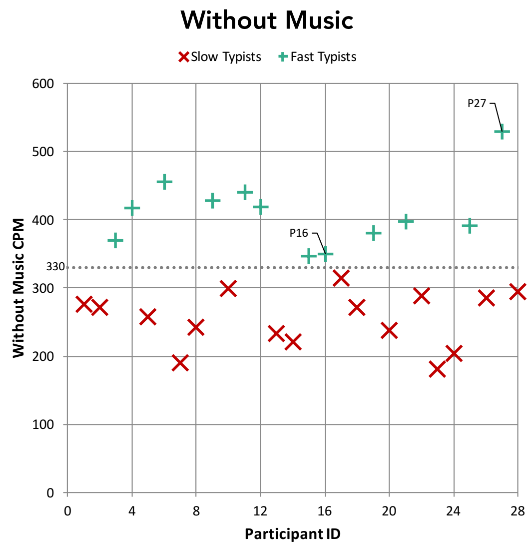


(b) n = 28, M = 313.35, SD = 91.48

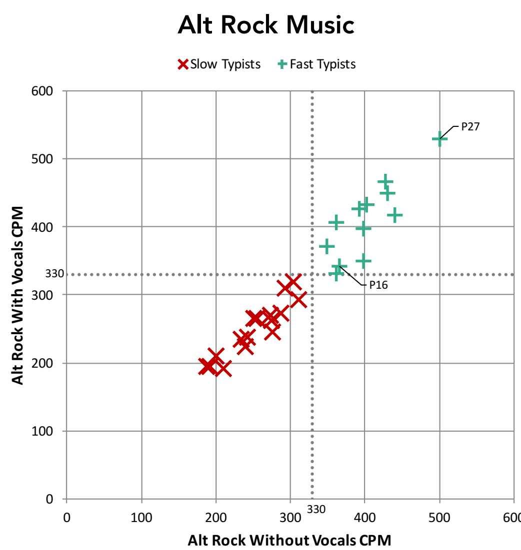
Figure 4.6: Histograms of CPM for pop rock music

Visual inspection of the histograms exposes some potentially interesting phenomena in the distributions. First, the histograms appear to contain an outlier at the top of each CPM distribution, with a single data point in excess of 480 CPM in all of the experiment conditions. The second phenomenon to emerge is the suggestion that rather than a single normal distribution of CPM, it might be better to approximate the distribution as bimodal, comprising of two normal distributions, one representing slower typists and the other for faster typists. There is a fairly clear drop between 320 and 340 CPM in all the histograms, so it may be appropriate to separate the distribution into two different speed groups for the analysis.

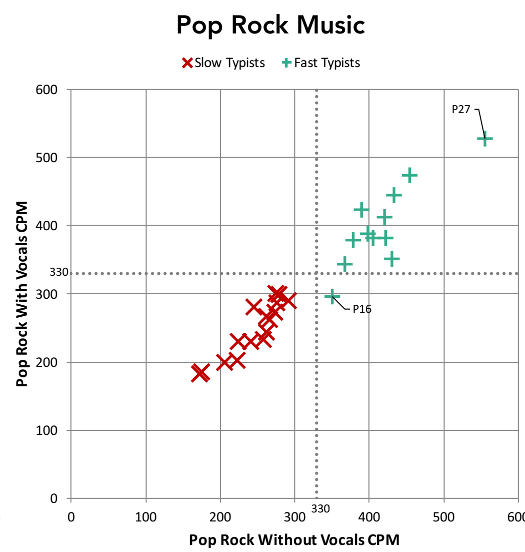
Although the histograms include a gap between 320 and 340 CPM, it does not necessarily follow that it is the same participants always achieving speeds above or below the gap in all of the conditions. A preliminary classification structure was applied to the sample, using 330 CPM as the threshold to determine if a participant should be assigned to the slow or fast typists group. To check the suitability of the categorisation of participants using a threshold of 330 CPM scatterplots including this classification structure were created (Figure 4.7). The CPM scatterplots highlight the two different speed groups and the 330 CPM threshold.



(a) Plotted against Participant ID



(b)



(c)

Figure 4.7: Scatterplots of CPM

In the without music condition (Figure 4.7, a), there is a clear gap between 310 and 340 CPM. The classification structure holds through the alt rock scatterplot (Figure 4.7, b). There is another clear gap in the without vocals condition between 310 and 340 CPM (horizontal axis) and a gap in the with vocals condition between 320 and 330 CPM (vertical axis), but it is considerably smaller as one participant achieves almost exactly 330 CPM. With pop rock music the gap between slow and fast typists is retained in the without music condition (Figure 4.7, c – horizontal axis). But, participant 16 has been highlighted as they drop below the threshold of 330 CPM in the with vocals condition (vertical axis) by some margin, achieving only 294.67 CPM. Despite achieving a lower speed in one of the transcription tasks P16 was classified as a fast typist, because they reached speeds in excess of 330 CPM in the four other tasks. Participant 27 is also highlighted as the possible outlier, confirming it was a single participant who achieved speeds in excess of 480 CPM in all of the transcription tasks.

I previously suggested any effect due to a verbal stimulus may differ according to the typists' skill, though this experiment had not intended to investigate that hypothesis. As there is evidence of a bimodal distribution which holds when scrutinising the CPM achieved by each participant, the inferential analysis will treat the data as a single group and also separated into two groups based on the classification used in Figure 4.7. It is possible that any effect due to the music may differ according to the typist's skill level, so performing the analysis using both approaches may result in some interesting and contrasting results and should be pursued even though it was not part of the original aims.

Box plots of the data were generated (see Appendix D.1) and show that with the data treated as a single group P27 was not an outlier. When the data was split into the two speed groups, P27 became an outlier in the without music and pop rock without vocals conditions. The inferential analysis using the speed classification was performed with and without the inclusion of P27 in the data. This participant's data did not affect the outcomes from the analysis so they were retained in the data, as per the procedure described in Chapter 3, Section 3.5.1.3. Table 4.1 shows descriptive statistics for all the typists together and with separation into the fast and slow groups.

Music Condition		**All Typists** (N = 28)		**Slow Typists** (n = 16)		**Fast Typists** (n = 12)	
		M (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)
Without Music		321.13	90.43	254.10	40.43	410.50	50.95
Alt Rock	Without Vocals	316.36	85.68	251.99	39.60	402.19	42.63
	With Vocals	318.46	93.72	249.53	40.07	410.37	56.93
Pop Rock	Without Vocals	319.09	96.78	245.47	36.28	417.24	52.96
	With Vocals	313.35	91.48	247.96	40.20	400.54	61.86

Table 4.1: Descriptive statistics of CPM

Shapiro-Wilk's tests were performed to assess the normality of the individual distributions before the inferential analysis. Appendix D.1.1 contains the results from these tests. None of the CPM distributions deviated significantly from normality, either with or without speed classification applied as a between groups factor.

For the inferential analysis of CPM, the sample was first treated as if it was from a single group, by performing a one way repeated measures ANOVA using the 1 by 5 experiment paradigm. Then, a mixed ANOVA was performed with the speed classification included as a between participants factor.

All of the underlying assumptions for repeated measures ANOVA were met when the data was treated as a single group, including that of normality of distribution. But, when the speed classification was included as a between groups factor an outlier was introduced. However, including the data from the outlier had no effect on the outcomes of the analysis. All other assumptions for repeated measures and mixed ANOVAs were met. Table 4.2 presents the results from the inferential analysis of CPM using experiment paradigm 1 both with and without inclusion of the speed classification.

Design	Test	F	df	p	η_p^2
1 x 5	Music	0.944	4, 108	n.s.	-
1 x 5 x 2*	Music	1.195	4, 104	n.s.	-
	Music x Speed*	1.986	4, 104	n.s.	-
	Speed*	90.420	1, 26	**< 0.0005**	0.777

Table 4.2: Inferential analysis of CPM
(* between groups factor)

There were no significant omnibus effects or interactions when the data was analysed as a single group or with the speed classification applied.

Speed classification was a significant between groups factor, $F(1, 26) = 90.420$, $p < 0.0005$, $\eta_p^2 = 0.777$. The slow typists' CPM was lower than the fast typists' (slow: M = 249.81, SD = 39.32; fast: M = 408.17, SD = 53.07).

CPM was then analysed using a 2 by 2 repeated measures ANOVA and a mixed ANOVA with the speed classification applied. Table 4.3 presents the results from these analyses.

When the data was treated as a single group there were no significant effects of interactions. With the speed classification applied, neither style, $F(1, 26) = 0.058$, *n.s.*, nor the presence or absence of vocals, $F(1, 26) = 0.322$, *n.s.*, had significant omnibus effects on CPM.

Again, speed classification was a significant between groups factor, $F(1, 26) = 90.278$, $p < 0.000$, $\eta_p^2 = 0.776$. The slow typists' CPM was lower than the fast typists' (slow: $M = 248.74$, $SD = 39.04$; fast: $M = 407.58$, $SD = 53.60$).

Design	Test	F	df	p	η_p^2	Figure
2 x 2	Style	0.164	1, 27	n.s.	-	-
	Vocals	0.247	1, 27	n.s.	-	-
	Style x Vocals	0.188	1, 27	n.s.	-	-
2 x 2 x 2*	Style	0.058	1, 26	n.s.	-	-
	Vocals	0.322	1, 26	n.s.	-	-
	Style x Vocals	3.831	1, 26	0.061**	0.128	4.80
	Style x Speed*	1.265	1, 26	n.s.	-	-
	Vocals x Speed*	0.326	1, 26	n.s.	-	-
	Style x Vocals x Speed*	8.571	1, 26	**0.007**	0.248	4.90
	Speed*	90.278	1, 26	**< 0.0005**	0.776	-

Table 4.3: Inferential analysis of CPM
(* between groups factor, ** non-significant result where $p < 0.1$)

The two-way interaction between style and vocals approached significance, $F(1, 26) = 3.831$, $p = 0.061$, $\eta_p^2 = 0.128$, (Figure 4.8).

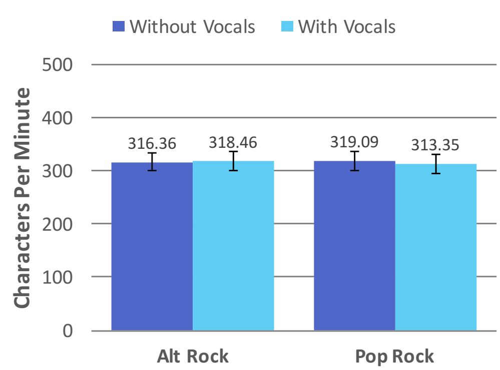


Figure 4.8: Trend towards significant interaction between style and vocals ($p = 0.061$)

The simple main effect of vocals was not significant for either style (alt rock: $F(1, 26) = 0.424$, *n.s.*; pop rock: $F(1, 26) = 2.332$, *n.s.*).

The three-way interaction between style, vocals and speed group was significant, $F(1, 26) = 8.571$, $p = 0.007$, $\eta_p^2 = 0.248$, (Figure 4.9).

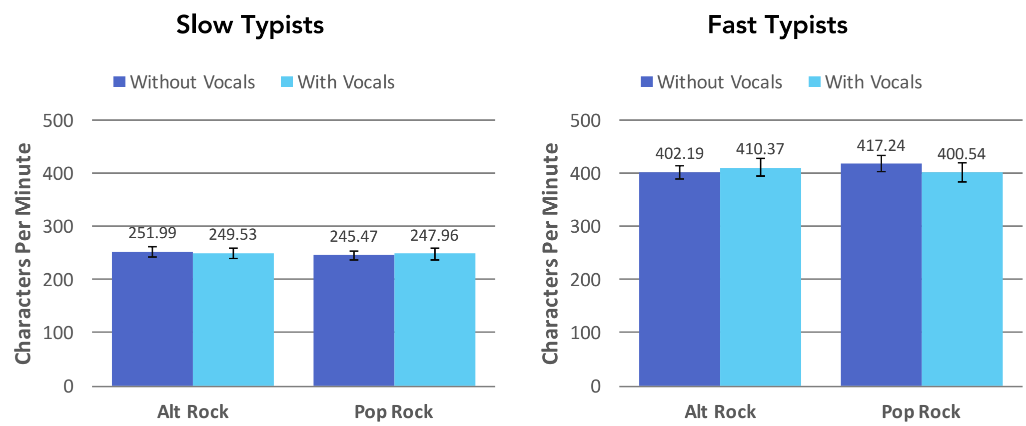


Figure 4.9: Significant interaction between style, vocals and speed group (p = 0.007)

For the simple effects analysis, this interaction can be conceptualised as the two-way interaction between style and vocals being different between the two speed groups. The simple two-way interaction between style and vocals was significant for fast typists, $F(1, 11) = 7.504$, $p = 0.019$, $\eta_p^2 = 0.406$, but not for slow typists, $F(1, 15) = 0.77$, *n.s.* The effect size of the interaction for fast typists is large with 40.6% of the variance attributed to the interaction between vocals and style.

For fast typists, the simple simple main effect of vocals was not significant for the alt rock music, $F(1, 11) = 0.83$, *n.s.*, but showed a trend towards significance for the pop rock style with a large effect size, $F(1, 11) = 3.27$, $p = 0.098$, $\eta_p^2 = 0.229$. With the pop rock music, CPM without vocals was higher than with vocals (with vocals: $M = 400.54$, $SD = 61.86$; without vocals: $M = 417.24$, $SD = 52.96$).

Again, for the fast typists, the simple simple main effect of style approached significance without vocals, $F(1, 11) = 1.904$, $p = 0.054$, $\eta_p^2 = 0.148$, but not with vocals, $F(1, 11) = 1.904$, *n.s.* In the without vocals conditions, CPM was higher with pop rock music and lower with alt rock style (pop rock: $M = 417.24$, $SD = 52.96$; alt rock: $M = 402.19$, $SD = 42.63$).

4.2.1.2 Error Rate

Typing accuracy was established by calculating the error rate percentage using Equation 2, Chapter 3, Section 3.3.1.2. A consolidated histogram of error rate percentage across all five experiment conditions was created (Figure 4.10).

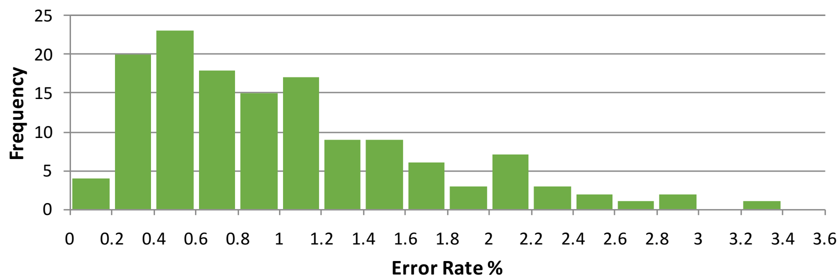
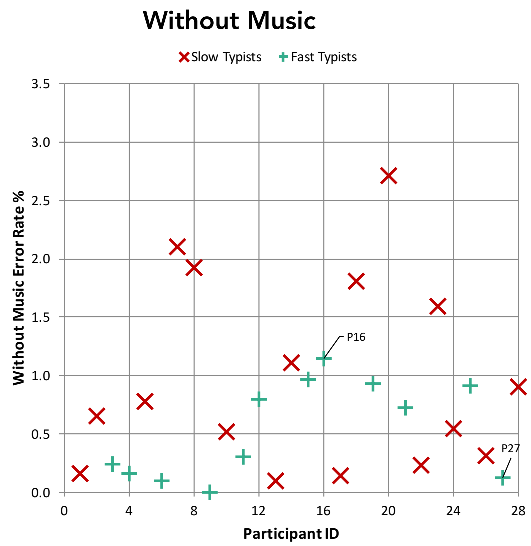


Figure 4.10: Histogram of error rate percentage for all conditions
N = 140, *M* = 0.828, *SD* = 0.672

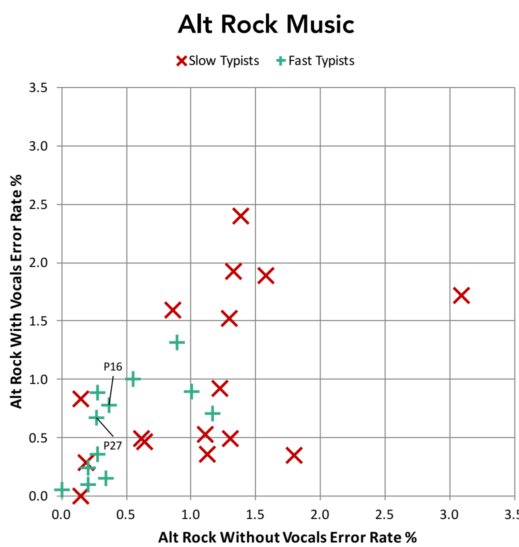
Visual inspection of the histogram shows that the error rate percentage data is positively skewed, with a long tail. This distribution shape is quite typical for error data, as generally lots of people make a few errors, while only a few people make lots of errors.

Analysis of the data for normality using Shapiro-Wilk's tests showed that all five of the transcription task conditions deviated significantly from a normal distribution (see Appendix D.1.2). Four of the tasks had strong violations of the normality of distribution assumption (p = 0.005), whilst the alt rock with vocals condition had a moderate violation of normality (p = 0.029).

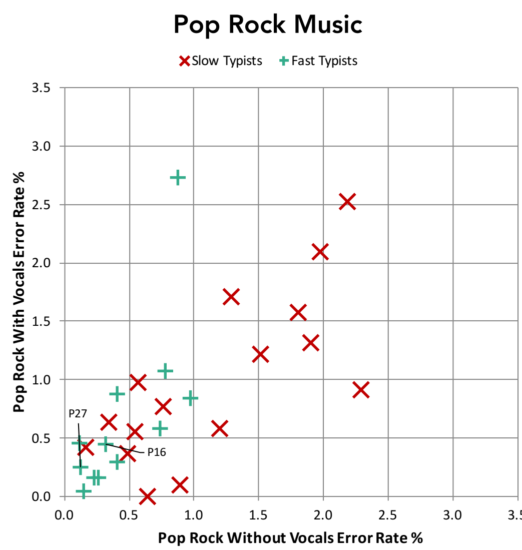
It is also evident from the histogram that the bimodality of distribution identified in typing speed data does not persist through to error rate percentage measure. Figure 4.11 shows scatterplots of the error rate percentage including participant classification based on speed group. These scatterplots show the fast typists' error rate percentages are typically clustered around lower values while the range of error rate percentages achieved by the slow typists is larger. Some slow typists' error rates were comparatively high, with other slow typists achieving error rates similar to those achieved by the fast typists.



(a) Plotted against Participant ID



(b)



(c)

Figure 4.11: Scatterplots of error rate percentage

As the error rate percentage data had a strong positive skew, it is not really suitable for analysis using ANOVA. A square root transformation was applied to the error rate percentage data which improved the normality of the distribution, as shown by the histogram in Figure 4.12. The results from Shapiro-Wilk's tests of normality were all non-significant with the square root transformation applied, both when the data was considered as single group.

Figure 4.12: Histogram of transformed error rate percentage for all conditions
N = 140, *M* = 0.8277, *SD* = 0.380

Box plots were created to assess for outliers (included in Appendix D.1.3). When the transformed data was treated as a single group there were no outliers. With the speed classification applied outliers were introduced. Both outliers had higher transformed error rates than the rest of their group. The outliers did not affect the outcomes from the inferential analysis so were retained in the data. Table 4.4 shows descriptive statistics of transformed error rate percentage, both with and without the speed classification applied.

Music Condition		All Typists (*N* = 28)		Slow Typists (*n* = 16)		Fast Typists (*n* = 12)	
		M (%)	*SD* (%)	*M* (%)	*SD* (%)	*M* (%)	*SD* (%)
Without Music		0.79	0.40	0.90	0.42	0.65	0.34
Alt Rock	**Without Vocals**	0.83	0.38	0.99	0.37	0.62	0.29
	With Vocals	0.83	0.38	0.91	0.41	0.72	0.30
Pop Rock	**Without Vocals**	0.85	0.36	1.02	0.35	0.63	0.23
	With Vocals	0.83	0.40	0.90	0.41	0.73	0.38

Table 4.4: Descriptive statistics of transformed error rate percentage

There were no assumption violations when the transformed data was treated as a single group using either the 1 by 5 or 2 by 2 analysis paradigms. When the speed classification was applied, two outliers were introduced but they did not affect the outcomes. There were no other violations of assumptions for repeated measures or mixed ANOVAs. Table 4.5 presents the results from the statistical analysis using the 1 by 5 paradigm.

Design	Test	F	df	p	η_p^2
1 x 5	Music	0.246	4, 108	n.s.	-
1 x 5 x 2*	Music	0.187	4, 104	n.s.	-
	Music x Speed*	1.220	4, 104	n.s.	-
	Speed*	6.153	1, 26	**0.020**	0.191

Table 4.5: Inferential analysis of transformed error rate percentage
(* between groups factor)

There were no significant omnibus effects or interactions when the transformed error rate data was treated as a single group, or when the speed classification was applied. However, speed classification was a significant between groups factor, $F(1, 26) = 6.152$, $p = 0.020$, $\eta_p^2 = 0.191$. The mean transformed error percentage for the slow typists was higher than for the fast typists (slow: $M = 0.95$, $SD = 0.39$; fast: $M = 0.67$, $SD = 0.31$), demonstrating that the fast typists made significantly fewer errors. Translating this back to the original measure of error rate percentage, the fast group average error rate percentage was 0.45% while the slow group's error rate percentage was 0.90%, i.e. it was twice as large.

Table 4.6 shows the results from the analysis using the 2 by 2 paradigm.

Design	Test	F	df	p	η_p^2	Figure
2 x 2	Style	0.052	1, 27	n.s.	-	-
	Vocals	0.153	1, 27	n.s.	-	-
	Style x Vocals	0.037	1, 27	n.s.	-	-
2 x 2 x 2*	Style	0.046	1, 26	n.s.	-	-
	Vocals	0.002	1, 26	n.s.	-	-
	Style x Vocals	0.027	1, 26	n.s.	-	-
	Style x Speed*	0.003	1, 26	n.s.	-	-
	Vocals x Speed*	7.238	1, 26	**0.012**	0.218	4.13
	Style x Vocals x Speed*	0.024	1, 26	n.s.	-	-
	Speed*	7.039	1, 26	**0.013**	0.213	-

Table 4.6: Inferential analysis of transformed error rate percentage
(* between groups factor)

When the data was treated as a single group there were no significant effects or interactions. With the data split into two speed groups there were no significant omnibus effects. But, speed classification was a significant between groups factor, $F(1, 26) = 7.039$, $p = 0.013$, $\eta_p^2 = 0.213$. The fast typists made fewer errors than the slow typists (fast: $M = 0.67$, $SD = 0.30$; slow: $M = 0.96$, $SD = 0.38$).

The two-way interaction between vocals and speed group was significant, $F(1, 26) = 7.238$, $p = 0.012$, $\eta_p^2 = 0.218$, (Figure 4.13).

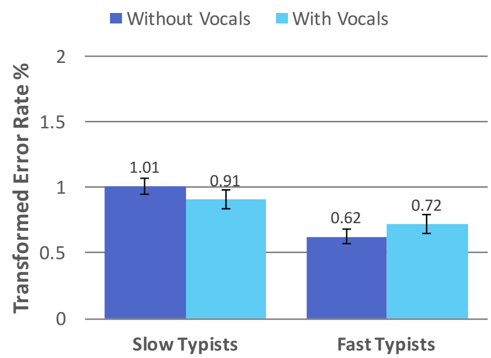


Figure 4.13: Significant interaction between vocals and speed group (p = 0.012)

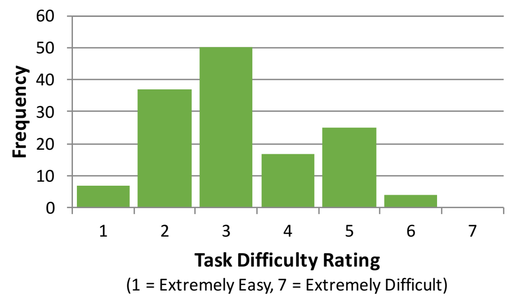
The simple main effect of vocals was significant for the fast typists, $F(1, 11) = 7.730$, $p = 0.018$, $\eta_p^2 = 0.413$, but not for slow typists, $F(1, 15) = 3.025$, *n.s.* The effect size for the fast typists was large, with 41.3% of the variance occurring due to vocals. The fast typists made fewer errors when accompanied by instrumental music and more errors when the music contained vocals (without vocals: $M = 0.62$, $SD = 0.26$; with vocals: $M = 0.72$, $SD = 0.34$).

4.2.2 Typing Experience

Two objective measures of typing experience were taken, task difficulty across all five tasks, and how distraction ratings for the music from the four music conditions.

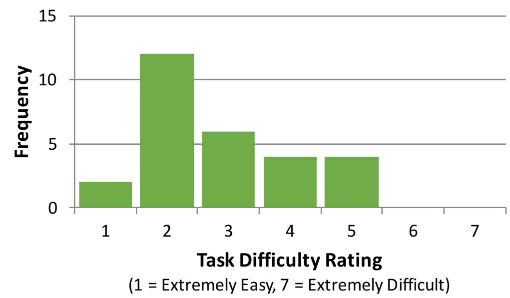
4.2.2.1 Task Difficulty

The participants rated how easy they found the transcription typing tasks using a 7-point Likert item, ranging from 1 = extremely difficult to 7 = extremely easy. The data was reversed for analysis so low values mapped to a perception that the task was easier, with higher values mapping to increased difficulty. A consolidated histogram of task difficulty ratings across all five experiment conditions was created (Figure 4.14) as well as histograms for each of the transcription typing tasks separately (Figures 4.15 to 4.17).



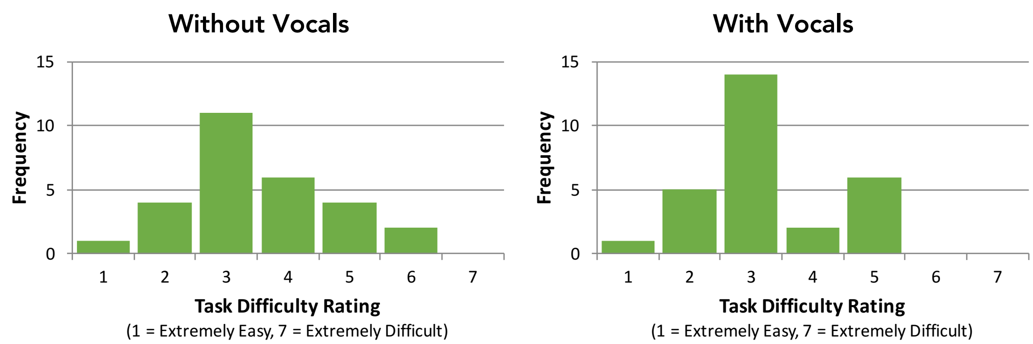
N = 140, *Q1* = 2.00, *Mdn* = 3.00, *Q3* = 4.00

Figure 4.14: Histogram of task difficulty ratings for all conditions



n = 28, *Q1* = 2.00, *Mdn* = 2.50, *Q3* = 4.00

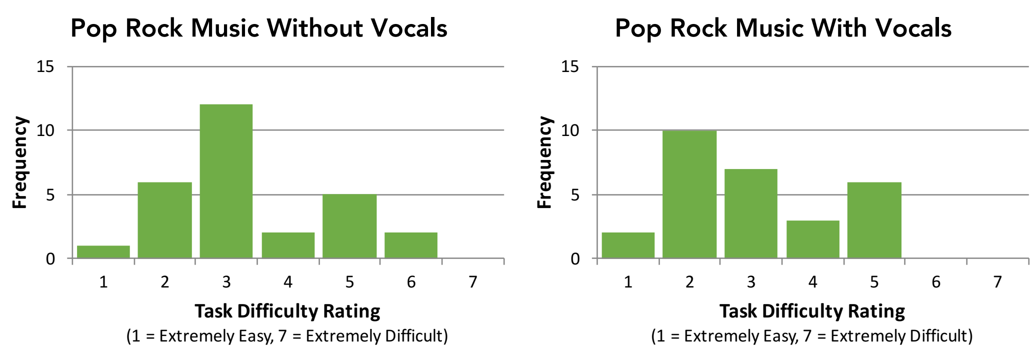
Figure 4.15: Histogram of task difficulty ratings for without music condition



(a) *n* = 28, *Q1* = 3.00, *Mdn* = 3.00, *Q3* = 4.00

(b) *n* = 28, *Q1* = 3.00, *Mdn* = 3.00, *Q3* = 4.00

Figure 4.16: Histograms of task difficulty ratings for alt rock music conditions



(a) *n* = 28, *Q1* = 2.25, *Mdn* = 3.00, *Q3* = 4.75

(b) *n* = 28, *Q1* = 2.00, *Mdn* = 3.00, *Q3* = 4.00

Figure 4.17: Histograms of task difficulty ratings for pop rock music conditions

The histogram of the alt rock without vocals condition (Figure 4.16, a) resembles a normal distribution, however the other histograms look to deviate from normality. It is also important to note that, with the exception of rating 7 (i.e. extremely difficult) the full range of ratings is used in most of the distributions.

As this ratings data is not really suitable for parametric tests, the approach taken was to perform a parametric test first and then follow up any significant or almost significant results with the closest suitable non-parametric test, as described in Chapter 3, section

3.5.3. Claims based on the outcomes from the analysis are only made based on the results from non-parametric tests. Table 4.7 presents results from the repeated measures ANOVA using the 1 by 5 analysis paradigm with and without the speed classification applied.

Design	Test	F	df	p	η_p^2
1 x 5	Music	2.326	4, 108	0.061**	0.079
1 x 5 x 2*	Music	2.196	4, 104	0.075**	0.078
	Music x Speed*	0.109	4, 104	n.s.	-
	Speed*	1.725	1, 26	n.s.	-

Table 4.7: Inferential analysis of task difficulty ratings
(* between groups factor, ** non-significant result where p <0.1)

When the data was treated as a single group, there was a trend towards a significant effect of music, $F(4, 108) = 2.326$, $p = 0.061$, $\eta_p^2 = 0.079$. Using a 1 by 5 Friedman test (the non-parametric alternative to a one-way ANOVA) there was also a trend towards a significant effect, $\chi^2(4) = 9.049$, $p = 0.060$, resulting in a very similar p value to the ANOVA result. The similar p values suggest that the ANOVA assumption violations may not be having much impact on this analysis.

The interaction between music and speed classification from the mixed ANOVA was not significant, $F(4, 104) = 0.109$, *n.s.* So, a non-parametric analysis considering the effect of music for the slow and fast typists separately was not performed.

As there is a suggestion of a trend towards a significant effect of music between the five conditions, post hoc pairwise comparisons are warranted, but care should be taken not to overstate any outcomes. To avoid over testing, it was important to minimise the number of pairwise comparisons and use an alpha level than was adjusted using a Bonferroni correction. I previously predicted that typing when accompanied by music that contained vocals would be harder than typing with instrumental music, so I made pairwise comparisons between the without music and with vocals conditions for both styles of music using a Wilcoxon signed rank test. I also expected that typing when accompanied by music containing vocals would be harder than typing without music, so I compared the without music condition to the with vocals conditions from both styles. This limits the number of pairwise comparisons to four, so the critical value of alpha after the Bonferroni correction is $p < 0.0125$. Table 4.8 shows that none of the pairwise comparisons were significant.

Condition 1	Condition 2	z	p
Without Music	Alt Rock With Vocals	-1.455	n.s.
Without Music	Pop Rock With Vocals	1.868	n.s.
Alt Rock With Vocals	Alt Rock Without Vocals	1.064	n.s.
Pop Rock With Vocals	Pop Rock Without Vocals	1.112	n.s.

Table 4.8: Post hoc tests for main effect of music
(significance at $p < 0.0125$)

Due to the experiment design, the analysis of the task difficulty ratings using paradigm 2 requires a 2 by 2 repeated measures ANVOA. Table 4.9 presents the results from the ANOVAs using paradigm 2.

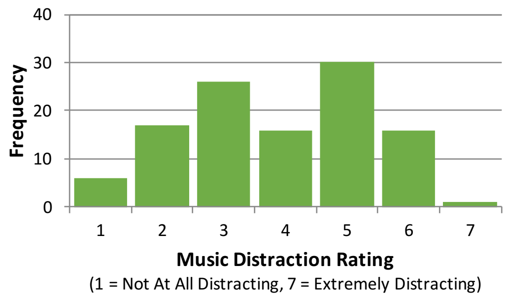
Design	Test	F	df	p	η_p^2
2 x 2	Style	1.596	1, 27	n.s.	-
	Vocals	3.927	1, 27	0.058**	0.127
	Style x Vocals	0.036	1, 27	n.s.	-
2 x 2 x 2*	Style	1.671	1, 26	n.s.	-
	Vocals	3.872	1, 26	0.060**	0.130
	Style x Vocals	0.046	1, 26	n.s.	-
	Style x Speed*	0.186	1, 26	n.s.	-
	Vocals x Speed*	0.079	1, 26	n.s.	-
	Style x Vocals x Speed*	0.046	1, 26	n.s.	-
	Speed*	1.765	1, 26	n.s.	-

Table 4.9: Inferential analysis of task difficulty ratings
(* between groups factor, ** non-significant result where $p < 0.1$)

Both with and without the speed classification applied there was a trend towards a significant effect of vocals (without classification: $F(1, 27) = 3.927$, $p = 0.058$, $\eta_p^2 = 0.127$; with classification: $F(1, 26) = 3.872$, $p = 0.060$, $\eta_p^2 = 0.130$). However, a Wilcoxon signed rank test comparing the effect of vocals overall was not significant, $z = -1.511$, *n.s.*, so this analysis was not taken further.

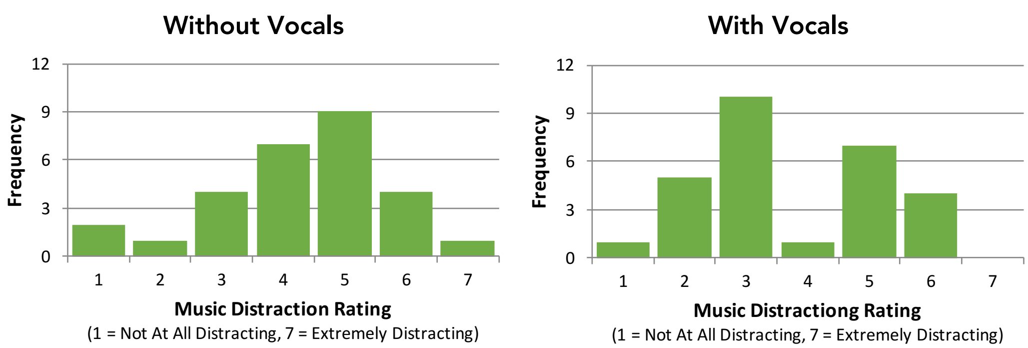
4.2.2.2 Music Distraction

The participants rated how distracting they found the music in each of the four music conditions using a 7-point Likert item, ranging from 1 = extremely distracting to 7 = not at all distracting. The data was reversed for analysis so that low values map to a perception that the music was less distracting and higher values mean it was more distracting. A consolidated histogram of music distraction ratings across all four experiment conditions was created (Figure 4.18) as well as histograms for each of the transcription typing tasks separately (Figures 4.19 and 4.20)



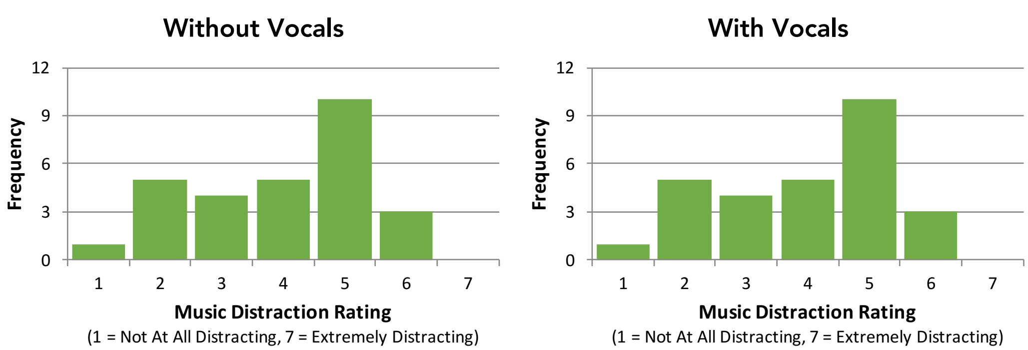
N = 112, *Q1* = 32.00, *Mdn* = 4.00, *Q3* = 5.00

Figure 4.18: Histogram of music distraction ratings for all conditions



(a) *n* = 28, *Q1* = 3.25, *Mdn* = 4.50, *Q3* = 5.00 (b) *n* = 28, *Q1* = 3.00, *Mdn* = 3.00, *Q3* = 5.00

Figure 4.19: Histograms of music distraction ratings for alt rock music conditions



(a) *n* = 28, *Q1* = 3.00, *Mdn* = 4.00, *Q3* = 5.00) (b) *n* = 28, *Q1* = 2.00, *Mdn* = 3.00, *Q3* = 5.00

Figure 4.20: Histograms of music distraction ratings for pop rock music condition

Similarly to the task difficulty ratings, the distributions are a mixture of those that resemble a normal distribution (e.g. the alt rock without vocals distribution in Figure 4.19 a) to the alt rock music with vocals distribution (Figure 4.19 b), which is clearly not normal. The full range of available values was again used. The approach of performing an ANOVA and then following up any significant effects or interactions was used. As there is no without music condition in the analysis of music distraction ratings, this data was only analysed using experiment paradigm 2. Table 4.10 shows the results using the 2 by 2 experiment paradigm.

Design	Test	F	df	p	η_p^2
2 x 2	Style	0.953	1, 27	n.s.	-
	Vocals	3.928	1, 27	0.058**	0.127
	Style x Vocals	0.177	1, 27	n.s.	-
2 x 2 x 2*	Style	0.917	1, 26	n.s.	-
	Vocals	4.134	1, 26	0.052**	0.137
	Style x Vocals	0.098	1, 26	n.s.	-
	Style x Speed*	0.004	1, 26	n.s.	-
	Vocals x Speed*	0.422	1, 26	n.s.	-
	Style x Vocals x Speed*	0.487	1, 26	n.s.	-
	Speed*	0.020	1, 26	n.s.	-

Table 4.10: Inferential analysis of music distraction ratings
(* between groups factor, ** non-significant result where p < 0.1)

Both with and without the speed classification applied there was a trend towards a significant effect for vocals (without classification: $F(1, 27) = 3.928$, $p = 0.058$, $\eta_p^2 = 0.127$; with classification: $F(1, 26) = 4.134$, $p = 0.052$, $\eta_p^2 = 0.137$). A Wilcoxon signed rank test verified a significant overall effect of vocals, $z = -2.100$, $p = 0.036$. The participants found the music with vocals more distracting than music without vocals (with vocals: *Mdn* = 3.00; without vocals: *Mdn* = 4.00).

If the participants found the music to be distracting, i.e. they selected a value between one and four on the music distraction Likert item, they were asked to explain what was distracting about the music. Twenty five of the 28 participants included at least one explanation of why they felt the music was distracting. A light weight approach to content analysis was performed to categorise the comments made. Table 4.11 shows the categories, a count of the frequency of comments made which fit into that category, and some examples of comments.

Category	Frequency	Examples
Inclusion of vocals	21	"The vocals, even the anticipation of vocals during the instrumental sections… kept gaining my attention" "It was more difficult when the man was signing 'INTO MYSELF!'" "Vocals confused what I was typing so couldn't really process what I was reading"
Too loud	12	"Too loud, couldn't hear keystrokes" "I thought it was quite loud" "Slightly too loud"
Rhythm	6	"Primarily no consistent rhythm" "Varying rhythms" "Changed beat a lot"
Too repetitive	6	"Too … repetitive" "It was really repetitive"
Non-specified Irregularity	5	"It kept changing quite a lot" "The lack of regularity kept gaining my attention"
Dislike	3	"Not the kind of music I like" "The singer was not very good"
Unfamiliar music	2	"As I didn't know the track I couldn't banish it to my subconscious to sing along"
Heard in previous task	2	"I was trying to work out whether it was the same track [as in an earlier task], or just similar"
Lack of vocals	1	"Lack of vocals made it hard to anticipate pattern"
Instrumentation	1	"The guitars kept changing, there'd be only one, then many, then one"

Table 4.11: Qualitative analysis of music distraction justifications

By far, the most frequent comments involved finding vocals in the music distracting. However, it is interesting that one participant contrastingly found the lack of vocals in the instrumental music distracting. Volume was the other dimension of the music that was most frequently referred to in the comments, with it being too loud. It is also interesting that both the repetitive nature and the irregularity of the music feature as aspects of the music that are distracting. With regard to the irregularity, there were five instances where the participants commented that the music was irregular, without specifying what aspect of the music they were referring to. Other comments specifically highlighted irregular rhythms, or changing instrumentation.

4.2.3 Summary of Results

The results section contains a lot of analyses, so a summary of the key results is included in the following sections.

4.2.3.1 Typing Performance

In terms of typing speed there were no significant effects caused by the music in the 1 by 5 analysis, both with and without the speed classification applied. Unsurprisingly, as speed was the factor that determined each individual participant's classification as either a fast or slow typist, there was a significant difference in CPM between the two groups.

With the 2 by 2 paradigm, the three-way interaction between style, vocals and speed group was significant. The simple interaction between style and vocals was significant for fast typists but not for the slow. The simple interaction between style and vocals had a large effect size for the fast typists, so it should follow that either the simple simple main effect of vocals or the simple simple main effect of style reached significance. However, this was not the case as none of the simple simple main effects, either for vocals or style, were significant.

Although not quite reaching significance, the simple simple main effects of vocals approached significance with pop rock but not alt rock music. With pop rock music, CPM was higher without vocals and lower with vocals. Further, the simple simple main effect of style approached significance in the without vocals conditions, but was not significant for the with vocals condition. When accompanied by music that contained vocals, CPM was higher with pop rock and lower with alt rock music.

In terms of error rate, speed classification was a significant between groups factor in both analysis paradigms, with lower error rates achieved by the fast typists than the slow. In the 1 by 5 paradigm, there were no other significant effects or interactions. But, with the 2 by 2 by 2 analysis paradigm the two-way interaction between vocals and speed group was significant, with a significant simple main effect of vocals for the fast typists but not the slow. The fast typists made fewer errors when accompanied by music that did not contain vocals.

4.2.3.2 Typing Experience

There were no clear effects identified on either of the typing experience metrics for perceived difficulty. However, music without vocals was perceived as less distracting than music with vocals.

4.3 DISCUSSION

Although there were no omnibus effects for typing speed caused by the presence of vocals in the music, or the style of music, there is strong evidence that the effect of music differs according to whether the participant was classified as a slow or fast typist. In this experiment, there was no evidence to suggest the slow typists were affected by the presence of music, the vocals or the style of music. But, for the fast typists it seems that, with the pop rock music the presence of vocals in the music may have had a negative effect on typing speed. The simple effects analysis for CPM did not quite reach the required value of $p < 0.050$ needed for significance, achieving $p = 0.098$ instead. However, with just 12 participants in the group and a large effect size there is a suggestion that the hypothesis that vocals in the music has a negative effect on typing speed may be correct, at least for the fast typists. This experiment has failed to provide clear evidence that vocals in the music has a negative impact on the speed of the faster, more skilled, typists, although the trends in the analysis indicate this does warrant further investigation.

The experiment did not provide any evidence to suggest that the music accompaniment affected typing accuracy differently to the without music condition. The participants did not make significantly more or fewer errors in the without music condition over any of the conditions with music playing. The slow typists made more errors during the tasks than the fast typists, which indicates the fast typists were better in terms of accuracy as well as speed and that typing fast did not automatically lead to an increase in uncorrected errors.

In the 2 by 2 analysis, similarly to the typing speed results, the effect of the speed classification was important. The effects of music vocals differed according to whether the focus was on the fast typists or the slow typists. For the fast typists, fewer uncorrected errors were identified in the without vocals conditions. But, for the slow typists there is insufficient evidence to suggest that vocals in the music reduced typing accuracy.

The initial aims of this experiment were taken with the assumption that the 28 participants would be treated as a single group. In this first experiment typing ability was not intended to be included as a factor under investigation, although there was a concern that the effect of music on typing performance might differ depending on the experience of the typist. However, one clear outcome from this experiment is that the participants can be classified as a slow or fast typist and that this classification has a large effect on the outcomes of the analysis.

The detailed analyses presented in the previous section have shown that classification into the two speed groups is important for two reasons. Firstly, inclusion of this between groups factor reduces the degrees of freedom for the error term by one, which seems to improve the fit of the statistical model to the collected data resulting in greater statistical power, and a more sophisticated analysis with better partitioning of the error variance. Secondly, and more importantly, while the omnibus effects of the objective measures of typing performance were typically not significant, the interactions between speed group and various IVs resulted in significant effects. Typically, significant effects were identified for the fast typists that were not evident for the slow typists.

The fact that there were big differences in effect depending on whether the participant had been categorised as either a fast or slow typist is particularly interesting as, in the difficulty and distraction ratings, there was no obvious difference between the effect of any of the music IVs between the two different groups of typists. Given there was a suggestion that vocals in the music was a distracting factor for both typist groups, it leads to the question why is this not reflected in the speed and accuracy data for both speed groups? Is this due to a lack of power? Or perhaps something else?

Further, speed classification was a significant between groups factor with both the typing speed and accuracy measures. The former is not surprising, given that the speed of typing was the deciding factor as to which group the participant was classified into. However, despite the error rate percentage scatterplots clearly showing a mix between slow and fast typists, there was still a significant difference in the transformed error rate between the slow and fast typists, with the fast group making fewer errors than the slow group. This tells us that the fast typists are not just faster and making more keypresses, but they are making fewer errors, so are better typists, i.e. they are more skilled as typists, and not just faster.

4.3.1 Discussion of Each Experiment Aim

The initial aims of this experiment were as follows:

1. to investigate whether transcription typing when accompanied by music affects typing performance and experience,
2. to investigate whether the presence of vocals in a piece of music reduces typing performance and experience,
3. to establish if any effects occur across two different styles of music,
4. to confirm the method is suitable for use in future experiments.

The results presented in the previous section are discussed in more detail in relation to the specific experiment aims in the following sections.

4.3.1.1 Experiment Aim 1

The first experiment aim was to compare whether typing performance and experience was different in the conditions where the tasks included a music accompaniment to a without music condition. To explore this aim, the omnibus effect of music in the 1 by 5 analysis paradigm needs to be examined.

The omnibus effect of music was not significant, even with the speed classification applied, so the results from this experiment do not provide any evidence that music can be used to improve typing performance or experience. Whilst there was no improvement in either typing speed or accuracy, there was also no significant decrease in performance due to the music. It is possible that type II errors are concealing any significant effects due to the modest participant numbers. But, within this experiment, there was no evidence of either positive or negative differences between any of the four music conditions and the without music condition.

As there is no evidence that performance or experience has been affected by the pieces of music, the first experiment aim has not been achieved.

4.3.1.2 Experiment Aim 2

The second aim of the experiment focused on investigating the effect of vocals on performance and experience, with the expectation that this dimension of music would probably reduce performance by making the task more difficult as the music with vocals should be more distracting than the music without vocals. To establish if the experiment was a success in terms of this aim the omnibus effect of vocals in the 2 by 2 experiment paradigm needs to be studied.

The omnibus effect of vocals on CPM was not significant, either with or without the speed classification included in the analysis. The two-way interaction between speed group and vocals was also non-significant. Therefore, this experiment has been unable to provide evidence that vocals in music affects typing speed when both pieces of music were considered together.

However, in terms of error rate, although the omnibus effect of vocals was not significant, there was a significant two-way interaction between speed group and vocals. The simple effects analysis revealed that vocals did have reduce the fast typists' performance in terms of typing accuracy. This result demonstrates a difference in how vocals affects performance between the typing speed and accuracy measures. There are two possible explanations for this difference. The first and simplest explanation is that the vocals in the music did affect the participants' typing speed and accuracy measures in different ways. This explanation reflects Jensen's (1931) experiment where a decrease in speed did not correspond with an increase in accuracy and vice versa. Another explanation is that the speed measure has less power than the accuracy measure so significant differences were revealed for accuracy but not for speed. To establish if this is the case, a similar experiment with larger numbers of participants is needed.

In terms of the experience of typing, this experiment has provided evidence that music with vocals is perceived as more distracting than music without vocals, with no dependence on speed classification. I would expect this result to be reflected in the objective measures of typing performance, but the increased distraction did not lead to reduced performance in terms of speed, only for accuracy and only for the fast typists.

This experiment has demonstrated that, for fast typists, the presence of vocals in the music had an overall negative effect on typing accuracy and that overall music with vocals was perceived as more distracting than music without vocals. Therefore, this experiment has partially achieved the second experiment aim.

4.3.1.3 Experiment Aim 3

The third aim of this experiment was to explore if any identified effect caused by the vocals was similar between two pieces of rock music from different styles. To explore success in terms of this aim the omnibus effect of style and the interaction between style and vocals needs to be studied.

The omnibus effect of style was non-significant for all four DVs, indicating that the style of rock music did not affect either typing performance or experience sufficiently to lead to a significant effect. If there was an overall effect of style, it was too small for a significant omnibus effect of style in any of the DVs.

The two-way interaction between style and vocals approached significance for the typing speed measure only. But, what is more interesting, is the significant three-way interaction between style, vocals and speed group for CPM. This result means the effect of vocals differs depending upon the style of music and the speed group. Rather than being consistent across the different styles of music, the results support the idea that the presence of vocals in the music has different effects depending on whether it was in the alt rock or the pop rock style music.

The simple interaction between style and vocals was not significant for slow typists for typing speed. For fast typists though, the simple interaction between style and vocals was significant with a large effect size. However, although some comparisons approached significance none of the simple simple main effects achieved p values less than 0.050. So, the precise cause of this large significant interaction for fast typists is not understood. Again, these results suggest a similar experiment with a larger sample size, particularly including more fast typists, is needed.

The same three-way interaction was not significant in the analysis of typing accuracy, which adds weight to the idea that the effect of music on typing speed differs to the effect of music on typing accuracy.

4.3.1.4 Experiment Aim 4

The final aim of the experiment focused on the method, to check whether experiments of this type were suitable to explore the underlying research question regarding affecting work related computing task performance with music. The method in this experiment was successful in that a large amount of data was collected and analysed with some interesting results, particularly with regard to the statistically significant interactions. However, the post hoc simple effects analyses, which are typically quite liberal tests, did not provide clear statistically significant results allowing me to identify why the interactions were significant with large effects sizes.

Given the lack of significant omnibus and simple main effects identified in this experiment, it is worth considering whether there are differences occurring as a result of the experiment interventions through the IVs but that they are being obscured by the large variance and the closeness in mean values across the conditions, i.e. a lack of power in the experiment. It is also possible that a ceiling effect is being encountered. For example, the typing task might be so easy that any impact the music is having on typing performance is masked because the task is too simple meaning that even though there may be variability in the individual typing speeds of the participants due to the music IVs, as the participants are all able to type at their fastest in all conditions because the task is simple, the effect is not observable in the results.

Given the importance of the speed classification, the approach of applying the classification post hoc based on the speeds achieved needs some investigation, as there may be another, better approach to achieve this grouping. It is possible that instead of defining the two groups based on the speed alone, they could be classified based on their method of typing. It may be that all the participants classified as fast typists are using a touch typing approach where they have learned the position of all keys and keep their eyes on the screen at all times. Alternatively, if not a formal touch typing method, these participants may have developed their own idiosyncratic but fast method of typing which does not require them to look for the keys, with most of their attention focused on the screen and not the keyboard. The slow typists may all be using something similar to a hunt and peck method which requires them to have to frequently look down to the keyboard between keypresses, thus reducing their overall speed, and switching their eye gaze between the screen and keyboard.

Unfortunately, video recording was not included in the experiment, and I was not watching the participants closely to identify their method of typing, so cannot report how the participants were typing. Therefore, in this experiment, it was more appropriate to classify the typists as either slow typists or fast typists based on their CPM collected rather than defining the groups based on an assumption of their typing approach. In future experiments, I planned to watch the participants to try to establish if the slow typists were using one particular method while the fast typists were using another.

4.3.2 Limitations of This Experiment

For a human computer interaction (HCI) experiment using a fully factorial repeated measures design, 28 participants would be considered a good sample size. However, the classification into two speed groups split the sample into one group of 16 slow typists, and another group of just 12 fast typists. The modest number of participants once the classification was applied is a limitation of this experiment.

The results suggest the music only affected typing performance of the fast typists, but more participants are needed in the both groups to verify this. The interaction effects involving speed classification indicate it is the fast typists group where effects are occurring due to the music, but the post hoc simple effects analyses did not result in significant *p* values. These non-significant results are probably because 12 participants in the fast group is an insufficient number for the experiment to have sufficient power to reach significant results in the post hoc analyses.

A further limitation of this experiment, which might have affected the results, is that the typing task itself might have been so easy that the impact of any of the IVs is marginal and was unable to be identified in the analysis. The inconsistency between distraction ratings and speed analysis, and the fact that when it came to distraction there was no between groups effect due to the speed classification, suggests there may be an effect for all typists, but it is not revealed in the speed and accuracy data because of large within group variance, or ceiling effects.

The variance in typing speeds, in particular, was large. For any effect due to the IVs to be noticeable with such a large variance, the effect would need to be very large. This possible ceiling effects require further investigation. It may be possible to alter the typing task so that the ceiling effect is removed allowing any impact caused by the music to be identified.

4.3.3 Moving Forwards

The fact that none of the pieces of music had a significant effect when compared to the without music typing tasks is important and requires further investigation as it is the foundation of the research in this thesis. If the music had a consistently negative effects when compared to the without music condition, it would have been hard to argue that in this research I could, potentially, identify parameters of music that could positively affect performance at work. One piece of music being better than the other is not much use if without music is clearly the condition that leads to best performance.

The similarity between the music and without music conditions may be due to ceiling effects. So, to provide further evidence that music does not reduce typing performance when compared to typing without music, and to minimise the potential of a ceiling effect, it is important to try to vary the difficulty of the task. Changing the difficulty of the task also has the potential to also increase the levels of concentration required by all participants, which should ensure the fast typists are having to concentrate on the task. It is possible that the effects were noticeable in the fast typist group because the effect is large, but the effect for slow typists is smaller and was concealed by the large variance, all of which requires further investigation.

This experiment has provided evidence that the distribution of typing speed may be bimodal, and how someone is affected by music when transcription typing may depend on whether they are a skilled or non-skilled typist. The post hoc classification of participants as either a slow or fast typist led to two modest sized groups (slow typists: $n = 16$; fast typists: $n = 12$) rather than a single group of 28 participants. These relatively small group sizes limit the generalisability of the results and may have led to type II errors due to reduced statistical power. As the music affected participants differently depending on their skill level, task difficulty is an interesting factor to explore further.

Chapter 5

Genre And Task Difficulty

Experiment 2

The primary aim of Experiment 2 was to investigate how music from different genres with very different musicological characteristics affect typing performance and experience across the two different speed groups. The approach of manipulating genre in this experiment mimics the design in Jensen's (1931) experiment. However, instead of comparing the effects of jazz and dirge music to a without music condition, this experiment looks at the effect of ambient, classical and rock music on typing performance when compared to a without music condition.

Experiment 2 also incorporated a task difficulty manipulation by using different source material for the presented text. The intention was to include easy, medium and hard difficulty tasks in the experiment through changing the text. The objective of the task difficulty manipulation was twofold. First, to reduce the impact of any ceiling effects, hopefully allowing the influence of the music manipulations to emerge in the objective data. Secondly, using three texts with different difficulty levels allows me to explore how the fast typists were affected by music in tasks that require high levels of concentration and attention in order to maintain strong performance. The results from Experiment 1 showed the slow typists were not affected by music in the same way as the fast typists. I hypothesized the difference in the scale of the impact of music between the slow and fast typists was because the slow typists found transcription typing harder and needed to concentrate more which reduced any impact due to the music so far that there was either no effect due to music at all, or it was too small to emerge in the objective data because of the large variance. In this experiment, if the fast typists' performance is only affected by music in the easier tasks and not in the more difficult tasks, there would be further evidence to support my hypothesis that the more someone needs to concentrate on the task, the less the music affects their performance.

5.1 Method

5.1.1 Aims

The aims of this experiment were:

1. to investigate if the genre of music affects typing performance and experience, including a comparison to a without music condition,
2. to investigate whether task difficulty has a moderating effect on typing performance and experience,
3. to gather further evidence that the effect of music on typing performance and experience differs between slow and fast typists.

5.1.2 Participants

The participants in this experiment were chosen through a convenience sample, recruited using email lists, adverts on university boards, and via word of mouth. The experiment had 42 participants (15 female, 27 male). Table 5.1 shows the participants' ages.

Age Group	**< 18**	**18 – 24**	**25 – 34**	**35 – 44**	**45 – 60**	**> 60**
Frequency	0	22 (52%)	16 (38%)	3 (7%)	1 (2%)	0

Table 5.1: Participant ages

To ensure the sample was not biased towards slow typists, towards the end of the recruitment process potential participants were asked to verify their typing speed was higher than 65 words per minute (WPM), which is equivalent to 325 characters per minute (CPM).

All participants were native English speakers who did not have any known hearing impairments and were not dyslexic. All the participants were involved in academia at the University of York. Forty students took part in the experiment, 20 were postgraduate research students (13 from Computer Science, 7 from Arts or Humanities subjects) and 20 participants were undergraduates studying Computer Science or Electronics. Two administrators also took part in the experiment. All participants were compensated with a £10 Amazon or Marks and Spencer's voucher.

5.1.3 DESIGN

This experiment used a non-factorial two-way mixed design with repeated measures. Half of the participants did the transcription typing tasks using easy and medium difficulty typing tasks and the other participants did the medium and hard difficulty tasks.

There were two repeated measures independent variables (IVs). The first IV was music (four levels: without, ambient, classical and rock) and the second IV was task difficulty (group 1: easy and medium; group 2: medium and hard). The IVs were counterbalanced to avoid fatigue and practice effects.

Each participant completed eight transcription typing tasks:

- without music using the advanced English text (music: without; task difficulty: medium),
- accompanied by ambient music using the advanced English text (music: ambient; task difficulty: medium),
- accompanied by classical music using the advanced English text (music: classical; task difficulty: medium),
- accompanied by rock music using the advanced English text (music: rock; task difficulty: medium),
- without music using *either* the simple English or Dutch text (music: without; task difficulty: easy or hard),
- accompanied by ambient music using *either* the simple English or Dutch text (music: ambient; task difficulty: easy or hard).
- accompanied by classical music using *either* the simple English or Dutch text (music: classical; task difficulty: easy or hard),
- accompanied by rock music using *either* the simple English or Dutch text (music: rock; task difficulty: easy or hard).

In this chapter, I refer to simple English, advanced English and Dutch tasks rather than easy, medium and hard tasks. This terminology is used to avoid confusion with the perceived task difficulty ratings dependent variable (DV). More information about the music and text IVs is included in Chapter 3, section 3.3.1.1.

There were three design paradigms in this experiment. The first paradigm looks at the effect of music (four levels: without, ambient, classical and rock) within the medium difficulty tasks only by taking data from the advanced English tasks from all 42 participants. The arrangement of IVs for this first design paradigm is shown in Figure 5.1.

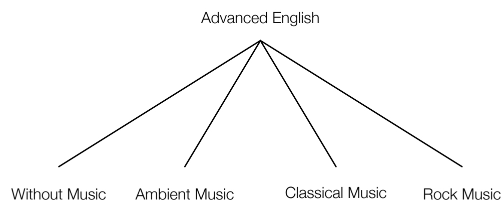


Figure 5.1: Structure of IVs for experiment paradigm 1

The second design paradigm takes data from the 21 participants who did transcription typing tasks using the simple and advanced English texts. The arrangement of IVs for the second design paradigm is shown in Figure 5.2.

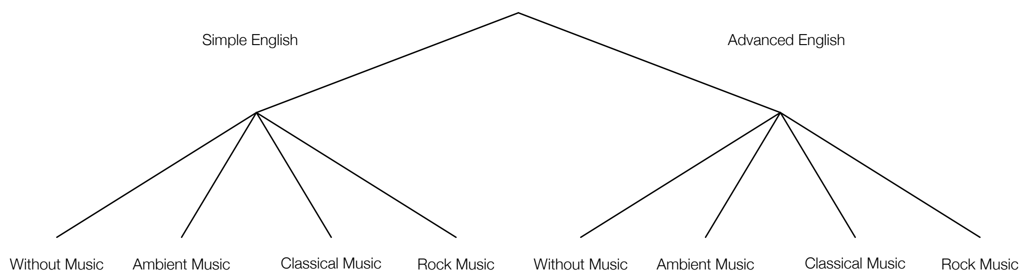


Figure 5.2: Structure of IVs for experiment paradigm 2

The final design paradigm is illustrated in Figure 5.3 and compared the four music conditions across the advanced English and Dutch texts.

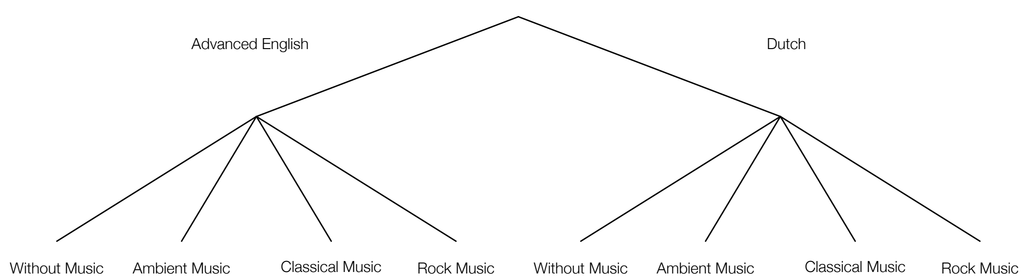


Figure 5.3: Structure of IVs for experiment paradigm 3

The DVs were typing performance and typing experience. Typing performance was measured in both speed (CPM) and accuracy (error rate percentage). Typing experience was measured quantitatively using 7-point Likert items for perceived task difficulty and music distraction and qualitatively by asking the participants to explain why they perceived the music as distracting. See Chapter 3, section 3.3.1.2 for more on how the DVs were measured and calculated.

5.1.4 Equipment and Materials

This experiment was performed using a laboratory based methodology in a quiet Usability Lab in the Department of Computer Science at the University of York. The participants used a desktop PC running Windows XP with Firefox version 12.

The music was controlled using Audacity (Open Source, 2012) running on my MacBook Pro, which was connected to a pair of Philips SPA 2210 2.0 laptop speakers. The speakers themselves and the audio level of the MacBook Pro were both set to maximum volume, while the playback volume in Audacity was set to 75%.

The rock music in this experiment was the without vocals version of *Atrophy* (The Doppler Shift, n.d.) (the alt rock style used in Experiment 1). The classical music came from *Allegro Assai* movement from Mozart's Piano Concerto No 20 in D minor (2005). The ambient music was *Continuity Pt. 1* (Liquid Mind, 2001). See the Music Independent Variables section in Chapter 3, Section 3.3.1.1 for more on how the music IVs were selected.

The texts used in this experiment were the simple and advanced English and Dutch texts. See the Presented Text IVs section in Chapter 3, Section 3.3.1.1 for more information on the task IVs.

Participants completed short questionnaires to collect subjective ratings of task difficulty and music distraction. These questionnaires included an open ended question asking why the music was distracting, as well as a question to establish if they had heard that particular piece of music before. The participants also completed a demographics questionnaire. All questionnaires are included in Appendices A.2, A.3 and A.4.

5.1.5 Procedure

Participants with an odd participant ID number were allocated to the group doing the simple and advanced English tasks. Participants with an even ID completed the advanced English and Dutch tasks. The procedure followed that described in Chapter 3, Section 3.3.3.1 exactly. All participants completed eight transcription typing tasks and spent 36 minutes typing during the experiment.

5.2 Results

The results section begins by carefully considering the histograms and scatterplots of CPM from the advanced English tasks using data from all 42 participants to establish an appropriate speed classification structure. Given the importance of the speed classification shown in Experiment 1, it is necessary to explore and adopt a single, consistent approach to classification across the two task groups. Once an agreed classification structure is defined based on the CPM data, the results section continues by presenting the descriptive and inferential analysis using the three different experiment design paradigms described in Section 5.1.3.

5.2.1 Speed Classification

A histogram of CPM data was created from all advanced English tasks across both experiment groups and all four music conditions, (Figure 5.4).

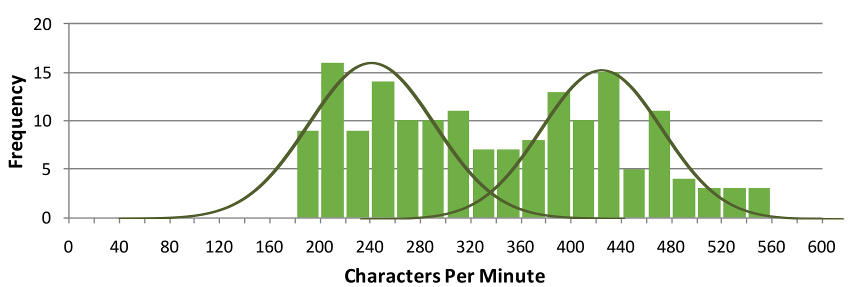
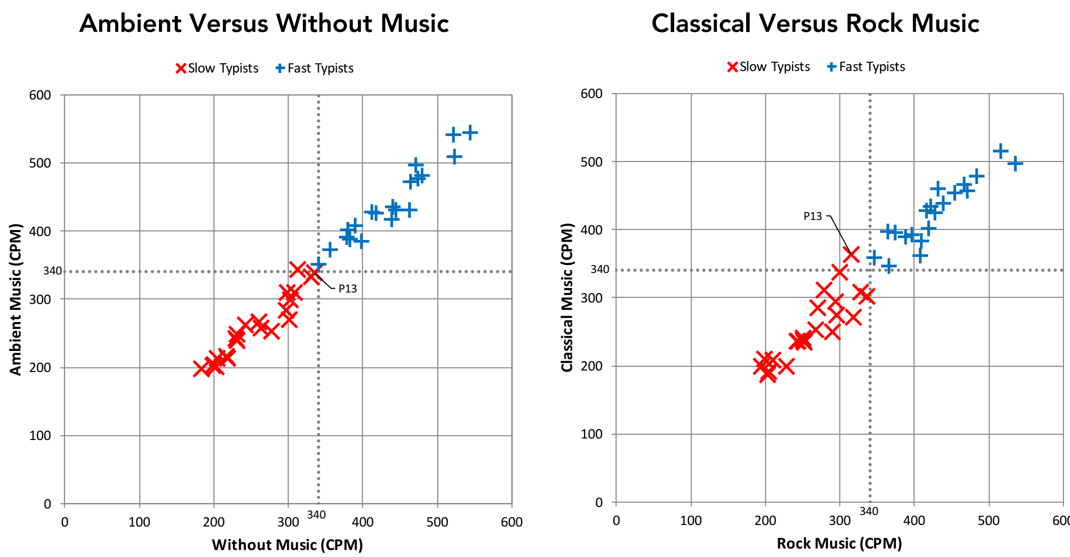


Figure 5.4: Histogram of CPM for the advanced English tasks
n = 168, *M* = 341.02, *SD* = 99.47

Figure 5.4 shows that the bimodal characteristic identified in Experiment 1 was retained, though, it is not as prominent. Approximations of normal distributions have been overlaid to show the crossing point between the two distributions is at roughly 340 CPM. Two scatterplots of CPM were created to explore whether classifying the participants into slow and fast groups based on a threshold of 340 CPM was appropriate. The first scatterplot compares CPM without music to ambient music (Figure 5.5 a). The second scatterplot compares CPM when accompanied by classical music to rock music (Figure 5.5 b).



(a) (b)

Figure 5.5: Scatterplots of CPM

Participant P13 is highlighted in the scatterplots as they were classified as a slow typist despite achieving above the 340 CPM threshold in the rock music condition because they were below the threshold in the other conditions. All other participants fitted neatly into the classification structure using a 340 CPM threshold. The classification resulted in 22 participants designated as slow and 20 as fast typists and was adopted in the data analysis in all three experiment paradigms.

5.2.2 Analysis Using Design Paradigm 1

Experiment design paradigm 1 looks at the effect of music in the advanced English task conditions only, using a 1 by 4 repeated measures design. The results from the analysis of typing performance are presented first, followed by the analysis of typing experience.

5.2.2.1 Typing Performance

This section presents descriptive and inferential statistics for CPM and error rate percentage. Before the inferential analysis, all distributions were assessed for a) the presence of outliers which may distort the results using box plots and b) deviations from normality using Shapiro-Wilk's tests. Box plots and results from Shapiro-Wilk's tests for the analysis using paradigm 1 can be found in Appendix E1 .1. Outliers which affected the outcomes from the analysis were capped using the procedure described in Chapter 3, Section 3.5.1.3.

Characters Per Minute

Inferential analysis of the CPM data was performed using a 1 by 4 repeated measures ANOVA treating the data as a single group of 42 participants and as a mixed ANOVA including the speed classification as a between groups factor. When treated as a single group, two of the four CPM distributions (50%) had moderately significant deviations from normality. With the speed classification applied none of the eight CPM distributions deviated from normality. All other assumptions for repeated measures and mixed ANOVAs were met. Table 5.2 shows descriptive statistics of CPM for all typists and separated into the slow and fast groups.

Music Condition	All Typists (N = 42)		Slow Typists (n = 22)		Fast Typists (n = 20)	
	M (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)
Without	345.34	103.97	259.60	47.10	439.66	54.73
Ambient	341.76	104.52	256.10	47.35	435.99	57.05
Classical	340.73	95.15	262.46	44.01	426.82	50.01
Rock	336.25	97.30	256.24	49.05	424.27	46.94

Table 5.2: Descriptive statistics of CPM

Table 5.3 presents the results from both ANOVAs.

Design	Test	F	df	p	η_p^2	Figure
1 x 4	Music	2.753	3, 123	**0.045**	0.063	5.6
1 x 4 x 2*	Music	3.120	3, 120	**0.029**	0.072	5.6
	Music x Speed*	3.405	3, 120	**0.020**	0.078	5.7
	Speed*	136.156	1, 40	< **0.0005**	0.773	-

Table 5.3: Inferential analysis of CPM
(* between participants factor)

When the data was treated as a single group, the omnibus effect of music was significant, $F(3, 123) = 2.753$, $p = 0.045$, $\eta_p^2 = 0.063$, (Figure 5.6). Similarly, with the speed classification applied, the omnibus effect of music was significant, $F(3, 120) = 3.120$, $p = 0.029$, $\eta_p^2 = 0.072$, (Figure 5.6).

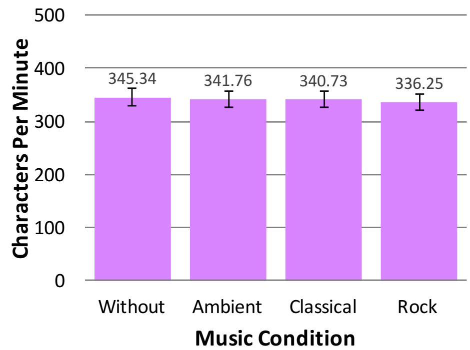


Figure 5.6: Significant omnibus effect of music ($p = 0.029$)

To follow up the significant omnibus effect of music from the mixed ANOVA post hoc pairwise comparisons were performed. The results of this analysis are shown in Table 5.4.

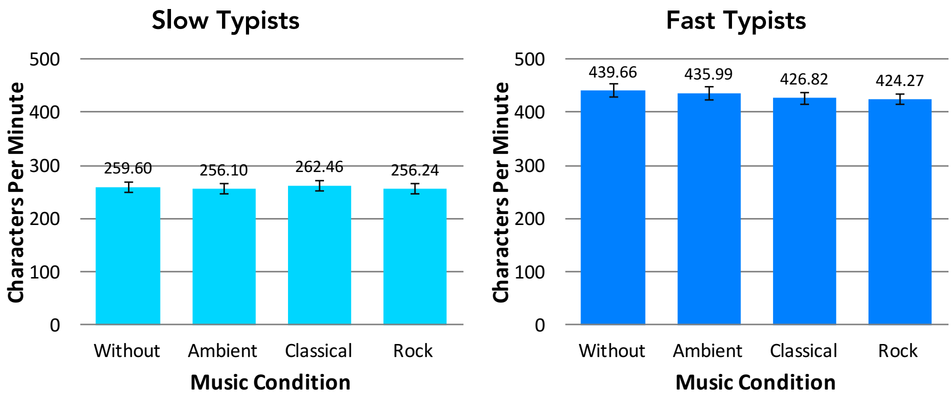
Condition 1	Condition 2	Mean Difference	p
Without	Ambient	3.581	n.s.
Without	Classical	4.982	n.s.
Without	Rock	9.371	**0.005**
Ambient	Classical	1.402	n.s.
Ambient	Rock	5.790	n.s.
Classical	Rock	4.389	n.s.

Table 5.4: Post hoc tests for omnibus effect of music
(significance at $p < 0.0083$)

CPM without music was significantly higher than CPM with rock music playing (without music: $M = 345.34$, $SD = 103.97$; rock: $M = 336.25$, $SD = 97.31$).

Speed classification was a significant between groups factor, $F(1, 40) = 136.156$, $p < 0.0005$, $\eta_p^2 = 0.773$. The fast typists' CPM was significantly higher than the slow typists' (fast: $M = 431.67$, $SD = 52.18$; slow: $M = 258.60$, $SD = 46.88$).

The two-way interaction between music and speed group was significant, $F(3, 120) = 3.405$, $p = 0.020$, $\eta_p^2 = 0.078$, (Figure 5.7).



(a) (b)

Figure 5.7: Significant interaction between music and speed group (p = 0.019)

The simple main effect of music was significant for fast, $F(3, 57) = 6.520$, $p = 0.001$, $\eta_p^2 = 0.255$, (Figure 5.7 b), but not for slow, $F(3, 63) = 0.863$, *n.s.*, (Figure 5.7 a), typists.

Post hoc pairwise comparisons were performed for the fast typists, (Table 5.5).

Fast Typists			
Condition 1	**Condition 2**	**Mean Difference**	**p**
Without	Ambient	3.667	n.s.
Without	Classical	12.833	**0.002**
Without	Rock	15.389	**0.002**
Ambient	Classical	9.167	0.040**
Ambient	Rock	11.722	**0.008**
Classical	Rock	2.556	n.s.

Table 5.5: Post hoc tests for omnibus effect of music for fast typists
(significance at $p < 0.0083$, ** significant at the non-adjusted level)

The fast typists' CPM without music was significantly higher than when accompanied by either classical or rock music (without music: M = 439.66, SD = 54.73; classical: M = 426.82, SD = 50.01; rock: M = 424.27, SD = 46.94). Their CPM when accompanied by ambient music was also significantly higher than with rock music (ambient: M = 435.99, SD = 57.05; rock: M = 424.27, SD = 46.94).

For the fast typists, the difference in CPM between the ambient and classical music conditions was significant at the non-adjusted rather than Bonferroni adjusted level (ambient: M = 435.99, SD = 57.05; classical: M = 426.82, SD = 50.01).

Error Rate

In the first experiment, the error rate data had a strong positive skew and needed a square root transformation to improve the normality sufficiently for ANOVA to be an appropriate statistical test. A histogram of the non-transformed data across all music conditions shows a similar strong positive skew (Figure 5.8) and all of the non-transformed distributions deviated strongly from normality.

A square root transformation was applied. However, the analysis of normality of these transformed distributions still led to all distributions strongly deviating from normality when treated as a single group. With the speed classification applied, three of the eight distributions (37.5%) deviated significantly from normality. So, to improve the results of the transformation further logarithms were taken instead of the square root following the procedure described in Chapter 3, Section 3.5.2.

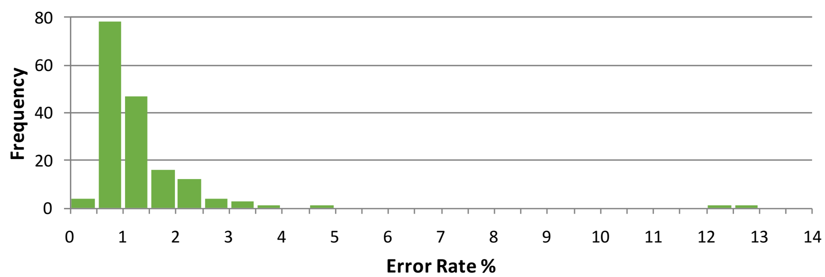


Figure 5.8: Histogram of error rate percentage for the advanced English text
n = 168, M = 0.854, SD = 1.41

Figure 5.9 shows the distribution of transformed error rate percentage using logarithms for all conditions combined in the advanced English tasks.

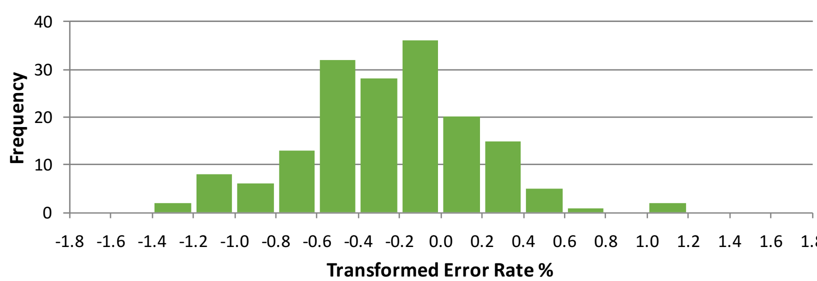
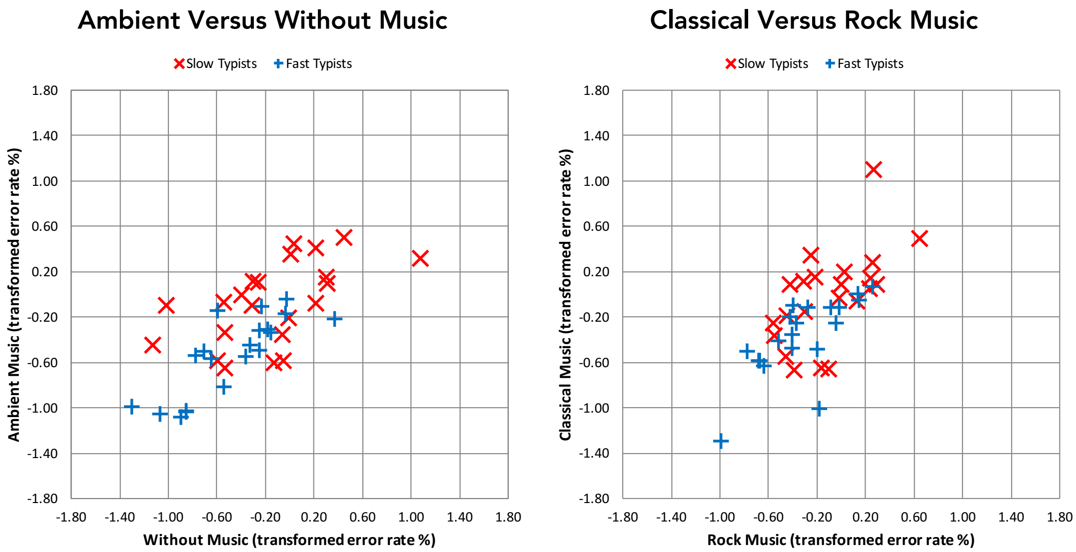


Figure 5.9: Histogram of logarithmically transformed error rate percentage for advanced English text
n = 168, M = -0.257, SD = 0.40

Scatterplots of transformed error rate percentage were created to investigate whether there was a relationship between speed group and typing accuracy (Figure 5.10).



(a) (b)

Figure 5.10: Scatterplots of transformed error rate percentage for advanced English text

Visual inspection of the scatterplots suggests typing performance in speed may not be a predictor of performance in accuracy. The scatterplots also suggest that there may be some outlying data points. Box plots verified that with the data treated as a single group there were two outliers and with the data was split into speed groups there were two different outliers. These outliers affected the results of the inferential analysis so were capped following the procedure described in Chapter 3, Section 3.5.1.3. After capping, none of the distributions deviated significantly from normality. Table 5.6 shows descriptive statistics of transformed error rate percentage.

Music Condition	All Typists (N = 42)		Slow Typists (n = 22)		Fast Typists (n = 20)	
	M (%)	SD (%)	M (%)	SD (%)	M (%)	SD (%)
Without	-0.295	0.42	-0.074	0.36	-0.538	0.35
Ambient	-0.324	0.45	-0.149	0.49	-0.486	0.41
Classical	-0.207	0.35	-0.097	0.33	-0.328	0.33
Rock	-0.203	0.38	-0.048	0.35	-0.359	0.31

Table 5.6: Descriptive statistics of transformed error rate percentage

The transformed error rate data was analysed both with and without the speed classification applied. Table 5.7 shows the results from the two ANOVAs.

Design	Test	F	df	p	η_p^2	Figure
1 x 4	Music	3.032	3, 123	**0.032**	0.069	5.11
1 x 4 x 2*	Music	2.816	3, 120	**0.042**	0.066	5.11
	Music x Speed*	1.825	3, 120	n.s.	-	-
	Speed*	12.066	1, 40	**0.001**	0.232	-

Table 5.7: Inferential analysis of transformed error rate percentage
(* between participants factor)

When the data was analysed as a single group, the effect of music was significant, $F(3, 123) = 3.032$, $p = 0.032$, $\eta_p^2 = 0.069$, (Figure 5.11).

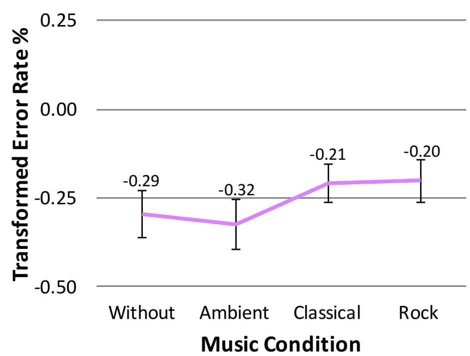


Figure 5.11: Significant omnibus effect of music ($p = 0.032$)

Post hoc pairwise comparisons were performed (Table 5.8). None of the comparisons reached the required *p* value for statistical significance using the Bonferroni adjustment.

Condition 1	Condition 2	Mean Difference	p
Without	Ambient	0.029	n.s.
Without	Classical	-0.088	n.s.
Without	Rock	-0.092	n.s.
Ambient	Classical	-0.117	0.020**
Ambient	Rock	-0.122	0.013**
Classical	Rock	0.005	n.s.

Table 5.8: Post hoc tests for omnibus effect of music
(significance at $p < 0.0083$, ** significant at the non-adjusted level)

At the non-adjusted significance level, the transformed error rate when accompanied by ambient music was lower than both classical music and rock music (ambient: $M = -0.324$, $SD = 0.45$; classical: $M = -0.207$, $SD = 0.35$; rock: $M = -0.203$, $SD = 0.38$).

With the speed classification included as a between participants factor the omnibus effect of music was significant in the mixed ANOVA, $F(3, 120) = 2.816$, $p = 0.042$, $\eta_p^2 = 0.066$, (Figure 5.11).

Post hoc pairwise comparisons were not performed as this result is weaker than when the analysis was performed using a single group, and the interaction between music and speed group was non-significant, $F(3, 120) = 1.825$, *n.s.*, which indicates that the speed classification does not interact with the effect of music on transformed error rate.

Speed classification was a significant between groups factor, $F(1, 40) = 12.066$, $p = 0.001$, $\eta_p^2 = 0.232$. The fast typists' transformed error rate was lower than the slow typists' (fast: $M = -0.428$, $SD = 0.36$; slow: $M = -0.092$, $SD = 0.38$).

5.2.2.2 Typing Experience

Typing experience was established by asking the participants to rate how difficult they found the task and how distracting they found the music.

Task Difficulty

Ratings of task difficulty were taken using a 7-point Likert item, ranging from 1 = extremely difficult to 7 = extremely easy. The mapping was reversed for analysis so that lower values indicate the task was perceived as easier. Table 5.9 presents descriptive statistics including the median (*Mdn*), lower and upper quartile (*Q1* and *Q3* respectively) and mean (*M*) for all typists. Table 5.10 includes separation into slow and fast groups.

Music Condition	All Typists (N = 42)			
	Q1	*Mdn*	*Q3*	*M*
Without	2.00	2.00	3.00	2.60
Ambient	2.00	3.00	4.00	2.79
Classical	2.00	3.00	4.00	3.29
Rock	2.75	3.00	4.25	3.55

Table 5.9: Descriptive statistics of task difficulty ratings for all typists

Music Condition	Slow Typists (n = 22)				Fast Typists (N = 20)			
	Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
Without	2.00	2.50	3.25	2.77	2.00	2.00	3.00	2.40
Ambient	2.00	3.00	4.00	3.05	2.00	2.00	3.00	2.50
Classical	2.00	3.00	5.00	3.32	2.00	3.00	4.00	3.25
Rock	2.00	3.00	4.25	3.50	3.00	4.00	4.75	3.60

Table 5.10: Descriptive statistics of task difficulty ratings by speed classification

The non-parametric Friedman test is the most appropriate to analyse the data when it is considered as a single group. The effect of music was significant, $\chi^2(3) = 14.550$, $p = 0.002$, (Figure 5.12).

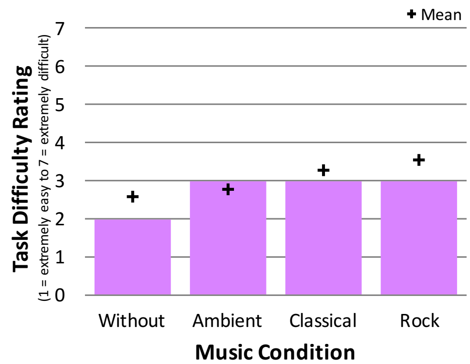


Figure 5.12: Significant main effect of music ($p = 0.002$)

Post hoc Wilcoxon signed rank tests were performed (Table 5.11).

Condition 1	Condition 2	z	p
Without	Ambient	1.338	n.s.
Without	Classical	2.862	**0.004**
Without	Rock	3.080	**0.002**
Ambient	Classical	2.238	0.025**
Ambient	Rock	2.554	0.011**
Classical	Rock	0.784	n.s.

Table 5.11: Post hoc tests for main effect of music
(significance at $p < 0.0083$, ** significant at the non-adjusted level)

At the Bonferroni adjusted level the rating of task difficulty was significantly lower without music than when accompanied by both the classical and rock music (without music: *Mdn* = 2.00; classical: *Mdn* = 3.00; rock: *Mdn* = 3.00).

The difference in task difficulty ratings between the ambient and classical music, and ambient and rock music, did not reach significance at the Bonferroni adjusted level. But, are potentially interesting as both comparisons were significant at the non-adjusted level. The median for all three of these conditions was 3.00. The quartiles for ambient and classical music were the same, but a lower mean in the ambient music condition suggests tasks accompanied by ambient music were perceived as easier than those with rock music (ambient: $Q1 = 2.00$, $Mdn = 3.00$, $Q3 = 4.00$, $M = 2.79$; classical: $Q1 = 2.00$, $Mdn = 3.00$, $Q3 = 4.00$, $M = 3.29$). The Q1 and Q3 values for rock music were lower than ambient music indicating that the tasks accompanied by ambient music were perceived as easier than those with rock music (ambient: $Q1 = 2.00$, $Mdn = 3.00$, $Q3 = 4.00$; rock: $Q1 = 2.75$, $Mdn = 3.00$, $Q3 = 4.25$).

To establish if there were differences in ratings due to the speed classification it was necessary to first use a mixed ANOVA test despite the non-continuous data as there is no non-parametric test which can accommodate a between participants factor, following the procedure described previously in Chapter 3, Section 3.5.3. Table 5.12 shows the results from this mixed ANOVA.

Test	F	df	p	η_p^2
Music	6.792	2.334, 93.376	**< 0.0005**	0.145
Music x Speed*	0.725	2.334, 93.376	n.s.	-
Speed*	0.747	1, 40	n.s.	-

Table 5.12: Inferential analysis of task difficulty ratings
(* between participants factor)

The main effect of music was significant, $F(2.334, 93.376) = 6.792$, $p < 0.0005$, $\eta_p^2 = 0.145$, which aligns well with the result from the non-parametric test without the between participants partition applied.

The interaction between music and speed group was not significant, $F(2.334, 93.376) = 0.725$, *n.s.*, and speed classification was not a significant between groups factor, $F(1, 40) = 0.747$, *n.s.* These results show the overall perception of task difficulty did not differ between the speed groups and the effect of music on task difficulty ratings was not affected by the speed classification.

Music Distraction

In the conditions where music accompanied the transcription task the participants were asked to rate how distracting they found the music using a 7-point Likert item ranging from 1 = extremely distracting to 7 = not at all distracting. The mapping of the scale was reversed for the analysis. Table 5.13 shows descriptive statistics of music distraction ratings for all typists and Table 5.14 includes separation into the slow and fast groups.

Music Condition	All Typists ($N = 42$)			
	Q1	*Mdn*	*Q3*	*M*
Ambient	1.00	2.00	4.00	2.52
Classical	2.00	3.00	5.00	3.33
Rock	4.00	5.00	5.00	4.29

Table 5.13: Descriptive statistics of music distraction ratings for all typists

Music Condition	Slow Typists (*n* = 22)				Fast Typists (*n* = 20)			
	Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
Ambient	1.00	2.50	4.25	2.82	1.00	1.50	3.00	2.20
Classical	2.00	2.00	5.00	3.09	2.00	4.00	5.00	3.60
Rock	3.00	5.00	5.00	4.27	4.00	4.50	5.00	4.30

Table 5.14: Descriptive statistics of music distraction ratings by speed classification

A Friedman test is the most appropriate test to analyse this data as a single group. The effect of music was significant, $\chi^2(2) = 23.739$, $p < 0.0005$, (Figure 5.13).

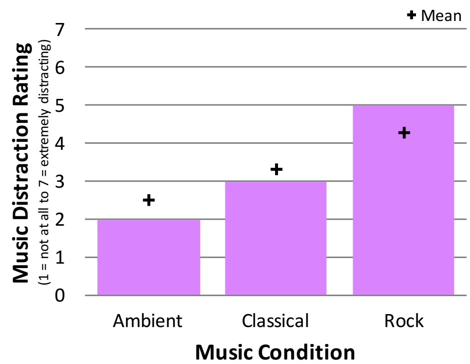


Figure 5.13: Significant main effect of music ($p < 0.0005$)

Post hoc pairwise comparisons were performed using the Wilcoxon signed rank test (Table 5.15).

Condition 1	Condition 2	z	p
Ambient	Classical	2.399	**0.016**
Ambient	Rock	4.454	**< 0.0005**
Classical	Rock	2.938	**0.003**

Table 5.15: Post hoc tests for main effect of music
(significance at $p < 0.0166$)

Statistically significant differences were revealed between all three pairs at the Bonferroni adjusted level. Ambient music was perceived as less distracting than both classical and rock music (ambient: *Mdn* = 2.00; classical: *Mdn* = 3.00; rock: *Mdn* = 5.00). Classical music was also perceived as less distracting than rock music (classical: *Mdn* = 3.00; rock: *Mdn* = 5.00).

To include speed classification as a between participants factor the analysis has to be performed using a mixed ANOVA. Table 5.16 presents the results of this mixed ANOVA.

Test	F	df	p	η_p^2
Music	17.214	2, 80	**< 0.0005**	0.301
Music x Speed*	1.741	2, 80	n.s.	-
Speed*	0.007	1, 40	n.s.	-

Table 5.16: Inferential analysis of music distraction ratings
(* between participants factor)

The main effect of music was significant, $F(2, 80) = 17.214$, $p < 0.0005$, $\eta_p^2 = 0.301$, again aligning well with the result from the non-parametric analysis without the speed classification applied. Speed classification was not a significant between groups factor, $F(1, 40) = 0.007$, *n.s.*, and the interaction between music and speed group was not significant, $F(2, 80) = 1.741$, *n.s.*

5.2.2.3 Summary of Analysis Using Paradigm 1

The results from the analysis of typing performance and experience using paradigm 1 are summarised in the next sections.

Typing Performance

In terms of CPM, the interaction between music and the speed group was significant. The simple main effect of music was significant for the fast typists but not the slow typists. The fast typists' CPM without music was lower than with both classical and rock music, and CPM with ambient music was lower than with rock music. There was also a suggestion that CPM was lower with classical than ambient music, however this latter comparison did not reach significance at the Bonferroni adjusted level. Speed classification was unsurprisingly a significant between groups factor with lower speeds achieved by the slow typists and higher speeds by the fast typists.

The omnibus effect of music was significant with the speed classification applied and showed a trend towards significance without the classification applied. Pairwise comparisons revealed CPM was significantly higher without music and lower when the transcription task was accompanied by rock music. All other pairwise comparisons when the CPM data was treated as a single group were non-significant.

With the error rate DV a logarithmic transformation needed to be applied before the inferential analysis could be performed. There was a significant omnibus effect of music both with and without the speed classification applied, with the effect strongest when the data was treated as a single group. However, all of the pairwise comparisons were non-significant at the Bonferroni adjusted level. At the non-adjusted level, the transformed error rate with ambient music was lower than with both classical and rock music. With the speed classification applied the pairwise comparisons also revealed potentially interesting differences when the transformed error rate without music was compared to the classical and rock music conditions. The transformed error rate without music was lower, although these pairwise comparisons were not significant, but p was less than 0.1. Speed classification was a significant between groups factor. The fast typists' transformed error rate was lower than the slow typists'.

Typing Experience

Classification as a slow or fast typist did not affect whether the task was perceived as difficult or whether the music was perceived as distracting. There were significant main effects of music for both measures. In terms of task difficulty, the tasks without music or accompanied by ambient music were perceived as easier than tasks where either classical or rock music was playing. While the ambient music was perceived as significantly less distracting than both the classical and rock music and the classical music was, in turn, perceived as significantly less distracting than the rock music.

5.2.3 Analysis Using Design Paradigm 2

Experiment design paradigm 2 focuses on how the effect of music differs between the simple and advanced English tasks. This group comprised 21 participants, 11 of whom had been classified as slow and 10 as fast typists.

5.2.3.1 Typing Performance

Again, typing performance was measured in CPM and error rate percentage. The distributions were assessed for normality and the presence of outliers using box plots and Shapiro-Wilk's tests for each distribution which can be found in Appendix E.1.2. Outliers that affected the results of the analysis were capped following the procedure described in Chapter 3, Section 3.5.1.3. The error rate data was transformed using logarithms following the procedure described in Chapter 3, Section 3.5.2.

Characters Per Minute

When the CPM data was treated as a single group none of the distributions deviated significantly from normality and there were no outliers. With the speed classification applied, none of the distributions deviated significantly from normality. However, an outlier was introduced, but as they did not affect the results, so the data was retained in the sample. Table 5.17 shows descriptive statistics of CPM.

Text Condition	Music Condition	All Typists (n = 21)		Slow Typists (n = 11)		Fast Typists (n = 10)	
		M (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)
Simple English	Without	349.95	94.85	271.13	48.60	436.64	37.71
	Ambient	349.76	98.83	270.44	58.36	437.00	42.16
	Classical	345.39	93.04	265.64	42.16	433.11	30.64
	Rock	346.17	90.27	270.38	46.32	429.53	31.36
Advanced English	Without	347.74	92.10	270.32	44.74	432.89	35.28
	Ambient	345.82	90.27	269.94	42.42	429.29	36.46
	Classical	341.92	88.79	267.60	44.20	423.67	34.21
	Rock	336.67	89.03	262.87	45.96	417.84	35.93

Table 5.17: Descriptive statistics of CPM

The data was first analysed as a single group, using a 2 by 4 repeated measures ANOVA. There were no assumption violations for this ANOVA. There were no significant effects or interactions, as shown in Table 5.18.

Test	F	df	p	η_p^2
Music	1.833	3, 60	n.s.	-
Text	2.493	1, 20	n.s.	-
Music x Text	0.703	3, 60	n.s.	-

Table 5.18: Inferential analysis of CPM

The speed classification was applied and a mixed 2 by 4 by 2 ANOVA performed. Aside from the outliers mentioned previously, there were no other assumption violations for mixed ANOVA. Table 5.19 presents the results from the mixed ANOVA.

Test	F	df	p	η_p^2
Music	1.838	3, 57	-	-
Music x Speed*	0.343	3, 57	-	-
Text	2.666	1, 19	-	-
Text x Speed*	1.133	1, 19	-	-
Music x Text	0.666	3, 57	-	-
Music x Text x Speed*	0.229	3, 57	-	-
Speed*	88.671	1, 19	**< 0.0005**	0.824

Table 5.19: Inferential analysis of CPM
(* between participants factor)

Speed classification was a significant between groups factor, $F(1, 19) = 88.671$, $p < 0.0005$, $\eta_p^2 = 0.824$. The fast typists' CPM was higher than the slow typists' (fast: $M = 430.00$, $SD = 35.47$; slow: $M = 268.54$, $SD = 46.59$). No other effects or interactions were significant.

Error Rate

The error rate data was transformed using logarithms before analysis. Both when treated as a single group and with the speed classification applied, a number of outlying data points were identified whose inclusion in the data did affect the outcomes of the analysis so were capped. After capping, in the single group analysis one distribution (12.5%) had a marginally significant deviation from normality ($p = 0.049$), but none of the 16 distributions deviated significantly from normality when the speed classification was applied. Table 5.20 presents descriptive statistics of transformed error rate percentage.

Text	Music Condition	All Typists (N = 21)		Slow Typists (n = 11)		Fast Typists (n = 10)	
		M (%)	SD (%)	M (%)	SD (%)	M (%)	SD (%)
Simple English	Without	-0.433	0.47	-0.218	0.35	-0.669	0.49
	Ambient	-0.477	0.53	-0.306	0.58	-0.665	0.41
	Classical	-0.239	0.33	-0.035	0.29	-0.523	0.40
	Rock	-0.303	0.28	-0.188	0.23	-0.417	0.29
Advanced English	Without	-0.325	0.40	-0.155	0.34	-0.512	0.38
	Ambient	-0.336	0.42	-0.203	0.30	-0.481	0.49
	Classical	-0.236	0.32	-0.141	0.29	-0.336	0.23
	Rock	-0.195	0.28	-0.032	0.16	-0.404	0.37

Table 5.20: Descriptive statistics of transformed error rate percentage

The analysis was first performed using a 2 by 4 repeated measures ANOVA treating the data as a single group. Aside from the one marginally non-normal distribution discussed previously, there were no other violations of assumptions for repeated measures ANOVA. Table 5.21 presents the results from this repeated measures ANOVA.

Test	F	df	p	η_p^2	Figure
Music	3.949	3, 60	**0.012**	0.165	5.14
Text	2.735	1, 20	n.s.	-	-
Music x Text	0.760	3, 60	n.s.	-	-

Table 5.21: Inferential analysis of transformed error rate percentage

Music had a significant effect on transformed error rate, $F(3, 60) = 3.949$, $p = 0.012$, $\eta_p^2 = 0.165$, (Figure 5.14).

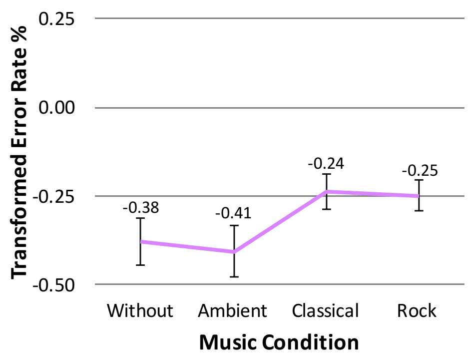


Figure 5.14: Significant omnibus effect of music ($p = 0.012$)

Post hoc pairwise comparison tests were performed (Table 5.22).

Condition 1	Condition 2	Mean Difference	p
Without	Ambient	0.28	n.s.
Without	Classical	-0.141	0.038**
Without	Rock	-0.130	0.072***
Ambient	Classical	-0.168	0.009**
Ambient	Rock	-0.157	0.043**
Classical	Rock	0.011	n.s.

Table 5.22: Post hoc tests for omnibus effect of music
(significance at $p < 0.0083$, ** significant at the non-adjusted level, *** not significant at Bonferroni or non-adjusted levels, but $p < 0.1$)

None of the pairwise comparisons were significant at the Bonferroni adjusted level. But, the difference in transformed error rate when accompanied by ambient music to classical music was extremely close to the required adjusted p value of 0.0083. The transformed error rate was lower with ambient music than classical (ambient: $M = -0.406$, $SD = 0.47$; classical: $M = -0.238$, $SD = 0.32$).

Using the non-adjusted level, the transformed error rate when typing without music was lower than when accompanied by classical music (without music: *M* = -0.379, *SD* = 0.43; classical: *M* = -0.238, *SD* = 0.32). Also, the transformed error rate when accompanied by ambient music was lower than with rock music (ambient: *M* = -0.406, *SD* = 0.47; rock: *M* = -0.249, *SD* = 0.28).

Finally, the difference in transformed error rate between the without music and rock conditions was not significant, but the resulting *p* value was less than 0.1 so is potentially interesting. The transformed error rate without music was lower than with rock music (without music: *M* = -0.379, *SD* = 0.43; rock: *M* = -0.249, *SD* = 0.28).

The analysis then included speed classification as a between groups factor. With the classification applied and outlier capping, none of the distributions deviated significantly from normality. However, there was a violation of the homogeneity of variance assumption for the advanced English task when accompanied by rock music, $F(1, 19) = 5.757$, $p = 0.027$. The group sizes were roughly equal (slow: $n = 11$; fast: $n = 10$), so the mixed ANOVA should be robust to this heteroscedasticity and the analysis can proceed. Table 5.23 presents the results from the mixed ANOVA.

Test	**F**	**df**	**p**	η_p^2
Music	3.482	3, 57	**0.022**	0.155
Music x Speed*	0.262	3, 57	n.s.	-
Text	2.966	1, 19	n.s.	-
Text x Speed*	0.547	1, 19	n.s.	-
Music x Text	0.337	3, 57	n.s.	-
Music x Text x Speed*	1.431	3, 57	n.s.	-
Speed*	9.572	1, 19	**0.006**	0.335

Table 5.23: Inferential analysis of transformed error rate percentage
(* between participants factor)

The effect of music was again significant, $F(3, 57)$, $p = 0.022$, $\eta_p^2 = 0.155$. Though, with the speed classification applied this omnibus effect is weaker with a smaller effect size than when the analysis was performed on the data treated as from a single group.

Speed classification was a significant between groups factor, $F(1, 19) = 9.572$, $p = 0.006$, $\eta_p^2 = 0.335$. The fast typists' error rate was lower than the slow typists' (fast: *M* = -0.501, *SD* = 0.160; slow: *M* = -0.160, *SD* = 0.33). As there were no significant effects or interactions which involved the music IV and the speed group, the analysis was not taken further as it would not lead to any further interesting outcomes beyond those that have already been identified.

5.2.3.2 Typing Experience

The experience of typing when accompanied by these different pieces of music in the simple and advanced English conditions was again measured in terms of the perception of task difficulty and how distracting the music was.

Task Difficulty

Ratings of task difficulty were taken using a 7-point Likert item ranging from 1 = extremely difficult to 7 = extremely easy. For analysis the mapping was reversed. Table 5.24 provides descriptive statistics for all typists with Table 5.25 incorporating the separation into the slow and fast groups.

Text Condition	Music Condition	All Typists (N = 21)			
		Q1	*Mdn*	*Q3*	*M*
Simple English	Without	1.50	2.00	3.50	2.52
	Ambient	2.00	3.00	4.00	2.86
	Classical	3.00	4.00	5.00	3.57
	Rock	2.00	3.00	5.00	3.24
Advanced English	Without	2.00	2.00	3.50	2.71
	Ambient	2.00	2.00	4.00	2.71
	Classical	2.50	3.00	5.00	3.38
	Rock	3.00	3.00	5.00	3.62

Table 5.24: Descriptive statistics of task difficulty ratings for all typists

Text Condition	Music Condition	Slow Typists (n = 11)				Fast Typists (n = 10)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
Simple English	Without	2.00	3.00	4.00	2.91	1.00	2.00	2.50	2.10
	Ambient	2.00	3.00	4.00	3.00	1.75	2.50	4.00	2.70
	Classical	3.00	4.00	4.00	3.36	2.75	4.00	5.00	3.80
	Rock	2.00	3.00	5.00	3.27	2.00	3.00	5.00	3.20
Advanced English	Without	2.00	2.00	3.00	2.64	2.00	2.50	4.00	2.80
	Ambient	2.00	3.00	4.00	2.91	2.00	2.00	3.25	2.50
	Classical	3.00	3.00	5.00	3.45	2.00	3.00	4.50	3.30
	Rock	2.00	3.00	4.00	3.36	3.00	4.00	5.00	3.90

Table 5.25: Descriptive statistics of task difficulty ratings by speed classification

Non-parametric tests cannot accommodate a 2 by 4 repeated measures design, so the analysis was first performed using a repeated measures ANOVA. Any significant effects or interactions were followed up with the most appropriate non-parametric test. Table 5.26 presents results from the repeated measures ANOVA.

Test	F	df	p	η_p^2
Music	5.505	3, 60	**0.002**	0.216
Text	0.101	1, 20	n.s.	-
Music x Text	2.351	3, 60	n.s.	-

Table 5.26: Inferential analysis of task difficulty ratings

The omnibus effect of music was significant, $F(3, 60) = 5.05$, $p = 0.002$, $\eta_p^2 = 0.216$. While, the effect of text was not significant, $F(1, 20) = 0.101$, *n.s.*, and neither was the interaction between music and text, $F(3, 60) = 2.351$, *n.s.* As there is no evidence that text is an important factor, the text separation was removed, and the analysis proceeded using a 1 by 4 Friedman test with the data from both text conditions combined. The effect of music was significant, $\chi^2(3) = 16.148$, $p = 0.001$, (Figure 5.15).

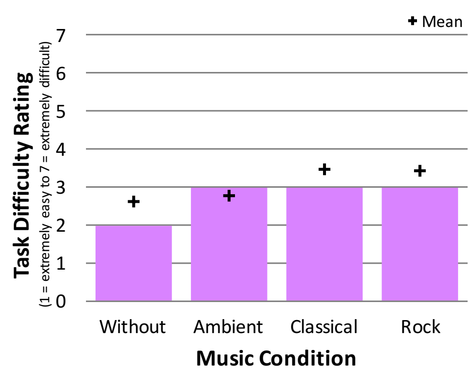


Figure 5.15: Significant main effect of music ($p = 0.001$)

Post hoc pairwise comparisons were performed using Wilcoxon signed rank tests (Table 5.27).

Condition 1	Condition 2	z	p
Without	Ambient	0.979	n.s.
Without	Classical	3.203	**0.001**
Without	Rock	2.535	0.011**
Ambient	Classical	3.145	**0.002**
Ambient	Rock	2.159	0.031**
Classical	Rock	0.344	n.s.

Table 5.27: Post hoc tests for main effect of music
(significance at $p < 0.0083$, ** significant at the non-adjusted level)

The pairwise comparisons revealed significant differences in task difficulty ratings at the Bonferroni adjusted level between the without music and classical music conditions (without music: *Mdn* = 2.00; classical: *Mdn* = 3.00) and between the ambient and classical music conditions (ambient: *Mdn* = 3.00; classical: *Mdn* = 3.00). For the latter pair, the medians are equal so do not provide sufficient information to establish which music condition led to significantly lower task difficulty ratings. But, inspection of the quartiles and means indicates the transcription typing tasks accompanied by ambient music were perceived as easier than those with classical music (ambient: $Q1$ = 2.00, $Q3$ = 4.00, M = 2.79; classical: $Q1$ = 2.00, $Q3$ = 5.00, M = 3.48).

Using a non-adjusted significant level, the differences in task difficulty ratings between the without music and rock music conditions were significant (without music: *Mdn* = 2.00; rock: *Mdn* = 3.00) and so was the difference in ratings between the ambient and rock music conditions (ambient: *Mdn* = 3.00; rock: *Mdn* = 3.00). Again, with the latter pair the medians are equal so to establish which has the lower task difficulty ratings the quartiles and means need to be inspected. Tasks accompanied by ambient music were generally rated as easier than tasks accompanied by rock music (ambient: $Q1$ = 2.00, $Q3$ = 4.00, M = 2.79; rock: $Q1$ = 2.00, $Q3$ = 5.00, M = 3.43).

To assess whether the speed classification affected the task difficulty ratings a mixed ANOVA was performed. Table 5.28 presents the results from this analysis.

Test	**F**	**df**	**p**	η_p^2
Music	5.573	3, 57	**0.002**	0.227
Music x Speed*	0.655	3, 57	n.s.	-
Text	0.115	1, 19	n.s.	-
Text x Speed*	0.334	1, 19	n.s.	-
Music x Text	0.741	3, 57	n.s.	-
Music x Text x Speed*	1.108	3, 57	n.s.	-
Speed*	0.048	1, 19	n.s.	-

Table 5.28: Inferential analysis of task difficulty ratings
(* between participants factor)

The omnibus effect of music was significant, $F(3, 57) = 5.573$, $p = 0.002$, $\eta_p^2 = 0.227$, which aligns with the analysis without the speed classification applied. Pairwise comparisons were not performed as the main effect of music on task difficulty ratings was investigated before. There were no other significant omnibus effects or interactions and speed classification was not a significant between groups factor, $F(1, 19) = 0.048$, *n.s.*

Music Distraction

To establish how distracting the different pieces of music were perceived as a 7-point Likert item ranging from 1 = extremely distracting to 7 = not at all distracting was used in all conditions where music accompanied the transcription tasks. For this analysis the mapping was reversed so that low values mapped to less distracting music. Descriptive statistics including the median, lower and upper quartile and mean for all typists is provided in Table 5.29 and with separation into slow and fast groups in Table 5.30 .

Text Condition	Music Condition	All Typists (N = 21)			
		Q1	*Mdn*	*Q3*	*M*
Simple English	Ambient	1.00	2.00	3.50	2.33
	Classical	2.50	3.00	5.00	3.43
	Rock	1.50	3.00	4.00	3.19
Advanced English	Ambient	1.00	2.00	4.00	2.24
	Classical	2.00	3.00	5.00	3.43
	Rock	3.50	5.00	5.00	4.29

Table 5.29: Descriptive statistics of music distraction ratings for all typists

Text Condition	Music Condition	Slow Typists (n = 11)				Fast Typists (n = 10)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
Simple English	Ambient	1.00	2.00	4.00	2.55	1.00	2.00	3.25	2.10
	Classical	1.00	3.00	4.00	2.91	2.75	5.00	5.00	4.00
	Rock	1.00	3.00	4.00	2.73	2.00	4.00	5.25	3.70
Advanced English	Ambient	1.00	2.00	4.00	2.27	1.00	1.50	4.00	2.20
	Classical	2.00	3.00	5.00	3.27	1.75	3.50	5.25	3.60
	Rock	3.00	4.00	5.00	4.00	4.00	5.00	6.00	4.60

Table 5.30: Descriptive statistics of task difficulty ratings by speed classification

A 2 by 3 repeated measures ANOVA was performed to explore the differences in music distraction ratings across the different texts and pieces of music. Table 5.31 presents the results of this analysis.

Test	F	df	p	η_p^2
Music	10.403	2, 40	**< 0.0005**	0.342
Text	2.763	1, 20	n.s.	-
Music x Text	3.955	2, 40	**0.027**	0.165

Table 5.31: Inferential analysis of music distraction ratings

The omnibus effect of music was significant, $F(2, 40) = 10.403$, $p < 0.0005$, $\eta_p^2 = 0.342$, and so was the interaction between music and text, $F(2, 40) = 3.955$, $p = 0.027$, $\eta_p^2 = 0.165$. However, the omnibus effect of text was not significant, $F(1, 20) = 2.764$, n.s.

It is not possible to replicate an interaction effect with a non-parametric test. But, to accommodate the fact that, when it comes to the distraction ratings, the text may have an effect the follow up non-parametric analysis looking at the effect of music overall was performed using Friedman tests on the two texts separately.

When typing in the simple English text conditions, the music had a significant effect on distraction ratings, $\chi^2(2) = 9.781$, $p = 0.008$, (Figure 5.16).

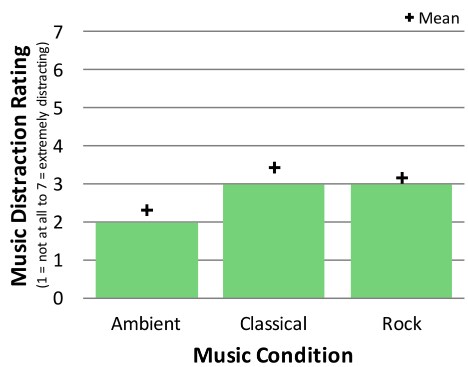


Figure 5.16: Significant main effect of music in simple English tasks ($p = 0.008$)

Post hoc pairwise comparisons using Wilcoxon signed rank tests were performed (Table 5.32).

Simple English Text			
Condition 1	**Condition 2**	**z**	**p**
Ambient	Classical	2.564	**0.010**
Ambient	Rock	2.229	0.026**
Classical	Rock	0.842	n.s.

Table 5.32: Post hoc tests for main effect of music in simple English tasks
(significance at $p < 0.0166$, ** significant at the non-adjusted level)

The pairwise comparisons revealed a significant difference between distraction ratings when accompanied by the ambient and classical music (ambient: *Mdn* = 2.00; classical: *Mdn* = 3.00). The difference in music distraction ratings between the ambient and rock music conditions was significant at the non-adjusted level, but not at the Bonferroni adjusted level (ambient: *Mdn* = 2.00; rock: *Mdn* = 3.00).

When using the advanced English text for the transcription tasks, the effect of music on distraction ratings was also significant, $\chi^2(2) = 15.881$, $p < 0.0005$, (Figure 5.17).

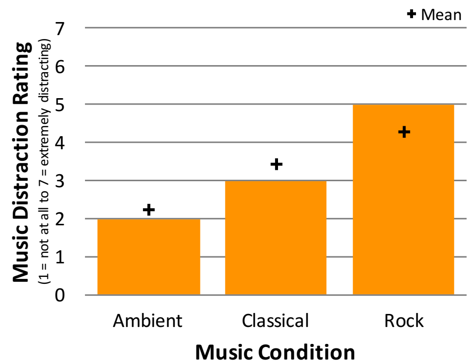


Figure 5.17: Significant main effect of music in advanced English tasks ($p < 0.0005$)

Post hoc pairwise comparisons using Wilcoxon signed rank tests were performed (Table 5.33).

Advanced English Text			
Condition 1	**Condition 2**	**z**	**p**
Ambient	Classical	2.520	**0.012**
Ambient	Rock	2.913	**0.004**
Classical	Rock	2.831	**0.014**

Table 5.33: Post hoc tests for main effect of music in advanced English tasks
(significance at $p < 0.0166$)

The music distraction ratings when accompanied by ambient music were significantly lower than both the classical and rock music ratings (ambient: *Mdn* = 2.00; classical: *Mdn* = 3.00; rock: *Mdn* = 5.00). The difference in music distraction ratings between the classical and rock music was also significant. The ratings with classical music were lower than when accompanied by rock music (classical: *Mdn* = 3.00; rock: *Mdn* = 5.00).

Speed classification was included as a between groups factor in a mixed ANOVA to explore whether speed group affected music distraction ratings. Table 5.34 presents the results.

Test	F	df	p	η_p^2
Music	11.049	2, 38	**< 0.0005**	0.368
Music x Speed*	1.546	2, 38	n.s.	-
Text	2.576	1, 19	n.s.	-
Text x Speed*	0.390	1, 19	n.s.	-
Music x Text	3.857	2, 38	**0.030**	0.169
Music x Text x Speed*	0.742	2, 38	n.s.	-
Speed*	0.776	1, 19	n.s.	-

Table 5.34: Inferential analysis of music distraction ratings
(* between participants factor)

The omnibus effect of music was significant, $F(2, 38) = 11.049$, $p < 0.0005$, $\eta_p^2 = 0.368$. The interaction between music and text was also significant, $F(2, 38) = 3.857$, $p = 0.030$, $\eta_p^2 = 0.169$. There were no other significant effects or interactions, including speed classification being a non-significant between participants factor, $F(1, 19) = 0.776$, *n.s.* As there were no additional significant effects or interactions and speed classification was not a significant between groups factor no further analysis was performed.

5.2.3.3 Summary of Analysis Using Paradigm 2

Typing Performance

The main effect of music was not significant for typing speed both with and without classification into fast and slow typist groups. The main effect of text was also not significant, however speed classification was a significant between groups factor.

In terms of error rate, the main effect of music was significant with both the speed classification applied and when the data was treated as a single group. The transformed error rate when accompanied by ambient music was lower than with classical music both with and without the speed classification applied, though, for the latter comparison this was only at the non-adjusted rather than Bonferroni adjusted level. Other pairwise comparisons were significant at the non-adjusted level both with and without the speed classification applied. The transformed error rate without music was lower than both classical music and rock music and the transformed error rate with ambient music was lower than with rock music accompanying the transcription typing task. The text did not have a significant effect on transformed error rate. Speed classification was a significant between groups factor. The participants who were categorised as fast typists achieved lower transformed error rates than those classified as slow typists.

Typing Experience

As with analysis paradigm 1, speed classification was not an important factor in the analysis of the two typing experience measures.

For task difficulty ratings the main effect of music was significant while the effect of text and the interaction between music and text was not significant. The experience of typing when accompanied by classical music was perceived as significantly harder than without music or with an ambient music accompaniment at the Bonferroni adjusted level. While the experience of typing when accompanied by rock music was harder than both an ambient music accompaniment or without music at the non-adjusted level.

For the music distraction ratings, the main effect of music was significant and so was the interaction between music and text, though the effect of text itself was not significant. The effect of music on distraction ratings was significant in the simple English transcription tasks. The ambient music was perceived as significantly less distracting than the classical music at the Bonferroni adjusted level, while it was also less distracting than rock music at the non-adjusted level. There was no difference in distraction ratings between the classical and rock music in the simple English tasks. The effect of music on distraction ratings was significant in the advanced English tasks as well. All three pairwise comparisons were significant at the Bonferroni adjusted level. The ambient music was perceived as least distracting and the rock music as most distracting. The classical music was perceived as both more distracting than the ambient music and less distracting than the rock music.

5.2.4 Analysis Using Design Paradigm 3

Experiment design paradigm 3 compares typing in the advanced English tasks to typing in Dutch language tasks. The classification process described in section 5.2.1 categorised 11 participants as slow typists and 10 as fast typists.

5.2.4.1 Typing Performance

Typing performance was again measured in CPM and error rate. Box plots and the results from Shapiro-Wilk's tests were used to assess for outliers and normality of distribution prior to inferential analysis which are included in Appendix E.1.1. Outliers that affected the results were capped following the procedure described in Chapter 3, Section 3.5.1.3.

The error rate data was transformed using logarithms following the procedure described in Chapter 3, Section 3.5.2.

Characters Per Minute

When the CPM data was treated as a single group there were no outliers or significant deviations from normality. With the speed classification applied, there were no significant deviations from normality. However, outliers were introduced which did affect the outcomes from the analysis, so the values were capped.

The separation of participants into fast and slow typist groups with outlier capping when the classification was applied, led to the descriptive statistics shown in Table 5.35.

Text Condition	Music Condition	All Typists (N = 21)		Slow Typists (n = 11)		Fast Typists (n = 10)	
		M (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)
Advanced English	Without	342.94	116.89	248.87	49.04	446.42	70.55
	Ambient	337.70	119.22	242.26	49.88	442.69	73.77
	Classical	339.54	103.31	257.33	45.34	429.98	63.93
	Rock	335.84	107.15	249.62	53.33	430.69	57.17
Dutch	Without	231.14	84.04	166.61	35.32	302.13	60.16
	Ambient	225.87	88.58	160.95	37.77	300.91	65.79
	Classical	226.24	86.38	160.53	23.40	300.44	66.31
	Rock	227.86	84.96	155.23	13.72	301.98	60.70

Table 5.35: Descriptive statistics of CPM

When performing the analysis as a single group, there was a violation of the assumption of sphericity for the effect of music, $W(5) = 0.477$, $p = 0.017$, so the Greenhouse-Geisser correction was applied. All other assumptions for repeated measures ANOVA were met. Table 5.36 shows the results from this analysis.

Test	F	df	p	η_p^2
Music	0.667	2.069, 41.373	n.s.	-
Text	233.561	1, 20	**< 0.0005**	0.921
Music x Text	0.521	3, 60	n.s.	-

Table 5.36: Inferential analysis of CPM

There was a significant effect of text, $F(1, 20) = 233.561$, $p < 0.0005$, $\eta_p^2 = 0.921$. CPM in the Dutch tasks was significantly lower than in the advanced English tasks (Dutch: M = 227.78, SD = 85.99; advanced English: M = 339.01, SD = 111.64).

The main effect of music was not significant, $F(2.069, 41.373) = 0.667$, *n.s.*, and neither was the interaction between music and text, $F(3, 60) = 0.521$, *n.s.*

With the speed classification applied there were a number of assumption violations for mixed ANOVA after the outlier capping process had been followed. The sphericity assumption was violated for music, $W(5) = 16.484$, $p = 0.006$, so the Greenhouse-Geisser correction was applied. The assumption of homogeneity of variance was violated in the Dutch text classical and rock music conditions. However, the group sizes are roughly equal (slow typists: $n = 11$; fast typists: $n = 10$) so the analysis can proceed as mixed ANOVA is robust to heteroscedasticity with equal group sizes. Table 5.37 shows the results from the mixed ANOVA.

Test	**F**	**df**	**p**	**η_p^2**	**Figure**
Music	1.301	1.836, 34.889	n.s.	-	-
Music x Speed*	1.494	1.836, 34.889	n.s.	-	-
Text	491.668	1, 19	**< 0.0005**	0.963	-
Text x Speed*	21.853	1, 19	**< 0.0005**	0.535	5.18
Music x Text	0.062	3, 57	n.s.	-	-
Music x Text x Speed*	4.242	3, 57	**0.009**	0.182	5.19
Speed*	54.660	1, 19	**< 0.0005**	0.742	-

Table 5.37: Inferential analysis of CPM
(* between participants factor)

The main effect of music was not significant, $F(1.836, 34.889) = 1.301$, *n.s.* However, the main effect of text was significant, $F(1, 19) = 491.668$, $p < 0.0005$, $\eta_p^2 = 0.963$. CPM was higher in the advanced English tasks than in the Dutch tasks (advanced English: $M = 339.01$, $SD = 111.64$; Dutch: $M = 227.78$, $SD = 85.99$).

Speed classification was a significant between groups factor, $F(1, 19) = 54.660$, $p < 0.0005$, $\eta_p^2 = 0.742$. The fast typists' CPM was significantly higher than the slow typists' (fast: $M = 369.41$, $SD = 64.80$; slow: $M = 205.17$, $SD = 38.47$).

The two-way interaction between text and speed group was significant, $F(1, 19) = 21.853$, $p < 0.0005$, $\eta_p^2 = 0.535$, (Figure 5.18).

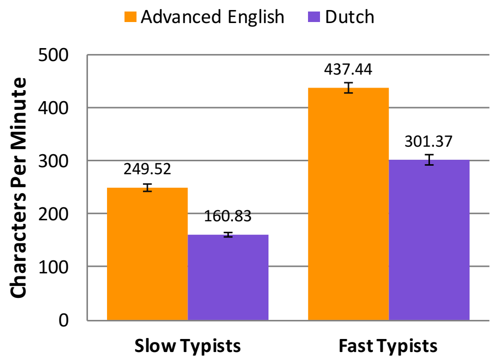


Figure 5.18: Significant interaction between text and speed group ($p < 0.0005$)

The simple main effect of text was significant for both fast, $F(1, 9) = 499.010$, $p < 0.0005$, $\eta_p^2 = 0.982$, and slow typists, $F(1, 10) = 125.651$, $p < 0.0005$, $\eta_p^2 = 0.926$. In both groups, CPM was higher in the advanced English tasks and lower in the Dutch conditions (slow typists, advanced English: $M = 249.52$, $SD = 49.40$; slow typists, Dutch: $M = 160.83$, $SD = 27.55$; fast typists, advanced English: $M = 473.44$, $SD = 66.36$; fast typists, Dutch: $M = 301.37$, $SD = 63.24$).

The three-way interaction between music, text and speed group was significant, $F(3, 57) = 4.242$, $p = 0.009$, $\eta_p^2 = 0.182$, (Figure 5.19).

The simple interaction between music and text was not significant for either the slow typists, $F(3, 30) = 2.157$, *n.s.*, (Figure 5.19, a and b), or the fast typists, $F(3, 27) = 2.164$, *n.s.*, (Figure 5.19, c and d).

Given the strength of the significant three-way interaction it is worth reconceptualising this to explore how the interaction between music and speed group differs between the two different texts, rather than how the interaction between music and text differs between the two speed groups.

The simple interaction between music and speed group was not significant in the Dutch tasks, $F(3, 57) = 0.704$, *n.s.*, (Figure 5.19, b and d), but was significant in the advanced English tasks, $F(3, 57) = 4.014$, $p = 0.012$, $\eta_p^2 = 0.174$, (Figure 5.19, a and c).

The simple simple main effect of music was significant for the fast typists, $F(3, 27) = 3.320$, $p = 0.035$, $\eta_p^2 = 0.270$, but was not significant for the slow typists, $F(1.953, 19.527) = 1.686$, *n.s.*

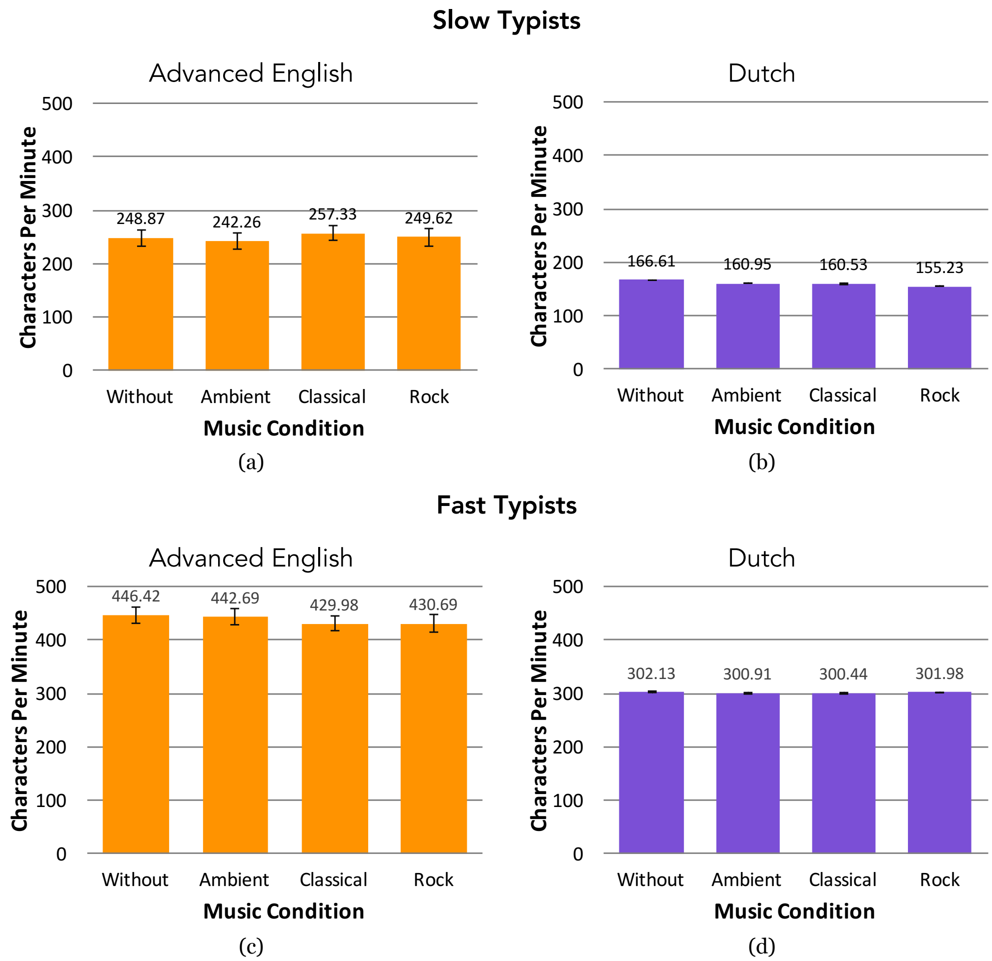


Figure 5.19: Significant interaction between music, text and typing speed group (p = 0.009)

Using the Bonferroni adjustment none of the pairwise comparisons for fast typists were significant, (Table 5.38).

Fast typists, advanced English text			
Condition 1	**Condition 2**	**Mean Difference**	**p**
Without	Ambient	3.733	n.s.
Without	Classical	16.444	0.026**
Without	Rock	15.733	0.040**
Ambient	Classical	12.711	0.067***
Ambient	Rock	12.000	n.s.
Classical	Rock	0.711	n.s.

Table 5.38: Post hoc tests for simple simple main effect of music for fast typists in the advanced English tasks
(significance at $p < 0.0083$, ** significant at the non-adjusted level, *** not significant at Bonferroni or non-adjusted levels, but $p < 0.1$)

Using a non-adjusted *p* level for significance, CPM without music was lower than when accompanied by either classical or rock music (without music: *M* = 446.42, *SD* = 70.55; classical: *M* = 429.98, *SD* = 63.93, rock: *M* = 430.69, *SD* = 57.17).

The differences in CPM between the without music and rock conditions was not significant at the non-adjusted level, but was less than 0.1 so may also be contributing to the overall significant simple simple main effect (ambient: *M* = 442.69, *SD* = 73.77; classical: *M* = 429.98, *SD* = 63.93).

Error Rate

A logarithmic transformation was applied to the error rate data. After transformation, when the data was considered as a single group, there were eight outlying data points, only one of which was in a task with the advanced English text. The outliers affected the results of the inferential analysis, so the outlier capping process was followed. After outlier capping, none of the distributions deviated significantly from normality. With the between groups speed classification applied there was a single outlier but none of the distributions deviated significantly from normality. This outlier again affected the results of the analysis so was capped. After capping, all 16 distributions retained their normality characteristics.

The descriptive statistics including outlier capping for transformed error rate percentage are shown in Table 5.39.

Text Condition	Music Condition	All Typists (N = 21)		Slow Typists (n = 11)		Fast Typists (n = 10)	
		M (%)	*SD* (%)	*M* (%)	*SD* (%)	*M* (%)	*SD* (%)
Advanced English	Without	-0.265	0.46	0.007	0.38	-0.564	0.33
	Ambient	-0.313	0.48	-0.095	0.64	-0.490	0.34
	Classical	-0.178	0.37	-0.053	0.37	-0.316	0.34
	Rock	-0.163	0.38	0.029	0.55	-0.313	0.24
Dutch	Without	-0.017	0.33	0.079	0.32	-0.122	0.33
	Ambient	-0.031	0.29	0.114	0.25	-0.191	0.25
	Classical	0.121	0.46	0.255	0.43	-0.076	0.36
	Rock	0.064	0.31	0.230	0.30	-0.118	0.21

Table 5.39: Descriptive statistics of transformed error rate percentage

The data was first analysed as a single group using a repeated measures ANOVA. There were no assumption violations. The results of this analysis are shown in Table 5.40.

Test	F	df	p	η_p^2	Figure
Music	3.646	3, 60	**0.017**	0.154	5.20
Text	19.189	1, 20	**< 0.0005**	0.490	-
Music x Text	0.201	3, 60	n.s.	-	-

Table 5.40: Inferential analysis of transformed error rate percentage

The omnibus effect of music was significant, $F(3, 60) = 3.646$, $p = 0.017$, $\eta_p^2 = 0.154$, (Figure 5.20).

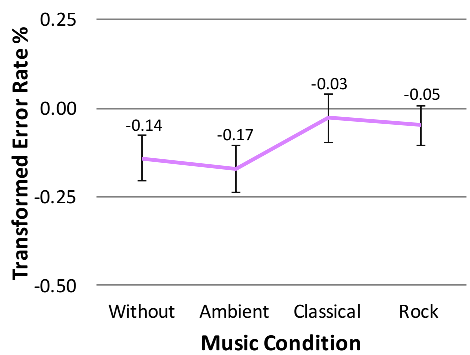


Figure 5.20: Significant omnibus effect of music ($p = 0.017$)

Pairwise comparisons were performed, (Table 5.41).

Condition 1	Condition 2	Mean Difference	p
Without	Ambient	0.031	n.s.
Without	Classical	-0.112	0.035**
Without	Rock	-0.091	0.097***
Ambient	Classical	-0.143	0.022**
Ambient	Rock	-0.123	**0.001**
Classical	Rock	0.021	n.s.

Table 5.41: Post hoc tests for omnibus effect of music
(significance at $p < 0.0083$, ** significant at the non-adjusted level, *** non-significant at Bonferroni and non-adjusted levels, but $p < 0.1$)

A significant difference in transformed error rates was revealed between the ambient and rock music conditions at the Bonferroni adjusted level. The transformed error rate when accompanied by ambient music was lower than with rock music (ambient: $M = -0.172$, $SD = 0.42$; rock: $M = -0.049$, $SD = 0.36$).

Further, using non-adjusted p level the transformed error rate without music was different to that when accompanied by classical music (without music: $M = -0.141$, $SD = 0.41$; classical: $M = -0.029$, $SD = 0.44$). The transformed error rate when accompanied by ambient music was also lower than the classical music condition (ambient: M = -0.172, SD = 0.42; classical: M = -0.029, SD = 0.44).

The difference in transformed error rate between the without music and rock music conditions was not significant at either the Bonferroni or non-adjusted level, but $p < 0.1$ so may indicate a difference. The transformed error rate without music was lower than with rock music (without music: M = -0.141, SD = 0.41; rock: M = -0.049, SD = 0.36).

With the speed classification applied as a between participants factor, there was a violation of the assumption of homogeneity of variance in the Dutch text rock music condition, $F(1, 19) = 6.282$, $p = 0.021$. The group sizes were balanced so the impact of this heteroscedasticity violation should be small. There were no other assumption violations for mixed ANOVA. Table 5.42 presents the results from this analysis.

Test	F	df	p	η_p^2
Music	3.063	3, 57	**0.035**	0.139
Music x Speed*	0.244	3, 57	n.s.	-
Text	14.353	1, 19	**0.001**	0.430
Text x Speed*	0.557	1, 19	n.s.	-
Music x Text	0.210	3, 57	n.s.	-
Music x Text x Speed*	1.783	3, 57	n.s.	-
Speed*	8.381	1, 19	**0.009**	0.306

Table 5.42: Inferential analysis of transformed error rate percentage
(* between participants factor)

The omnibus effect of music was significant, $F(3. 57) = 3.063$, $p = 0.035$, $\eta_p^2 = 0.139$. Again, with the speed classification applied the omnibus effect is weaker than when the data was treated as a single group. The omnibus effect of text was significant, $F(1, 19) = 14.353$, $p = 0.001$, $\eta_p^2 = 0.430$, but the effect is weaker than when the data was treated as a single group. Therefore, neither of these results were explored further.

Speed classification was a significant between groups factor, $F(1, 19) = 8.381$, $p = 0.009$, $\eta_p^2 = 0.306$. The fast typists made significantly fewer errors than the slow typists (fast: M = -0.274, SD = 0.34; slow: M = 0.071, SD = 0.42).

5.2.4.2 Typing Experience

Again, typing experience was measured quantitatively by measuring perceptions of task difficulty and music distraction.

Task Difficulty

Task difficulty was measured using 7-point Likert items ranging from 1 = extremely difficult to 7 = extremely easy for each of the transcription typing tasks. For this analysis the mapping was reversed so that low values indicate tasks that were perceived as easier. Table 5.43 presents descriptive statistics for all typists and Table 5.44 includes separation into the slow and fast groups.

Text Condition	Music Condition	All Typists (N = 21)			
		Q1	*Mdn*	*Q3*	*M*
Advanced English	Without	2.00	2.00	3.00	2.48
	Ambient	2.00	3.00	3.50	2.86
	Classical	2.00	3.00	4.00	3.19
	Rock	2.00	3.00	4.00	3.48
Dutch	Without	5.00	5.00	6.00	5.43
	Ambient	5.00	5.00	6.00	5.19
	Classical	4.50	5.00	6.50	5.38
	Rock	4.00	6.00	6.00	5.38

Table 5.43: Descriptive statistics of task difficulty ratings for all typists

Text Condition	Music Condition	Slow Typists (n = 11)				Fast Typists (n = 10)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
Advanced English	Without	2.00	3.00	4.00	2.91	2.00	2.00	2.00	2.00
	Ambient	2.00	3.00	4.00	3.18	2.00	2.50	3.00	2.50
	Classical	2.00	3.00	4.00	3.18	2.00	3.00	4.00	3.20
	Rock	2.00	3.00	5.00	3.64	2.00	3.50	4.00	3.30
Dutch	Without	5.00	5.00	6.00	5.27	5.00	5.50	6.00	5.60
	Ambient	5.00	5.00	6.00	5.09	5.00	5.50	6.60	5.30
	Classical	4.00	5.00	6.00	5.27	4.75	6.00	7.00	5.50
	Rock	4.00	5.00	6.00	5.27	4.00	6.00	7.00	5.50

Table 5.44: Descriptive statistics of task difficulty ratings by speed classification

To accommodate the design a 2 by 4 repeated measures ANOVA was performed. The results from this analysis are included in Table 5.45.

Test	F	df	p	η_p^2
Music	2.936	3, 60	**0.040**	0.128
Text	100.178	1, 20	**< 0.0005**	0.834
Music x Text	2.919	3, 60	**0.041**	0.127

Table 5.45: Inferential analysis of task difficulty ratings

The omnibus effect of music was significant, $F(3, 60) = 2.936$, $p = 0.040$, $\eta_p^2 = 0.128$, and so was the omnibus effect of text, $F(1, 20) = 100.189$, $p < 0.0005$, $\eta_p^2 = 0.834$. The interaction between music and text was also significant, $F(3, 60) = 2.919$, $p = 0.041$, $\eta_p^2 = 0.127$. This analysis clearly indicates that the text is an important factor in terms of the task difficulty. So, two 1 by 4 Friedman tests were performed on the task difficulty data separated into the two different texts.

In the advanced English tasks music had a significant effect on task difficulty ratings, $\chi^2(3) = 10.550$, $p = 0.014$, (Figure 5.21). However, in the Dutch tasks music did not have a significant effect on task difficulty ratings, $\chi^2(3) = 2.265$, *n.s.*

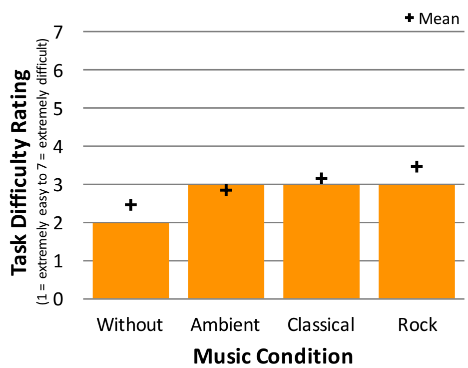


Figure 5.21: Significant main effect of music in advanced English tasks ($p = 0.014$)

Post hoc pairwise comparisons were performed for the advanced English tasks using Wilcoxon signed rank tests, (Table 5.46).

At the Bonferroni adjusted level none of the pairwise comparisons were significantly different. However, using the non-adjusted *p* level there were differences in the task difficulty ratings between the without music condition and all three music conditions (without music: *Mdn* = 2.00; ambient: *Mdn* = 3.00; classical: *Mdn* = 3.00; rock: *Mdn* = 3.00).

Advanced English Text			
Condition 1	**Condition 2**	**z**	**p**
Without	Ambient	2.000	0.046**
Without	Classical	2.430	0.015**
Without	Rock	2.610	0.009**
Ambient	Classical	1.126	n.s.
Ambient	Rock	1.883	0.060***
Classical	Rock	0.818	n.s.

Table 5.46: Post hoc tests for main effect of music in advanced English tasks
(significance at $p < 0.0083$, ** significant at the non-adjusted level, *** not significant at Bonferroni or non-adjusted levels, but $p < 0.1$)

To be able to include the speed classification as a between participants factor a mixed ANOVA was performed. Table 5.47 shows the results from this mixed ANOVA.

Test	F	df	p	η_p^2
Music	2.949	3, 57	**0.040**	0.134
Music x Speed*	0.679	3, 57	n.s.	-
Text	109.421	1, 19	**< 0.0005**	0.852
Text x Speed*	2.576	1, 19	n.s.	-
Music x Text	2.979	3, 57	**0.039**	0.136
Music x Text x Speed*	0.711	3, 57	n.s.	-
Speed*	0.104	1, 19	n.s.	-

Table 5.47: Inferential analysis of task difficulty ratings
(* between participants factor)

There was a significant omnibus effect of music, $F(3, 57) = 2.949$, $p = 0.040$, $\eta_p^2 = 0.134$, as well as a significant omnibus effect for text, $F(1, 19) = 109.421$, $p < 0.0005$, $\eta_p^2 = 0.852$. The interaction between music and text was significant, $F(3, 57) = 2.979$, $p = 0.039$, $\eta_p^2 = 0.136$. These significant effects and interactions align with those from the analysis without the between participants speed factor. As speed classification is not a significant between groups factor, $F(1, 19) = 0.104$, *n.s.*, and there are no significant effects or interactions involving the speed factor no further analysis was performed.

Music Distraction

Ratings of music distraction were taken using 7-point Likert items ranging from 1 = extremely distracting to 7 = not at all distracting for all of the transcription typing tasks where music was played. For the analysis this mapping was reversed. Table 5.48 presents descriptive statistics for all typists are shown and Table 5.49 includes separation into slow and fast groups.

Text Condition	Music Condition	All Typists (N = 21)			
		Q1	*Mdn*	*Q3*	*M*
Advanced English	Ambient	1.00	3.00	4.00	2.81
	Classical	2.00	2.00	5.00	3.24
	Rock	3.50	5.00	5.00	4.29
Dutch	Ambient	1.50	2.00	3.00	2.57
	Classical	3.00	3.00	5.00	3.62
	Rock	2.00	3.00	5.00	3.48

Table 5.48: Descriptive statistics of music distraction ratings for all typists

Text Condition	Music Condition	Slow Typists (n = 11)				Fast Typists (n = 10)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
Advanced English	Ambient	2.00	3.00	5.00	3.36	1.00	2.00	3.00	2.20
	Classical	2.00	2.00	5.00	2.91	2.00	4.00	5.00	3.60
	Rock	3.00	5.00	6.00	4.55	3.75	4.00	5.00	4.00
Dutch	Ambient	3.00	3.00	3.00	2.82	1.00	2.00	3.25	2.30
	Classical	3.00	3.00	3.00	3.27	2.75	4.00	6.00	4.00
	Rock	2.00	3.00	4.00	3.18	2.00	4.00	5.00	3.80

Table 5.49: Descriptive statistics of music distraction ratings by speed classification

A 2 by 3 repeated measures ANOVA was performed on ratings of music distraction. Table 5.50 presents the results of this analysis.

Test	F	df	p	η_p^2
Music	15.167	2, 40	**0.005**	0.235
Text	1.207	1, 20	n.s.	-
Music x Text	3.221	2, 40	0.050**	0.139

Table 5.50: Inferential analysis of music distraction ratings
(** non-significant outcome that is <0.1 and may indicate a trend towards significance)

The omnibus effect of music was significant, $F(2, 40) = 15.167$, $p = 0.005$, $\eta_p^2 = 0.235$. Although strictly speaking the interaction between music and text is non-significant, $F(2, 40) = 3.221$, $p = 0.050$, $\eta_p^2 = 0.139$, this result is extremely close to the required value. So, the analysis proceeded by considering the effect of music on the two texts separately.

In the advanced English tasks, the effect of music was significant using a 1 by 3 Friedman test, $\chi^2(2) = 11.606$, $p = 0.003$, (Figure 5.22).

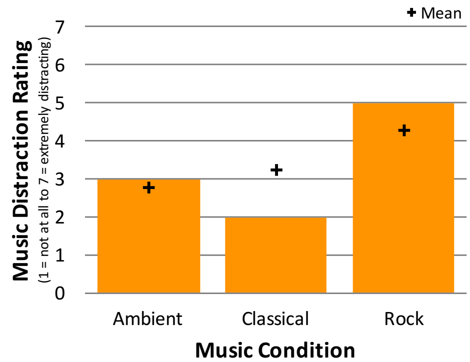


Figure 5.22: Significant main effect of music in advanced English tasks (p = 0.014)

Post hoc pairwise comparisons were performed using Wilcoxon signed rank tests, (Table 5.51).

Advanced English Text			
Condition 1	**Condition 2**	**z**	**p**
Ambient	Classical	0.933	n.s.
Ambient	Rock	3.322	**0.001**
Classical	Rock	2.477	**0.013**

Table 5.51: Post hoc tests for main effect of music in advanced English tasks
(significance at p < 0.0166)

The rock music was perceived as significantly more distracting than both the classical and ambient music (rock: *Mdn* = 5.00; classical: *Mdn* = 2.00; ambient: *Mdn* = 3.00).

In the Dutch tasks, there was a trend towards a significant effect of music, $\chi^2(2) = 5.765$, p = 0.056, (Figure 5.23).

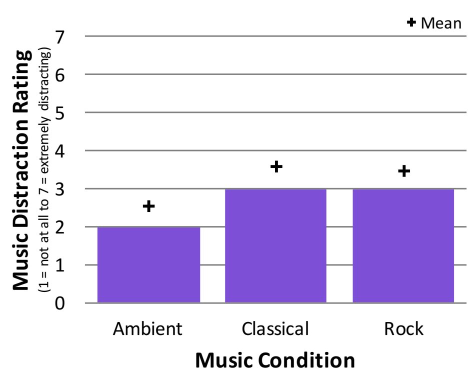


Figure 5.23: Trend towards a significant main effect of music in Dutch tasks (p = 0.056)

Post hoc pairwise comparisons were performed using Wilcoxon signed rank tests, (Table 5.52).

Dutch Text			
Condition 1	**Condition 2**	**z**	**p**
Ambient	Classical	2.198	0.028**
Ambient	Rock	1.989	0.035**
Classical	Rock	0.534	n.s.

Table 5.52: Post hoc tests for main effect of music in Dutch tasks
(significance at p < 0.0166, ** significant at the non-adjusted level)

None of the pairwise comparisons were significant at the Bonferroni adjusted level. Using a non-adjusted level, the ambient music was perceived as less distracting than both the classical and rock music (ambient: *Mdn* = 2.00; classical: *Mdn* = 3.00; rock: *Mdn* = 3.00).

To assess whether the speed classification affected the music distraction ratings a mixed ANOVA was performed. Table 5.53 shows the results of this mixed ANOVA.

Test	F	df	p	η_p^2
Music	6.969	2, 38	**0.003**	0.268
Music x Speed*	2.796	2, 38	0.074**	0.128
Text	1.128	1, 19	n.s.	-
Text x Speed*	2.477	1, 19	n.s.	-
Music x Text	3.026	2, 38	0.060**	0.137
Music x Text x Speed*	0.711	2, 38	n.s.	-
Speed*	0.006	1, 19	n.s.	-

Table 5.53: Inferential analysis of music distraction ratings
(* between participants factor)

With the speed classification applied, the omnibus effect of music was still significant, $F(2, 38) = 6.969$, $p = 0.003$, $\eta_p^2 = 0.268$, but there was also a trend towards a significant interaction between music and speed group, $F(2, 38) = 2.796$, $p = 0.074$, $\eta_p^2 = 0.128$. Speed classification itself was not a significant between groups factor, $F(1, 19) = 0.006$, *n.s.*, but the significant interaction indicates that the speed classification should be considered in this analysis. So, the analysis proceeded by looking at the two groups of typists separately, but collapsing the data from the two text conditions into a single task condition.

Using 1 by 3 Friedman tests, the effect of music was not significant for slow typists, $\chi^2(2) = 3.855$, *n.s.*, but was significant for fast typists, $\chi^2(2) = 13.571$, $p = 0.001$, (Figure 5.24).

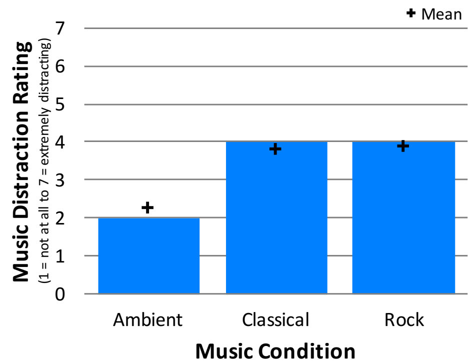


Figure 5.24: Significant main effect of music for fast typists (p = 0.001)

Post hoc pairwise comparisons were performed using Wilcoxon signed rank tests, (Table 5.54).

Fast Typists			
Condition 1	**Condition 2**	**z**	**p**
Ambient	Classical	3.245	**0.001**
Ambient	Rock	2.810	**0.005**
Classical	Rock	0.290	n.s.

Table 5.54: Post hoc tests main effect of music for fast typists
(significance at p < 0.0166)

At the Bonferroni adjusted level ambient music was significantly less distracting than both classical and rock music (ambient: *Mdn* = 2.00; classical: *Mdn* = 4.00; rock: *Mdn* = 4.00).

The interaction between music and text showed a trend towards significance, $F(2, 38)$ = 3.026, p = 0.060, η_p^2 = 0.137. As this interaction does not include the speed classification and is weaker than the interaction when the data was analysed as a single group, this analysis was not taken further.

5.2.4.3 Summary of Analysis Using Paradigm 3

Typing Performance

The omnibus effect of music was not significant for CPM, both with and without the speed classification applied. The omnibus effect of text was significant for CPM with the significantly lower values in the Dutch transcription tasks and higher values in the advanced English tasks.

With respect to the transformed error rate, the omnibus effect of music was extremely close to reaching significance (p = 0.050) when the data was considered as a single group.

None of the pairwise comparisons were significant at the Bonferroni adjusted level, but with the non-adjusted level the transformed error rate was lower without music than with classical music. Also, the transformed error rate when accompanied by ambient music was lower than both the classical and rock music. With the speed classification applied the omnibus effect of music was not significant and there were no significant interactions. The omnibus effect of text was significant with lower transformed error rates in the advanced English conditions than the Dutch tasks.

Typing Experience

Speed classification was not an important factor in the analysis of task difficulty, but did have a slight effect on the results from the music distraction analysis.

The omnibus effect of music was significant for the task difficulty ratings, the effect of text was significant and so was the interaction between music and text. So, the analysis proceeded with separation into the two task groups.

With the advanced English tasks, the choice of music had a significant effect on the task difficulty ratings. At the Bonferroni adjusted level none of the pairwise comparisons were significant. However, at the non-adjusted level the task difficulty ratings in the without music condition were lower than all of the music conditions.

When typing in Dutch, the effect of music on task difficulty ratings was not significant.

In terms of music distraction ratings, the omnibus effect of music was significant and the interaction between text and music was extremely close to significance ($p = 0.050$). So, again, the analysis proceeded by considering the effect of music for both of the texts separately. The overall effect of text, however, was not significant.

With the advanced English text, the effect of music was significant. Both the ambient and classical music were perceived as significantly less distracting than the rock music. However, there was no difference in perceived distraction ratings between the ambient and classical music.

For the Dutch transcription tasks, there was a trend towards a significant effect of music. At the Bonferroni adjusted level none of the pairwise comparisons were significant. But, at the non-adjusted level the ambient music was less distracting than both the rock and classical music.

5.3 Discussion

The analysis of each experiment paradigm is complex, but some themes emerge from the results. First, when it comes to typing performance the speed classification affects the results for CPM. Although there was a significant difference in the transformed error rates between the slow and fast typists meaning that the classification can be a good predictor of the overall error rate, the effect of music on error rate was not dependent on the classification. Further, speed classification was not important in the analysis of either of the experience measures indicating that the ratings of how difficult the tasks are, or how distracting the music is, did not depend on the typists' speed classification.

Another important theme is that, loosely speaking, the without music and ambient conditions were similar. Typically, these two conditions were rated as easier and had better performance when compared to the rock and classical music conditions. This potentially reflects the different characteristics of the pieces of music, with the classical and rock music having prominent melodies, strong beats and fast tempos. The ambient music had repetitive tones, but there was no obvious melody to follow and although it would be possible to tap out the beat for this piece of music, it is not a dominant feature of the music. It is perhaps possible that completing the transcription tasks when accompanied by ambient music can actually be considered quite similar to a without music condition. In fact, there were some analyses where there was a significant difference in the pairwise comparison between the ambient music and classical or rock music, where there was no significant difference between the without music and classical or rock music. This may indicate that the ambient music may lead to an improvement in performance over the without music condition.

When the analysis focused on the advanced English text only across the 42 participants typing speed classification was important. A significant omnibus effect of music on CPM was achieved, but so was a significant interaction between music and speed group. When this interaction is considered, the effect of music was significant for the fast typists but not for the slow. This result can be explained in one of three ways. One interpretation is that the effect for the fast typists is so strong that the overall effect is reached despite the music not actually affecting the slow typists' CPM. Another explanation is that some of the slow typists are experiencing the same effect as the fast typists, but not all of them. This means the overall significant omnibus effect of music is enhanced, but the slow typists when taken as a separate group do not experience the same overall effect of music. Another explanation is that there is an overall effect for slow typists, but the effect size is smaller than for the fast typists, leading to the significant interaction. But, there is insufficient power in the slow group for the effect to emerge in the analysis of simple main effects.

With the participants split into the two separate task groups, the outcomes differed quite considerably from the analysis of the advanced English text conditions only, which in itself is interesting and warrants further investigation.

The effect of music for the 21 participants who completed the simple and advanced English transcription tasks was non-significant for CPM both with and without the speed classification applied. This contrasts with the analysis of the advanced English tasks across the 42 participants where the effect of music was significant. It is likely that this is again an issue of power.

In terms of error rate, the omnibus effect of music across the simple and advanced English text conditions was significant which aligns with the analysis across the 42 participants when the advanced English tasks were considered separately. The transformed error rate when accompanied by ambient music was lower than with classical music at the Bonferroni level. At the non-adjusted level, the ambient music also had lower error rates than the rock music and the without music conditions led to lower error rates than both classical and rock music. Again, with slightly more power in the experiment all of these pairwise comparisons may have been significant at the Bonferroni adjusted level.

In terms of the experience of typing, the analysis of this group of 21 participants aligns well with the analysis from the advanced English text conditions across the 42 participants.

In this analysis the text did not affect the ratings of task difficulty indicating the simple English text was not an easier transcription typing task than the advanced English text and that the intended manipulation of task difficulty here may not have been successful. However, the overall effect of music on task difficulty was significant. The transcription typing tasks that were accompanied by classical music were perceived as harder than both the ambient music and without music conditions. Although not significant at the Bonferroni level, the transcription typing tasks accompanied by rock music were perceived as harder than the ambient or without music conditions at the non-adjusted level. This outcome seems to align with the error rate analysis, with higher error rates associated with the more difficult tasks.

The analysis of music distraction ratings resulted in a significant interaction between text and music and a main effect of music. In the simple English conditions, the ambient music was significantly less distracting than the classical music at the Bonferroni adjusted level. The ambient music was also less distracting than rock music at the non-adjusted level. The difference in distraction ratings between classical and rock music were not significant. While in the advanced English tasks the differences in distraction ratings between ambient and classical, and ambient and rock music were significant at the Bonferroni adjusted level, with lower distraction ratings for ambient music. Again, the difference between classical and rock music distraction ratings was not significant. Although the interaction between music and text was significant, the results for both texts were similar.

Moving onto the analysis of the paradigm comparing the advanced English and Dutch texts. The main effect of music was not significant but there was a significant two-way interaction between music and speed group and a significant three-way interaction between music, text and speed group.

The significant three-way interaction should be interpreted first, and with most emphasis as it indicates that the two-way interaction between music and text should be different across the two speed groups. There is some evidence of a difference in interactions, as the simple interaction between music and text was not significant for the fast typists but approached significance for the slow typists. However, the simple simple main effect of music was not significant for either the advanced English or Dutch texts. So, it is difficult to properly interpret and understand this result meaning that I must look to the significant two-way interaction between music and speed group for further information. The simple effects analysis of this two-way interaction resulted in a significant main effect of music for the fast typists but not the slow. However, none of the pairwise comparisons were significant so again it is unclear what is causing this simple main effect to be significant. This analysis is somewhat frustrating as it is clear that something is happening between the effect of music on CPM across the two different texts and speed groups, but it is not possible to identify what is going on. This, again, is likely to be an issue with power as when data was used from the 42 participants together, rather than this subsection of 21, the effect of music on CPM was clearer.

With regard to the error rate analysis, there was almost a significant effect of music when the data was treated as a single group (p = 0.050) but with the between participants speed partition applied this effect disappears. Using non-adjusted p values with pairwise comparisons the differences in error rates revealed lower transformed error rates with ambient music than both classical and rock music, and lower transformed error rates without music than when accompanied by classical music. Although not significant at the Bonferroni adjusted level, these results suggest why the omnibus effect came so close to the required p value for significance.

The speed classification did affect transformed error rate, but did not interact with the music IV. This result is similar to that identified in the other paradigms, with the effect of music in terms of transformed error rate not dependent on the speed classification.

The experience analysis again identified a significant effect of music on task difficulty ratings. Further, there was a significant interaction between music and text, so the task difficulty analysis proceeded considering the effects on each text separately. For the advanced English text, the music had a significant effect on the task difficulty ratings. None of the pairwise comparisons were significant at the Bonferroni adjusted level, but at the non-adjusted level the without music condition was perceived as easier than the ambient, classical and rock music conditions. This is a particularly interesting result as it is the first that suggests that perhaps the typing tasks when accompanied by ambient music may be slightly harder than the without music conditions. Although this result was not replicated when the advanced English text was considered as a single group, so from this experiment as a whole there is insufficient evidence to make that claim. And, in most other regards, the ambient music and without music conditions behaved very similarly so this marginally significant pairwise comparison at the non-adjusted level may simply be a clear justification for why adjustments are made for multiple pairwise comparisons! For the Dutch transcription tasks, there was no effect of music on task difficulty ratings. This shows that the Dutch transcription tasks were so hard there was either no effect of music on task difficulty ratings, or if there was an effect it was so small it could not be identified.

Finally, the analysis of perceived music distraction ratings led to a significant main effect of music. The interaction between music and text was again extremely close to significance (p = 0.050) so the analysis proceeded considering both texts separately. For the advanced English tasks, there was a significant difference in distraction ratings across the three music conditions. Pairwise comparisons revealed that rock music was perceived as significantly more distracting than both the ambient and classical music. Though, there was no difference identified in distraction ratings between the classical and ambient music conditions. With the Dutch transcription tasks, the effect of music on distraction ratings was not significant, but was again close (p = 0.056). The pairwise comparisons suggested that the ambient music may have been perceived as less distracting than the classical and rock music, but only at the non-adjusted level.

5.3.1 Discussion of Each Experiment Aim

The initial aims of this experiment were:

1. to investigate if the genre of music affects typing performance and experience, including a comparison to a without music condition,
2. to investigate whether task difficulty has a moderating effect on typing performance and experience,
3. to gather further evidence that the effect of music on typing performance and experience differs between slow and fast typists.

The results presented in the previous section are discussed in more detail in relation to the specific experiment aims in the following sections.

5.3.1.1 Experiment Aim 1

The first experiment aim of the experiment was to compare transcription typing performance when accompanied by different pieces of music to a without music condition. The objective was to explore whether music improved or reduced performance or experience over the without music condition, and if any effects due to music were consistent between different genres.

This aim has been successfully achieved. There is sufficient evidence from this experiment to say that this piece of ambient music affected typing performance and experience differently to both the classical and rock music. There is insufficient evidence to categorically claim the ambient and without music conditions are the same, but there is a clear indication that the effect of ambient music on typing performance and experience is closer to the without music condition than either the rock or classical music. There is also insufficient evidence to claim that rock and classical music have exactly the same effects on performance and experience, but when the results of the experiment are considered as a whole this seems likely.

5.3.1.2 Experiment Aim 2

The second aim of the experiment focused on identifying whether the effects of music on typing performance and experience were the same across different task difficulties, as manipulated by changing the source of the presented text. The task difficulty manipulations themselves were somewhat successful, but may not have solved the ceiling effect problems identified in Experiment 1. Instead, the task difficulty manipulation has raised more questions which certainly require further investigation. One clear outcome is that the simple and advanced English texts were not drastically different in terms of difficulty. The task difficulty analysis itself did not identify a significant effect of text between this pair, and the outcomes from the rest of the analysis were not contingent on the text.

On the other hand, the comparison between the advanced English and Dutch tasks was successful, in that the Dutch transcription tasks were rated significantly harder than the advanced English tasks and led to a significant three-way interaction between music, text and speed classification for CPM. Unfortunately, the post hoc analysis could not explain this three-way interaction, so follow up work is needed here, especially as there were hints of differences in distraction ratings for the different pieces of music even with the Dutch transcription typing tasks.

In the discussion for Experiment 1 I hypothesized that the reason the slow typists did not seem to be affected by the music in the same manner as the fast typists was due to the less skilled slow typists needing to concentrate more on the task. However, if this was the case, it should follow that in the hard Dutch transcription tasks, which do require high levels of concentration, the participants would not be able to distinguish between distraction ratings for the different pieces of music. That said, it is clear the manipulation of task difficulty using Dutch language worked and that it is an interesting task condition to take further in this research.

5.3.1.3 Experiment Aim 3

The final aim of this experiment involved exploring whether the slow and fast typist classification was an important factor. When it comes to typing speed, this experiment has provided evidence that the effect of music does seem to differ across the two groups. However, with regard to the performance in terms of error rate and the experience measures the speed classification did not affect the results. Although the faster typists made fewer errors, the effect of music did not differ between the two speed groups. This outcome is particularly interesting given neither this experiment nor Experiment 1 were able to provide evidence that the slow typists performance, in terms of speed, was affected by the music. However, this experiment has shown that their performance in terms of accuracy was affected in the same way as the fast typists, or at least not sufficiently differently to lead to a significant interaction. This outcome can be explained in two ways. Either, there is higher power in the error rate measure, so in terms of speed there may be an effect for the slow typists caused by the music, but it is smaller than the effect for fast typists and, potentially, so small that it cannot be identified even with 24 participants classified as slow in this experiment. Alternatively, this result could show that speed and accuracy of typing are not linked, and that the effect of music is different depending on whether it is the slow or fast typists and whether it is the speed or accuracy measure being considered. Again, these are avenues that require further investigation through more experiments.

5.3.2 Limitations of This Experiment

The main limitation of this experiment relates to the power and numbers of participants assigned to each group. In the advanced English only analysis, which considered all 42 participants as a single group, the power was sufficient to identify omnibus effects and pairwise comparisons at the Bonferroni adjusted level. However, in the two other analysis paradigms with the split into 21 participants in each group, there was insufficient power to properly and clearly identify and interpret effects, particularly with the post hoc tests.

5.3.3 Moving Forwards

It is clear from this experiment that more participants are needed. Also, there is something interesting in how the effect of music differs between the advanced English and Dutch texts and the speed classification. However, it is not clear what this effect is. Therefore, the next experiment needs a larger sample using both the advanced English and Dutch transcription typing tasks.

Chapter 6

Vocals, Volume and Task Difficulty 1

Experiment 3

This third experiment compared the effect of vocals, volume and task difficulty on transcription typing performance and experience using a classroom, rather than laboratory, based methodology. This approach allowed me to perform an experiment with at least 50 participants in 4 hours, which is much more efficient than the laboratory methodology. However, the preapartion time for a classroom based experiment is higher than a typical laboratory based study because the level of experimenter control is significantly reduced in a classroom environment. So, to ensure validity of the method, a thorough pilot testing process was used (which was discussed previously in Chapter 3, Section 3.4.2).

The vocals IV was included in this experiment because there was evidence from Experiment 1 that this is an interesting dimension to pursue further but more data was needed. The with and without vocals versions of the alt rock style music from Experiment 1 were used as the music stimuli in this experiment. The alt rock style music was chosen over pop rock style because the results from the Experiment 1 suggested that vocals may have had a significant effect on performance in the pop rock music condition. I did not want to constrain the generalisations of the outcomes from these experiments to a single style of music and as there was no clear evidence that vocals in the alt rock style music had negative effects it is more challenging and, potentially, more interesting to use this piece as the stimuli.

Volume was included as an IV in this experiment for three reasons. First, the qualitative comments from Experiment 1 described the volume of music as a cause of distraction. Second, loudness has been shown to affect people's performance in various studies, (e.g. in Edworthy & Waring, 2006). Finally, running an experiment where the volume has been controlled will help me to establish whether an online experiment would be appropriate in future experiments. Performing an experiment using an online methodology would allow me to gather a much larger sample of data, but I would be unable to control the playback volume used by the participants. Even if I asked all participants to set the volume to maximum on their computers the variability in actual playback volume levels is likely to be large as they would be using their own equipment to participate in the experiment. If the results from this experiment show volume does not affect performance or experience when typing in a controlled environment, this will provide evidence that volume does not need to be controlled and an online study would be appropriate. However, if volume is shown to be an important factor in the results then careful consideration of how to accommodate the lack of control of volume would be needed before any future online study could be used to collect data from a large number of participants.

Task difficulty was also included as IV so that the experiment compares the impact of the two music IVs on tasks that require different levels of concentration because this was shown to be an interesting avenue in Experiment 2.

6.1 METHOD

6.1.1 AIMS

The aims of this experiment were:

1. to gather further evidence that the presence of vocals in a piece of music has a negative impact on typing performance and experience,
2. to investigate how the volume of music affects typing performance and experience,
3. to investigate if the previously identified effect of vocals is consistent across music played at different volume levels,

4. to see whether an increase in task difficulty reduces the effect of the music IVs on typing performance and experience,
5. to gather further evidence that the effect of music on typing performance and experience differs between slow and fast typists,
6. to validate the method is appropriate for a classroom based methodology.

6.1.2 Participants and Experimenters

Both the participants and experimenters were first year undergraduates studying Computer Science at the University of York, taking a module entitled *Human Aspects of Computer Science*. Fifty five students acted as participants in the experiment (8 female, 47 male), with a further 55 students acting as the experimenter. All participants were in the 18 to 24 age group, though it is likely that most were towards the lower end of the age group as they were all first year undergraduates.

Nine participants were non-native speakers of English, but all had demonstrated competency in English by achieving International English Language Test (IELT) scores in excess of 6.5 as a requirement for their acceptance onto the degree programme. None of the participants in this experiment were familiar with the Dutch language. The participants were not asked to report whether they had a hearing disability or dyslexia as this would have been inappropriate given the classroom context. After the experiment, five participants were randomly selected to win £10 Amazon gift vouchers. The three fastest typists received £30, £20 and £10 Amazon gift vouchers. These vouchers were offered in order to encourage the participants to try during the experiment and take it seriously.

6.1.3 Design

The experiment used a fully factorial two-way mixed design with repeated measures. The between groups IV was volume (two levels: low and high). The repeated measures IVs were vocals (two levels: without and with vocals) and task difficulty (two levels: medium and hard difficulty). The two task difficulty conditions were manipulated by changing the language of the text. The medium difficulty tasks used the advanced English text with the hard difficulty tasks using the Dutch text, as described in the Presented Text IVs section, in Chapter 3, Section 3.3.1.1. So, each participant was assigned to either the low or high volume group and completed four transcription typing tasks:

- with music that contained vocals in English (vocals: with; task difficulty: medium),
- with music that did not contain vocals in English (vocals: without; task difficulty: medium),
- with music that contained vocals in Dutch (vocals: with; task difficulty: hard),
- with music that did not contain vocals in Dutch (vocals: without; task difficulty: hard).

As this experiment used the classroom based methodology there was no without music condition because the environment could not be controlled sufficiently to eliminate potential confounds. Figure 6.1 shows the arrangement of IVs used in the experiment.

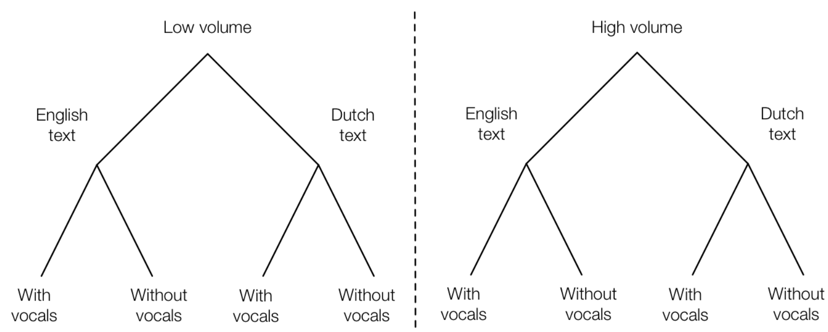


Figure 6.1: Structure of IVs for experiment

To avoid practice and fatigue effects, the order of presentation of tasks was counterbalanced.

The objective dependent variables (DVs) in this experiment were typing performance measured in both speed (CPM) and accuracy (error rate percentage). The subjective DVs of typing experience were measured using 7-point Likert items to collect ratings of perceived task difficulty and levels of music distraction.

6.1.4 Equipment and Materials

The experiment was performed using the classroom based methodology with bespoke websites created to collect the data and to guide and control the experiment. A thorough piloting process had been performed to ensure the data collected in this classroom based methodology was safe, which was described previously in Chapter 3, Section 3.4.2.

The experiment took place in an undergraduate PC lab in the Department of Computer Science at the University of York.

Participants used a desktop PC running Windows 7 with Firefox version 29. Inexpensive headphones (Astro Tools ATA 1144) were given to all participants.

This experiment used the with and without vocals versions of *Atrophy* (The Doppler Shift, n.d.) as the stimuli, which was the alt rock style music used in Experiment 1 (see Music Independent Variables section, in in Chapter 3, Section 3.3.1.1 for more details on how these versions were created). The presented text used the advanced English and Dutch material (see the Presented Text IVs section in Chapter 3, Section 3.3.1.1 for more details).

A short paper based questionnaire was distributed in order to collect subjective ratings of task difficulty and music distraction (see Appendix B.4). The participants also completed a demographics questionnaire (see Appendix B.5).

6.1.5 Procedure

All participants were given an ID number. Participants with an odd number were assigned to the high volume group, while participants with an even number completed the tasks listening to the low volume music. The procedure followed exactly that described in in Chapter 3, Section 3.3.3.2. The participants all completed four 4.5 minute transcription tasks so were typing for 18 minutes during the experiment.

6.2 Results

The first two experiments highlighted the importance of the speed classification in the analysis. This results section begins by following the classification process using the CPM data and then presents the analysis of typing performance and experience data.

6.2.1 Speed Classification

As in the previous experiment chapters, histograms and scatterplots of the CPM data were used to perform the speed classification.

The CPM data from the two volume groups was combined to create a consolidated histogram for the English transcription typing tasks (Figure 6.2). Approximations of two normal distributions have been overlaid on the histogram.

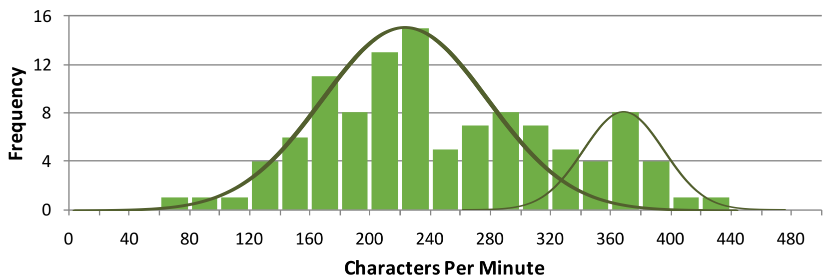


Figure 6.2: Histogram of CPM for English tasks
N = 110, *M* = 248.48, *SD* = 79.18

This histogram does not contain the clear gap between distributions that was evident in Experiment 1. There is a clear second peak in the distribution between 360 and 380 CPM. However, there is also a drop at 240 CPM which is not accommodated well by the bimodal explanation. It might be a trimodal distribution where a classification of slow, mid and fast typists would be more appropriate than just the bimodal groups for slow and fast typists. To explore this further, a scatterplot of CPM in the advanced English tasks was created (Figure 6.3). The scatterplot includes two possible thresholds at 250 and 330 CPM.

Visual inspection of the scatterplot shows using a value of 330 CPM is appropriate for the fast typists as all the participants in that group achieved speeds above 330 CPM in both conditions. But, is there a group of mid speed typists as well? Using a lower threshold of 250 CPM and an upper limit of 330 CPM would mean the typists in the central region, between the two threshold lines, would be classified as mid speed typists. However, this classification would introduce a number of problems, as there are four participants, (Participants P9, 51, 41 and 55 who are highlighted in the plot), who achieved speeds above 250 CPM in one condition, but below 250 CPM in the other. As the classification of mid speed typists is unclear, instead of forcing a trimodal classification onto the distribution, a simple bimodal grouping of slow and fast typists was used instead.

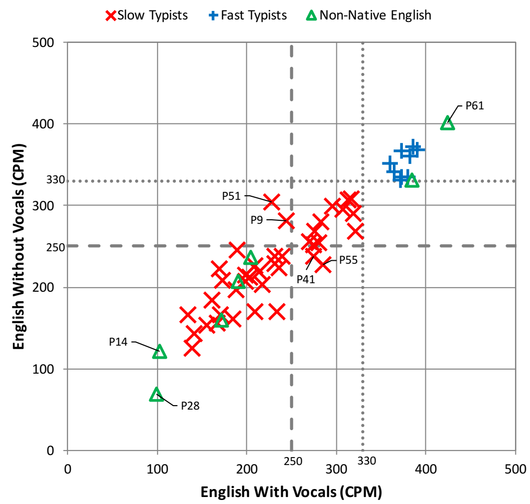


Figure 6.3: Scatterplot of CPM in advanced English tasks

Participants 14 and 28, both non-native English speakers, are also highlighted in the scatterplot as they were the slowest typists in the sample. These participants were removed from the data before the inferential analysis as their Dutch and English typing speeds were similar implying that both languages were unfamiliar to them. The other non-native speakers of English achieved lower speeds in Dutch than English making them suitable for inclusion in the analysis. Participant 61, a non-native English speaker, was the fastest typist in this experiment.

The proportion of fast typists in this sample was smaller than in Experiments 1 and 2, with just 11 out of 55 participants classified as fast typists. Six of these 11 participants were assigned to the high volume condition. Of the 42 slow participants remaining after participants 14 and 28 were removed from the sample, 21 had been assigned to the high volume condition.

6.2.2 Typing Performance

Following the speed classification, the analysis of typing performance can proceed, both with and without the classification applied. Typing performance was established by measuring CPM and error rate percentage. Before the inferential analysis was performed, the distributions were analysed for normality using the Shapiro-Wilk's tests and reviewing the *z* scores for skewness and kurtosis. The distributions were also inspected for outliers using box plots. The results from these normality tests and the box plots are included in

Appendix F.1. Outliers which affected the results of the analysis were capped following the procedure described in Chapter 3, Section 3.5.1.3.

6.2.2.1 Characters Per Minute

With the participants treated as two groups, without the speed classification applied, three of the eight CPM distributions (37.5%) were strongly non-normal ($p \leq 0.016$). The Dutch task without vocals distribution for the high volume group was of particular concern as it was highly leptokurtic. However, the group sizes are large enough and balanced (low volume: $n = 26$; high volume: $n = 27$) so the violations of normality of distribution can be overlooked. With the participants classified as fast or slow typists just 1 of the 16 conditions (6.25%) had a moderately non-normal distribution ($p = 0.015$).

A number of participants were outliers in at least one condition both when treated as two volume groups and with the speed classification applied as well. The inferential analysis was performed both with and without the outliers included in the data. The outliers did not affect the results of the inferential analysis so were retained in the data.

The separation of participants into fast and slow typist groups, together with the removal of participants 14 and 28, led to the descriptive statistics shown in Tables 6.1 and 6.2, which includes the data from the outliers.

Low Volume							
Text	**Vocals**	**All Typists** ($n = 26$)		**Slow Typists** ($n = 21$)		**Fast Typists** ($n = 5$)	
		M (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)
English	Without	248.78	81.57	218.89	58.47	374.31	10.26
	With	245.49	74.62	218.94	55.38	356.98	14.93
Dutch	Without	164.33	54.85	142.63	32.24	255.47	27.48
	With	165.06	59.28	140.80	32.02	266.98	27.66

Table 6.1: Descriptive statistics of CPM for low volume group

High Volume								
Text	Vocals	All Typists (n = 27)		Slow Typists (n = 21)		Fast Typists (n = 6)	
		M (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)
English	Without	265.26	79.74	230.84	50.60	385.74	20.70
	With	256.58	65.35	229.51	43.46	351.33	28.33
Dutch	Without	161.14	38.25	149.10	25.22	203.26	48.23
	With	160.01	35.25	153.56	28.08	182.59	50.21

Table 6.2: Descriptive statistics for CPM for high volume group

A fully factorial mixed ANOVA was performed with volume as a single between participants IV. There were two within participants IVs, the presence of vocals and presented text language.

The assumption of homogeneity of variance was not met. The difference in variance between the low and high volume groups in the Dutch text with vocals conditions was significant by Levene's test, $F(1, 51) = 5.856$, $p = 0.019$. However, as group sizes are balanced the ANOVA is robust to this heteroscedasticity. The outliers had no effect on the outcomes of the analysis and all other assumptions for mixed ANOVA were met. Table 6.3 presents the results from this analysis.

Test	F	df	p	η_p^2
Vocals	1.601	1, 51	n.s.	-
Vocals x Volume*	0.551	1, 51	n.s.	-
Text	200.677	1, 51	**< 0.0005**	0.797
Text x Volume*	1.929	1, 51	n.s.	-
Text x Vocals	1.346	1, 51	n.s.	-
Text x Vocals x Volume*	0.125	1, 51	n.s.	-
Volume*	0.094	1, 51	n.s.	-

Table 6.3: Inferential analysis of CPM
(* between groups factor)

The text had a significant effect on CPM, $F\,(1, 51) = 200.677$, $p < 0.0005$, $\eta_p^2 = 0.797$. CPM was higher when typing in English and lower when typing in Dutch (English: $M = 254.16$, $SD = 74.84$; Dutch: $M = 162.60$, $SD = 47.20$). There were no other significant effects or interactions.

The second analysis added speed classification (slow or fast) as an additional between participants factor. Again, the assumption of homogeneity of variance was not met, with violations occurring this time in both of the English text conditions (with vocals: $F(3, 49)$

= 3.435, p = 0.024; without vocals: $F(3, 49) = 4.928$, p = 0.005). The classification into speed groups led to unequal group sizes (slow typists, low volume: n = 21; slow typists, high volume: n = 21; fast typists, low volume: n = 5; fast typists, high volume: n = 6) so this assumption violation requires further investigation.

When sample sizes are different, heterogeneity of error variance is problematic if the larger variance is associated with the smaller group as the resulting F statistic is inflated leading to a higher chance of a type I error due to a liberal test (Tabachnick & Fidell, 2007). If the larger variance occurs in the larger group, the F statistic is conservative, with risk of a type II error. Inspection of the standard deviations shown in Tables 6.1 and 6.2 reveals that in the English task conditions, the smaller *SD*s (and therefore, by extension the variances as calculated by squaring the *SD*) were associated with the fast group of typists which had fewer participants than the slow typists group. As such, the F statistic is at risk of being conservative rather than liberal, so the analysis can proceed. Inclusion of the outlier participants in the analysis did not affect the outcomes and all other assumptions for a mixed ANOVA were met.

Table 6.4 presents the results from this analysis.

Test	F	df	p	η_p^2	Figure
Vocals	7.746	1, 49	**0.008**	0.137	-
Vocals x Volume*	4.296	1, 49	**0.043**	0.081	6.4
Vocals x Speed*	8.467	1, 49	**0.005**	0.147	6.5
Vocals x Volume* x Speed*	6.410	1, 49	**0.015**	0.116	6.10
Text	306.715	1, 49	**< 0.0005**	0.862	-
Text x Speed*	24.797	1, 49	**< 0.0005**	0.336	6.6
Text x Volume*	8.558	1, 49	**0.005**	0.149	6.7
Text x Vocals	3.630	1, 49	0.061**	0.069	6.8
Text x Volume* x Speed*	7.799	1, 49	**0.007**	0.137	6.11
Text x Vocals x Speed*	2.516	1, 49	n.s.	-	-
Text x Vocals x Volume*	0.093	1, 49	n.s.	-	-
Text x Vocals x Volume* x Speed*	0.871	1, 49	n.s.	-	-
Volume*	0.905	1, 49	n.s.	-	-
Speed*	90.805	1, 49	**< 0.0005**	0.650	-
Volume* x Speed*	3.395	1, 49	0.071**	0.065	6.9

Table 6.4: Inferential analysis of CPM with speed classification
(* between groups factor, ** non-significant outcome where p < 0.1)

The omnibus effect of the presence of vocals on CPM was significant, $F(1, 49) = 7.746$, $p = 0.008$, $\eta_\rho^2 = 0.137$. CPM was higher with music that did not contain vocals and was lower with music that did contain vocals (without vocals: $M = 209.04$, $SD = 80.78$; with vocals: $M = 206.82$, $SD = 74.36$).

The language of the presented text had a significant omnibus effect on CPM, $F(1, 49) = 306.715$, $p < 0.0005$, $\eta_\rho^2 = 0.862$. CPM in the English text conditions was higher than in the Dutch conditions (English: $M = 254.16$, $SD = 74.84$; Dutch: $M = 162.60$, $SD = 47.20$).

The volume of music did not have significant effect on CPM, $F(1, 49) = 0.905$, *n.s.*

The two-way interaction between vocals and volume was significant, $F(1, 49) = 4.296$, $p = 0.043$, $\eta_\rho^2 = 0.081$, (Figure 6.4).

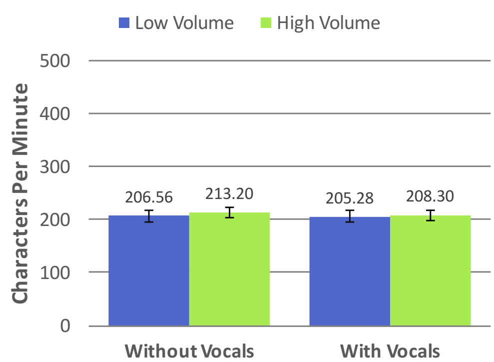


Figure 6.4: Significant interaction between vocals and volume ($p = 0.043$)

The simple main effect of vocals was significant for high volume music, $F(1, 25) = 7.711$, $p = 0.010$, $\eta_\rho^2 = 0.236$, but not for low volume music, $F(1, 24) = 0.724$, *n.s.* CPM was lower when the high volume music contained vocals and higher without vocals in the music (without vocals: $M = 213.20$, $SD = 59.00$; with vocals: $M = 208.30$, $SD = 50.30$).

The simple main effect of volume was not significant for either the without vocals, $F(1, 49) = 0.207$, *n.s.*, or with vocals, $F(1, 49) = 2.024$, *n.s.*, conditions.

The two-way interaction between vocals and speed classification was significant, $F(1, 49) = 8.467$, $p = 0.005$, $\eta_\rho^2 = 0.147$, (Figure 6.5).

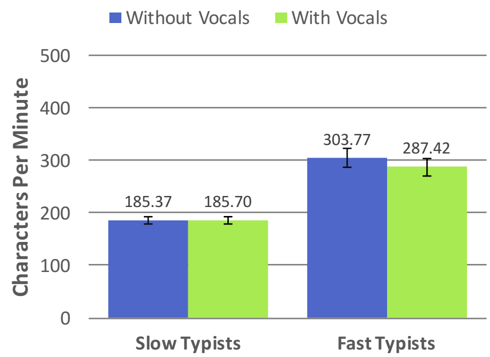


Figure 6.5: Significant interaction between vocals and speed group (p = 0.005)

The simple main effect of vocals was significant for fast, $F(1, 9) = 18.234$, $p = 0.002$, $\eta_\rho^2 = 0.670$, but not for slow, $F(1, 40) = 0.018$, *n.s.*, typists. The fast typists' CPM was higher when typing accompanied by music without vocals and lower when the music did contain vocals (without vocals: $M = 303.77$, $SD = 85.83$; with vocals: $M = 287.42$; $SD = 80.85$).

The two-way interaction between text and speed classification was significant, $F(1, 49) = 24.797$, $p < 0.0005$, $\eta_\rho^2 = 0.336$, (Figure 6.6).

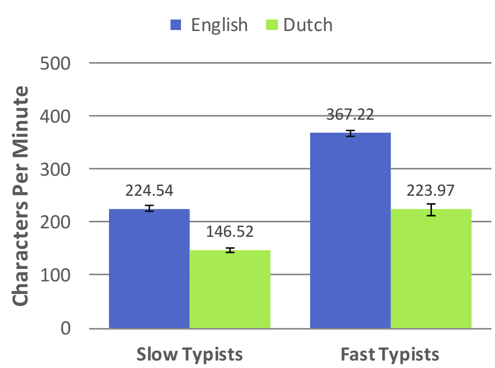


Figure 6.6: Significant interaction between text and speed group (p < 0.0005)

The simple main effect text was significant for both the slow, $F(1, 42) = 205.308$, $p < 0.0005$, $\eta_\rho^2 = 0.830$, and fast typists, $F(1, 9) = 87.320$, $p < 0.0005$, $\eta_\rho^2 = 0.907$. The slow typists' CPM was 78 CPM higher when typing in English than in Dutch (English: $M = 224.54$; Dutch: $M = 146.52$). The fast typists' CPM was 143 CPM higher when typing in English than Dutch (English: $M = 367.22$; Dutch: $M = 223.97$). The difference in CPM between the English and Dutch texts was much smaller for the slow typists than the fast, which explains why the two-way interaction between text and speed classification was significant when both the simple main effects were strongly significant with large effect sizes.

The two-way interaction between text and volume was significant, $F(1, 49) = 8.558$, $p = 0.005$, $\eta_\rho^2 = 0.149$, (Figure 6.7).

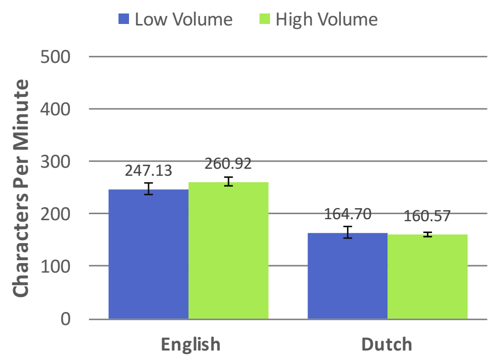


Figure 6.7: Significant interaction between text and volume ($p = 0.005$)

The simple main effect of volume as a between subjects factor was significant in Dutch, $F(1, 49) = 8.098$, $p = 0.006$, $\eta_\rho^2 = 0.142$, but not in English, $F(1, 49) = 0.204$, *n.s.* When typing in Dutch, CPM was higher when accompanied by low volume music and lower with high volume music (low volume: $M = 164.70$, $SD = 56.55$; high volume: $M = 160.57$, $SD = 36.44$).

The two-way interaction between text and vocals approached significance, $F(1, 49) = 3.630$, $p = 0.061$, $\eta_\rho^2 = 0.069$ (Figure 6.8).

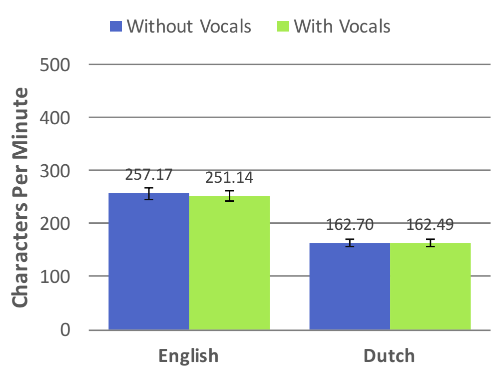


Figure 6.8: Trend towards significant interaction between text and vocals ($p = 0.061$)

The simple main effect of vocals was significant when typing in English, $F(1, 49) = 8.252$, $p = 0.006$, $\eta_\rho^2 = 0.144$, but not in Dutch, $F(1, 49) = 0.230$, *n.s.* In the English text conditions, CPM was higher when accompanied by music without vocals and lower when the music contained vocals (without vocals: $M = 257.17$, $SD = 80.30$; with vocals: $M = 251.14$, $SD = 69.59$).

The two-way interaction between volume and speed group also approached significance (Figure 6.9).

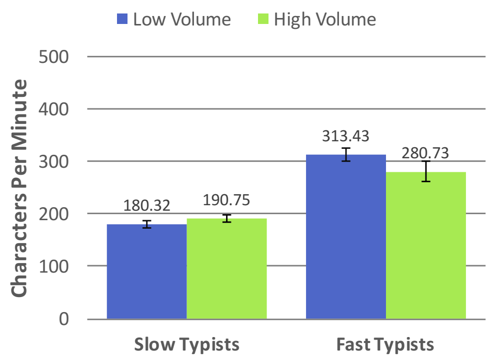


Figure 6.9: Trend towards significant interaction between volume and speed group (p = 0.071)

The simple main effect of volume was significant for fast typists, $F(1, 9) = 8.051$, $p = 0.019$, $\eta_\rho^2 = 0.472$, but not for slow typists, $F(1, 40) = 0.833$, *n.s.* The fast typists' CPM was higher when accompanied by low volume music and lower with high volume music (low volume: $M = 313.43$, $SD = 57.57$; high volume: $M = 280.73$, $SD = 97.86$).

The three-way interaction between vocals, volume and speed group was significant, $F(1, 49) = 6.410$, $p = 0.015$, $\eta_\rho^2 = 0.116$, (Figure 6.10). One interpretation of this significant interaction is that the two-way interaction between vocals and volume, which was significant, differing between the two speed groups[26].

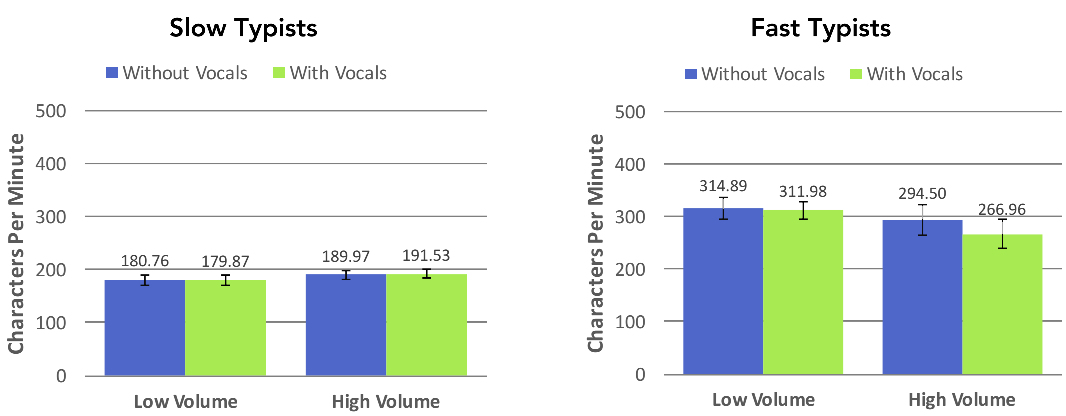


Figure 6.10: Significant three-way interaction between vocals, volume and speed group (p = 0.015)

The simple two-way interaction between vocals and volume was significant for fast typists, $F(1, 9) = 11.927$, $p = 0.007$, $\eta_\rho^2 = 0.570$, but not for slow typists, $F(1, 40) = 0.232$, *n.s.*

26 This interaction can also be conceptualized as the two-way interaction between volume and speed group, which was significant, differing between the two vocals levels. Or, it can be conceptualized as the two-way interaction between vocals and speed group, again a significant interaction, differing between the two volume group.

For the fast typists, the simple simple main effect of vocals was significant with high volume music, $F(1, 5) = 26.683$, $p = 0.004$, $\eta_\rho^2 = 0.842$, but not with low volume music, $F(1, 4) = 0.429$, *n.s.* When accompanied by high volume music, the fast typists' CPM was lower with music that contained vocals and higher when the music did not contain vocals (with vocals: $M = 266.96$, $SD = 96.31$; without vocals: $M = 294.50$, $SD = 101.66$).

Again, for fast typists, the simple simple main effect of volume was significant with music that did not contain vocals, $F(1, 9) = 14.474$, $p = 0.004$, $\eta_\rho^2 = 0.617$, but was not significant with vocals in the music, $F(1, 9) = 2.752$, *n.s.* When accompanied by music that contained vocals the fast typists' CPM was higher with low volume music and lower with high volume music (low volume: $M = 311.98$, $SD = 65.62$; high volume: $M = 266.96$, $SD = 96.31$).

The three-way interaction between text, volume and speed group was also significant, $F(1, 49) = 7.799$, $p = 0.007$, $\eta_\rho^2 = 0.137$, (Figure 6.11). This interaction can be conceptualized as the two-way interaction between text and volume, which was significant, differing between the two speed groups[27].

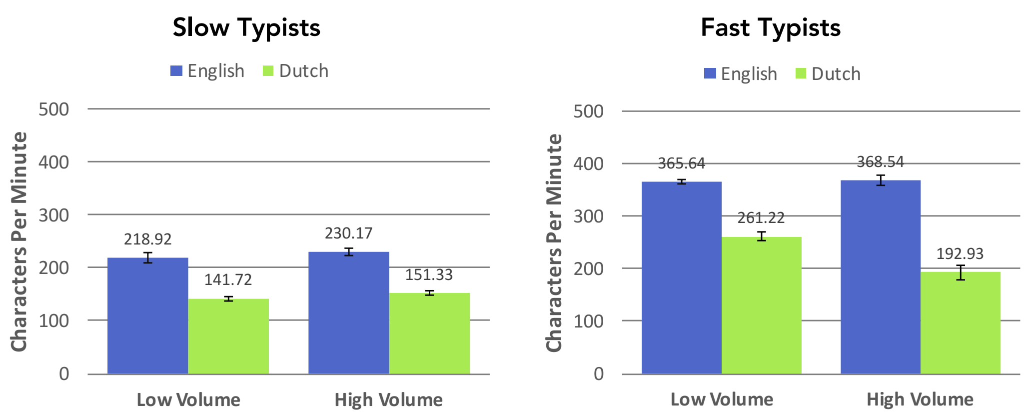


Figure 6.11: Significant three-way interaction between text, volume and speed group ($p = 0.007$)

The simple two-way interaction between text and volume was significant for the fast typists, $F(1, 9) = 5.643$, $p = 0.042$, $\eta_\rho^2 = 0.385$, but not for the slow typists, $F(1, 40) = 0.026$, *n.s.*

For fast typists, the simple simple main effect of volume was significant when typing in Dutch, $F(1, 9) = 8.056$, $p = 0.019$, $\eta_\rho^2 = 0.472$, but not in English, $F(1, 9) = 0.062$, *n.s.* When typing in Dutch, the fast typists' CPM was higher with low volume music and lower

27 This interaction can also be conceptualized as the two-way interaction between volume and speed group, which showed a trend towards significance, differing between the two texts. Or, it can be conceptualized as the two-way interaction between text and speed group, which was significant, differing between the two volume groups.

when accompanied by high volume music (low volume: *M* = 261.22, *SD* = 26.65; high volume: *M* = 192.93, *SD* = 48.17).

For fast typists, the simple simple main effect of text was significant for both the low, $F(1, 4) = 168.64$, $p < 0.0005$, $\eta_p^2 = 0.977$, and high volume groups, $F(1, 5) = 44.587$, $p = 0.001$, $\eta_p^2 = 0.899$. At both volume levels, the fast typists' CPM was higher when typing in English (low volume: *M* = 365.64, *SD* = 15.14; high volume: *M* = 368.54, *SD* = 29.71) and lower when typing in Dutch (low volume: *M* = 261.22, *SD* = 26.65; high volume: *M* = 192.93, *SD* = 48.17).

6.2.2.2 Error Rate

A histogram of error rate percentage for the English text using the data from both volume groups combined was created (Figure 6.12). This histogram shows that the error rate percentage data was positively skewed. Shapiro-Wilk's tests were performed, both with and without the typing speed classification applied (see Appendix F.1.2). These normality tests showed that, when the data was treated as two volume groups without the speed classification applied, all eight distributions of error rate percentage were very strongly non-normal with $p < 0.0005$. With the speed classification applied, all of the slow typists' distributions were again very strongly non-normal with $p < 0.0005$, while for the fast typists, three distributions were very strongly non-normal with $p \leq 0.002$ and one distribution had a moderate deviation from normality, $p = 0.017$.

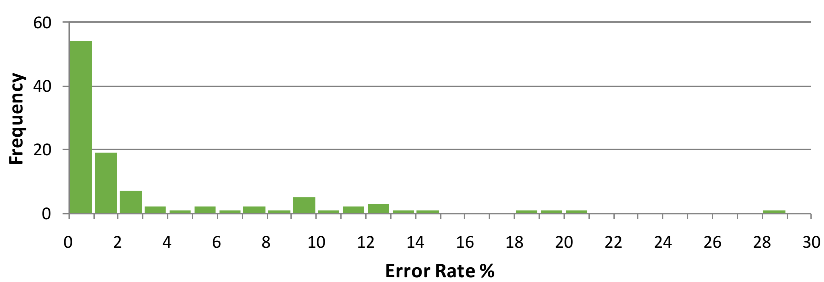


Figure 6.12: Histogram of error rate percentage for English tasks

As the error rate data had such severe deviations from normality, a logarithmic transformation was applied following the procedure described in Chapter 3, Section 3.5.2. The histogram of the transformed data appears to be bimodal (Figure 6.13).

Figure 6.13: Histogram of transformed error rate percentage for English tasks

A scatterplot of transformed error rate data was created (Figure 6.14) to explore whether a classification based on low and high accuracy could be used but it did not show sufficient clustering to infer a clear relationship, so no separation into ability groups based on accuracy was attempted.

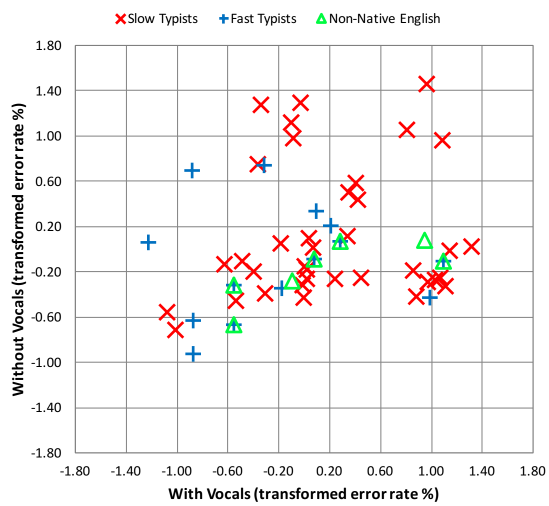


Figure 6.14: Scatterplot of transformed error rate percentage for English tasks

After transformation, box plots were created of the distributions (see Appendix F.1.3) and a number of data points in the sample became outliers. These outliers affected the outcomes from the analysis so were capped, following the procedure described in Chapter 3, Section 3.5.1.3. After outlier capping, when the data was treated as two groups without the speed classification applied, three of the eight distributions deviated from normality (37.5%). In the low volume group, the English task without vocals condition had a moderate deviation from normality (p = 0.031), and in the high volume group, two of the distributions had strong deviations from normality ($p \le 0.008$). With the speed classification applied, none of the fast typists' transformed error rate distributions deviated significantly from normality. But, two of the slow typists' distributions had strong deviations from normality ($p \le 0.009$) and two had moderate deviations from normality

($0.027 \leq p \leq 0.031$). Although these distributions violated normality by the Shapiro-Wilk's tests, inspection of the skewness and kurtosis characteristics of all the distributions showed they were suitable for analysis using ANOVAs as the *z* scores for all fell within the acceptable range, as specified in Chapter 3, Section 3.5.1.2.

Tables 6.5 and 6.6 show descriptive statistics for the transformed error rate, both with and without the speed classification applied.

Low Volume							
Text Condition	**Vocals Condition**	**All Typists** (n = 26)		**Slow Typists** (n = 21)		**Fast Typists** (n = 5)	
		M (%)	*SD* (%)	*M* (%)	*SD* (%)	*M* (%)	*SD* (%)
English	Without	0.054	0.52	-0.027	0.31	-0.044	0.65
	With	0.192	0.66	0.301	0.58	-0.268	0.86
Dutch	Without	0.185	0.52	0.333	0.45	-0.339	0.27
	With	0.499	0.58	0.513	0.56	0.440	0.71

Table 6.5: Descriptive statistics of transformed error rate percentage for low volume group after outlier capping

High Volume							
Text Condition	**Vocals Condition**	**All Typists** (n = 27)		**Slow Typists** (n = 21)		**Fast Typists** (n = 6)	
		M (%)	*SD* (%)	*M* (%)	*SD* (%)	*M* (%)	*SD* (%)
English	Without	0.027	0.55	0.047	0.58	-0.045	0.47
	With	0.042	0.68	0.096	0.68	-0.146	0.69
Dutch	Without	0.183	0.36	0.222	0.38	0.046	0.26
	With	0.341	0.66	0.332	0.69	0.217	0.33

Table 6.6: Descriptive statistics of transformed error rate percentage for high volume group after outlier capping

The analysis was first performed treating the transformed error rate data as two groups, without the speed classification applied. The between participants IV was volume (low and high), while the within participants IVs were presence of vocals (with and without) and presented text language (English and Dutch).

Aside from the violations of normality discussed previously, all other assumptions for mixed ANOVA were met. Table 6.7 reports the results from the mixed ANOVA.

Test	F	df	p	η_p^2
Vocals	5.151	1, 51	**0.027**	0.092
Vocals x Volume*	1.024	1, 51	n.s.	-
Text	8.156	1, 51	**0.006**	0.138
Text x Volume*	0.003	1, 51	n.s.	-
Text x Vocals	1.173	1, 51	n.s.	-
Text x Vocals x Volume*	0.013	1, 51	n.s.	-
Volume*	0.814	1, 51	n.s.	-

Table 6.7: Inferential analysis of transformed error rate percentage
(* between groups factor)

The omnibus effect of vocals was significant, $F(1, 51) = 5.151$, $p = 0.027$, $\eta_\rho^2 = 0.092$. The transformed error rate was higher when accompanied by music that contained vocals and lower when accompanied by instrumental music (with vocals: $M = 0.267$, $SD = 0.66$; without vocals: $M = 0.112$, $SD = 0.49$).

The omnibus effect of text was significant, $F(1, 51) = 8.156$, $p = 0.006$, $\eta_\rho^2 = 0.138$. The transformed error rate was higher when typing in Dutch and lower when typing in English (Dutch: $M = 0.301$, $SD = 0.55$; English: $M = 0.078$, $SD = 0.60$).

There were no other significant effects or interactions, so the analysis proceeded by including speed as an additional between participants factor and performing a mixed ANOVA. Aside from the normality violation discussed previously, all other assumptions for mixed ANOVA were met.

Table 6.8 presents the results from the analysis using a mixed ANOVA on the transformed error rate data with the speed classification applied and outlier values capped.

The omnibus effect of vocals approached significance, $F(1, 49) = 3.811$, $p = 0.057$, $\eta_\rho^2 = 0.072$. The transformed error rate was higher when the music contained vocals and lower with instrumental music (with vocals: $M = 0.258$, $SD = 0.65$; without vocals: $M = 0.104$, $SD = 0.47$).

The omnibus effect of text was significant, $F(1, 49) = 5.977$, $p = 0.018$, $\eta_\rho^2 = 0.109$. The transformed error rate was higher when typing in Dutch and lower when typing in English (Dutch: $M = 0.297$, $SD = 0.53$; English: $M = 0.065$, $SD = 0.59$).

Test	F	df	p	η_p^2	Figure
Vocals	3.811	1, 49	0.057**	0.072	-
Vocals x Volume*	1.591	1, 49	n.s.	-	-
Vocals x Speed*	0.004	1, 49	n.s.	-	-
Vocals x Volume* x Speed*	0.043	1, 49	n.s.	-	-
Text	5.977	1, 49	**0.018**	0.109	-
Text x Speed*	0.023	1, 49	n.s.	-	-
Text x Volume*	0.025	1, 49	n.s.	-	-
Text x Vocals	3.196	1, 49	0.080**	0.061	6.15
Text x Volume* x Speed*	0.070	1, 49	n.s.	-	-
Text x Vocals x Speed*	4.195	1, 49	**0.046**	0.079	6.16
Text x Vocals x Volume*	0.613	1, 49	n.s.	-	-
Text x Vocals x Volume* x Speed*	1.992	1, 49	n.s.	-	-
Volume*	0.026	1, 49	n.s.	-	-
Speed*	5.003	1, 49	**0.030**	0.093	-
Volume* x Speed*	0.650	1, 49	n.s.	-	-

Table 6.8: Inferential analysis of transformed error rate percentage with speed classification
(* between groups factor, ** non-significant result where $p < 0.1$)

The interaction between text and vocals also approached significance, $F(1, 49) = 3.196$, $p = 0.080$, $\eta_\rho^2 = 0.061$, (Figure 6.15).

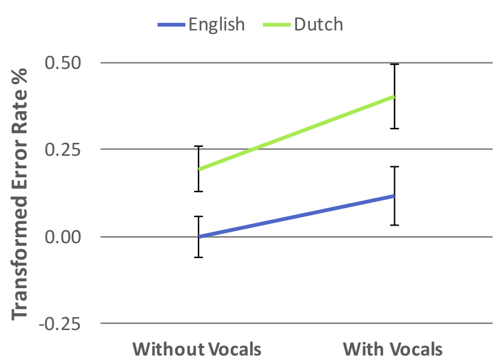


Figure 6.15: Trend towards a significant two-way interaction between text and vocals ($p = 0.080$)

The simple main effect of vocals was not significant in the English text conditions, $F(1, 49) = 0.011$, *n.s.* but was significant in the Dutch tasks, $F(1, 49) = 8.076$, $p = 0.007$, $\eta_\rho^2 = 0.141$. When typing in Dutch, the participants made fewer errors when the music was instrumental and more errors when the music contained vocals (without vocals: $M = 0.193$, $SD = 0.43$; with vocals: $M = 0.401$, $SD = 0.60$).

The three-way interaction between text, vocals and speed classification was significant, $F(1, 49) = 4.195$, $p = 0.046$, $\eta_\rho^2 = 0.079$, (Figure 6.16).

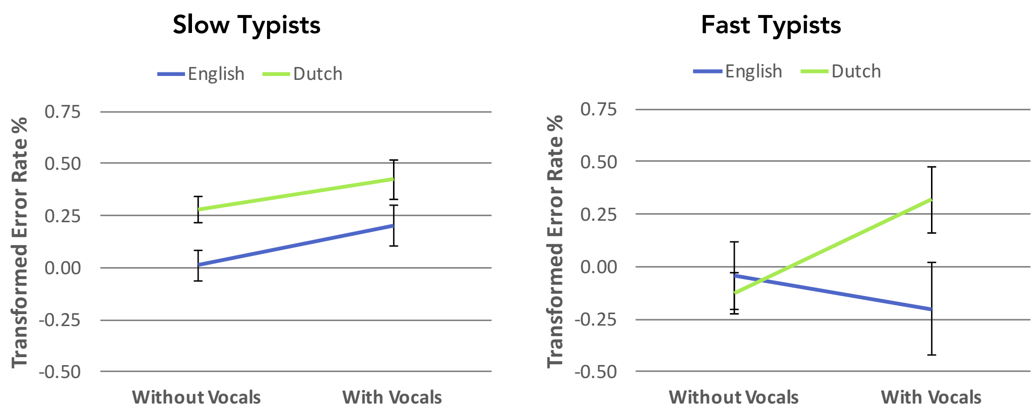


Figure 6.16: Significant three-way interaction between text, vocals and speed group (p = 0.044)

This interaction can be conceptualised as the two-way interaction between text and vocals, which approached significance, being different across the two levels of speed classification. However, the simple interaction between vocals and text was not significant for either the slow or fast typists (slow: $F(1, 40) = 0.092$, *n.s.*; fast: $F(1, 9) = 3.161$, *n.s.*).

Although these simple interactions are both not significant, they have quite different characteristics. For the slow typists, the p value was large with an extremely small effect size, while for the fast typists, the p value approached 0.1 with a large effect size (slow: p = 0.763, $\eta_\rho^2 = 0.002$; fast: $p = 0.109$, $\eta_\rho^2 = 0.260$). The considerable difference in characteristic between the two-way interactions between text and vocals at the different levels of speed group provides one possible explanation for why the three-way interaction was significant despite being unable to identify significant simple two-way interactions.

Another way to conceptualise the significant three-way interaction between text, vocals and speed group is to look at how the interaction between vocals and speed group differs between the two texts. However, this simple interaction was not significant for either the English, $F(1, 51) = 1.732$, *n.s.*, or Dutch, $F(1, 51) = 2.536$, *n.s.*, texts.

6.2.3 Typing Experience

Subjective ratings of task difficulty and music distraction were collected using 7-point Likert items. The design of this experiment cannot be accommodated by non-parametric tests, so as Likert item data is not suitable for parametric analysis the procedure described in Chapter 3, Section 3.5.3 was followed. The analysis was first performed using ANOVAs and significant effects or interactions followed by analysis using the most suitable non-parametric test.

6.2.3.1 Task Difficulty Ratings

The participants were asked to rate how easy they found the transcription typing task using a 7-point Likert item ranging from 1 = extremely difficult to 7 = extremely easy. The data was reversed for analysis so that low values map to a perception that the task is easier, with higher values mapping to a perception the task was harder.

Table 6.9 presents descriptive statistics of task difficulty ratings for all typists separated into the low and high volume groups, including the upper and lower quartiles (*Q1* and *Q3* respectively), the median (*Mdn*) and mean (*M*). Descriptive statistics separated into the two speed groups are included in Appendix F.2.1.

Text	Vocals	Low Volume (n = 26)				High Volume (n = 27)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
English	Without	2.00	3.00	3.25	2.69	2.00	2.00	4.00	2.89
	With	2.75	3.00	4.00	3.12	2.00	3.00	4.00	3.37
Dutch	Without	4.75	6.00	6.00	5.23	3.00	6.00	6.00	4.96
	With	5.00	5.00	6.00	5.38	5.00	5.00	6.00	5.26

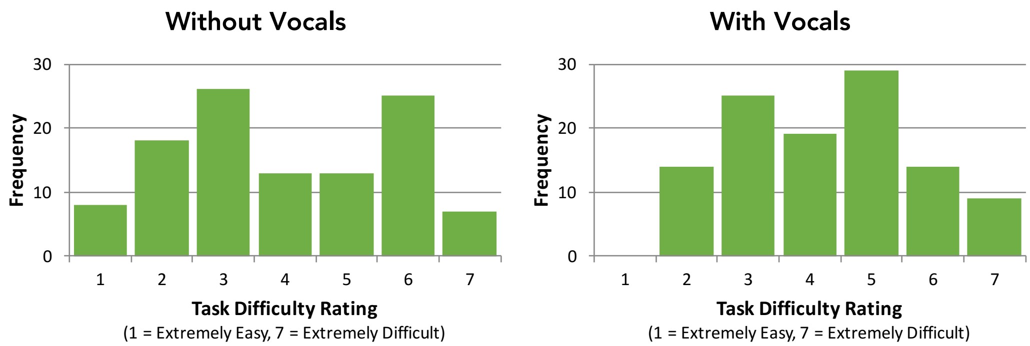
Table 6.9: Descriptive statistics of task difficulty ratings for low and high volume groups

The task difficulty ratings data was first analysed using a 2 by 2 by 2 mixed ANOVA. Volume (low and high) was the between participants variable and vocals (with and without) and text (English and Dutch) were the two repeated measures variables. Table 6.10 shows the results from this mixed ANOVA.

Test	F	df	p
Vocals	7.249	1, 51	**0.010**
Vocals x Volume*	0.159	1, 51	n.s.
Text	194.618	1, 51	**< 0.0005**
Text x Volume*	1.805	1, 51	n.s.
Text x Vocals	0.844	1, 51	n.s.
Text x Vocals x Volume*	0.023	1, 51	n.s.
Volume*	0.003	1, 51	n.s.

Table 6.10: Inferential analysis of task difficulty ratings
(* between groups factor)

The omnibus effect of vocals was significant, $F(1, 51) = 7.249$, $p = 0.010$, so this result needs to be investigated further using a non-parametric test. The data was restructured from four columns (one for each of the four tasks) to two columns (one column for the English and Dutch without vocals conditions combined, and the other column for the English and Dutch with vocals conditions combined). This data restructuring allowed me to perform a Wilcoxon signed rank test to compare the effect of vocals on task difficulty ratings. The test revealed a statistically significant effect of vocals, $z = -2.791$, $p = 0.005$, $r = 0.27$. The median of both conditions was 4.00, so to identify whether the with or without vocals condition was perceived as easier, two histograms were created (Figure 6.17).



(a) $n = 106$, $Q1 = 2.75$, $Mdn = 4.00$, $Q3 = 6.00$, $M = 3.94$

(b) $n = 106$, $Q1 = 3.00$, $Mdn = 4.00$, $Q3 = 5.00$, $M = 4.28$

Figure 6.17: Histograms of task difficulty ratings for alt rock music conditions

These histograms show that the without vocals condition (Figure 6.17, a) was perceived as easier than the with vocals condition (Figure 6.17, b), as there are more ratings below the neutral mid-point in the without vocals condition than with vocals.

The omnibus effect of text was significant in the mixed ANOVA as well. To investigate this further, the data was again restructured into two columns (one column containing the task difficulty ratings from both the English task conditions and the other column containing the task difficulty ratings from both the Dutch task conditions). A Wilcoxon signed rank test with the data condensed into two groups revealed a significant effect of text, $z = -8.369$, $p < 0.0005$, $r = 0.81$. The task difficulty ratings in the English tasks was lower than the ratings in the Dutch tasks (English: $Mdn = 3.00$; Dutch: $Mdn = 5.00$).

To investigate whether speed classification affected task difficulty ratings another mixed ANOVA was performed, this time including speed classification as a further between participants factor. Table 6.11 shows the results from this 2 by 2 by 2 by 2 mixed ANOVA.

Test	F	df	p
Vocals	7.684	1, 49	**0.008**
Vocals x Volume*	0.854	1, 49	n.s.
Vocals x Speed*	0.650	1, 49	n.s.
Vocals x Volume* x Speed*	4.425	1, 49	**0.041**
Text	111.803	1, 49	**< 0.0005**
Text x Speed*	1.033	1, 49	n.s.
Text x Volume*	0.801	1, 49	n.s.
Text x Vocals	2.634	1, 49	n.s.
Text x Volume* x Speed*	0.067	1, 49	n.s.
Text x Vocals x Speed*	2.479	1, 49	n.s.
Text x Vocals x Volume*	2.373	1, 49	n.s.
Text x Vocals x Volume* x Speed*	9.152	1, 49	**0.004**
Volume*	0.407	1, 49	n.s.
Speed*	9.412	1, 49	**0.004**
Volume* x Speed*	0.715	1, 49	n.s.

Table 6.11: Inferential analysis of task difficulty ratings with speed classification
(* between groups factor)

There were a few significant omnibus effects and interactions identified through this analysis. The omnibus effects of vocals and text were explored in the previous analysis without the speed factor, so these analyses were not repeated.

Speed classification was a significant between groups factor in the mixed ANOVA, $F(1, 49) = 9.412$, $p = 0.006$. To investigate this further, the data was restructured so that all the task difficulty ratings data was in a single column, with the speed classification grouping specified as a second column in the spreadsheet. A Mann-Whitney U test was then performed to investigate effect of speed classification on task difficulty ratings. The effect of speed classification was significant, $U = 2{,}590.00$, $z = -3.102$, $p = 0.002$, $r = 0.21$. The fast typists perceived the typing tasks as easier than the slow typists (fast: *Mdn* = 3.00; slow: *Mdn* = 4.00).

The three-way interaction between vocals, volume and speed group was significant, $F(1, 49) = 4.425$, $p = 0.041$, $\eta_p^2 = 0.083$, and so was the four-way interaction between text, vocals, volume and speed group, $F(1, 49) = 9.152$, $p = 0.004$. As the four-way interaction was stronger than the three-way interaction, rather than explore both, only the significant four-way interaction was explored further with a non-parametric analysis.

The significant four-way interaction between text, vocals, volume and speed group was further analysed by performing two sets of non-parametric tests. The first set of four Wilcoxon signed rank tests compared the effect of vocals on perceived task difficulty for each of the four separate speed and volume groups. Table 6.12 presents the results from these tests.

Effect of Vocals			
Volume	**Text**	**Slow Typists**	**Fast Typists**
Low	English	$z = -1.500$, *n.s.*	$z = -2.000$, $p = 0.046$, $r = 0.89$
	Dutch	$z = -0.489$, *n.s.*	$z = -1.732$, *n.s.*
High	English	$z = -0.684$, n.s.	$z = -2.333$, $p = 0.020$, $r = 0.95$
	Dutch	$z = -2.218$, $p = 0.027$, $r = 0.48$	$z = -1.069$, *n.s.*

Table 6.12: Wilcoxon signed rank tests for the effect of vocals on task difficulty ratings

The results in Table 6.12 clearly show that the effect of vocals on task difficulty ratings depends on the combination of speed classification and volume condition. The effect of vocals was significant for the fast typists in the English tasks at both volume levels, but not in the Dutch tasks. While, the effect of vocals was only significant for the slow typists who were exposed to high volume music and only in the Dutch language conditions. These results are explored further in the following paragraphs, beginning with the slow typists and moving onto the fast typists.

The slow typists that were exposed to low volume music did not perceive any difference in task difficulty due to the presence of vocals in the music, in either the English, $z = -1.500$, *n.s.*, or Dutch language tasks, $z = -0.489$, *n.s.*

However, when accompanied by high volume music, the slow typists perceived differences in task difficulty due to the presence of vocals in the music when typing in Dutch, $z = -2.218$, $p = 0.027$, $r = 0.48$, but not when typing in English, $z = -0.684$, *n.s.* When typing in Dutch accompanied by high volume music, the medians for the with and without vocals conditions were both 6.00, but inspection of the quartiles and means show that the slow typists perceived the task as easier without vocals in the music and harder when the music contained vocals (with vocals: $Q1 = 5.00$, $Mdn = 6.00$, $Q3 = 6.50$, $M = 5.62$; without vocals: $Q1 = 3.00$, $Mdn = 6.00$, $Q3 = 6.00$, $M = 4.95$).

The fast typists that were accompanied by low volume music perceived differences in task difficulty due to the presence of vocals in the music when typing in English, but not in the Dutch tasks (English: $z = -2.000$, $p = 0.046$, $r = 0.89$; Dutch: $z = -1.732$, *n.s.*). The fast typists who were exposed to low volume music perceived the English transcription task to be significantly harder when the accompanying music contained vocals and easier when the music did not contain vocals (with vocals: *Mdn* = 3.00; without vocals: *Mdn* = 2.00).

The fast typists who were accompanied by high volume music did not perceive differences in task difficulty due to the presence of vocals in the music when typing in Dutch, but the presence of vocals did affect perceived difficulty in the English tasks (Dutch: $z = -1.069$, *n.s.*; English: $z = -2.333$, $p = 0.020$). When typing in English, the fast typists who were exposed to high volume music found the task significantly harder when the music contained vocals and easier when it was instrumental (with vocals: *Mdn* = 3.00; without vocals: *Mdn* = 2.00).

The second set of tests involved performing four Mann Whitney *U* tests to compare the effect of volume on perceived task difficulty for each of the four separate speed groups and task conditions. Table 6.13 shows the results of these tests.

Effect of Volume			
Text	**Vocals**	**Slow Typists**	**Fast Typists**
English	Without	$U = 219.5$, $z = -0.026$, *n.s.*	$U = 11.0$, $z = -0.782$, *n.s.*
	With	$U = 215.5$, $z = -0.131$, *n.s.*	$U = 8.5$, $z = -1.308$, *n.s.*
Dutch	Without	$U = 196.5$, $z = -0.655$, *n.s.*	$U = 8.0$, $z = -1.357$, *n.s.*
	With	$U = 203.5$, $z = -0.451$, *n.s.*	$U = 10.0$, $z = -0.341$, *n.s.*

Table 6.13: Mann Whitney *U* tests for the effect of volume on task difficulty ratings

None of the pairwise comparisons investigating the effect of volume on task difficulty ratings were significant.

6.2.3.2 Music Distraction

The participants were also asked to rate how distracting they found the music using a 7-point Likert item ranging from 1 = extremely distracting to 7 = not at all distracting. For the analysis, the values were reversed so that the mapping was from 1 = not at all distracting to 7 = extremely distracting.

Table 6.14 shows descriptive statistics of task difficulty ratings for all typists separated into low and high volume groups. Descriptive statistics including the speed classification are included in Appendix F.2.2.

Text	Vocals	Low Volume (n = 26)				High Volume (n = 27)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
English	**Without**	1.00	1.50	3.00	2.00	2.00	2.00	4.00	2.74
	With	1.00	2.00	4.00	2.38	2.00	3.00	5.00	3.30
Dutch	**Without**	1.00	2.00	4.00	2.46	1.00	2.00	5.00	2.93
	With	1.00	2.00	3.00	2.31	2.00	3.00	5.00	3.19

Table 6.14: Descriptive statistics of music distraction ratings for all typists

A 2 by 2 by 2 mixed ANOVA was conducted as an initial analysis of the music distraction ratings. In this analysis, volume was the between participants IV, and vocals and text were the two repeated measures variables. The results from this mixed ANOVA are shown in Table 6.15.

Test	F	df	p
Vocals	1.723	1, 51	n.s.
Vocals x Volume*	0.538	1, 51	n.s.
Text	0.683	1, 51	n.s.
Text x Volume*	0.313	1, 51	n.s.
Text x Vocals	3.733	1, 51	0.059**
Text x Vocals x Volume*	0.314	1, 51	n.s.
Volume*	4.425	1, 51	**0.040**

Table 6.15: Inferential analysis of music distraction ratings
(* between groups factor, ** non-significant result where $p < 0.1$)

Volume was a significant between groups factor, $F(1, 51) = 4.425$, $p = 0.040$. To explore this result using a non-parametric test, the distraction ratings data was combined into a single group and the effect of volume was analysed using a Mann-Whitney U test, which was significant, U = 4,171.000, z = -3.328, $p = 0.001$, $r = 0.23$. The high volume music was perceived as more distracting than the low volume music (high volume: *Mdn* = 3.00; low volume: *Mdn* = 2.00).

The interaction between text and vocals was significant in the mixed ANOVA, $F(1, 51) = 3.733$, $p = 0.059$, but interactions cannot be accommodated by a non-parametric test. So, to explore this interaction further Wilcoxon signed rank tests were performed looking at the effect of vocals for the two texts separately. Table 6.16 shows the results from these tests.

Text	Effect of Vocals
English	z = -2.165, p = 0.030, r = 0.30
Dutch	z = -0.425, *n.s.*

Table 6.16: Wilcoxon signed rank tests of music distraction ratings

The effect of vocals was significant in the English task conditions, z = -2.165, p = 0.030, r = 0.30, but not in the Dutch, z = -0.425, *n.s.* In the English tasks, the medians and lower quartile distraction ratings of music distraction were the same in the with and without vocals conditions, but the upper quartile values were different (with vocals: $Q1$ = 1.00, *Mdn* = 2.00, $Q3$ = 5.00; without vocals: $Q1$ = 1.00, *Mdn* = 2.00, $Q3$ = 3.00). This difference in upper quartile values indicates the music with vocals was perceived as more distracting than the music without vocals.

There were no other significant effects or interactions from the analysis of the music distraction ratings data.

To explore whether the speed classification affected the music distraction ratings a mixed ANOVA was first performed. Table 6.17 shows the results from the analysis.

Test	F	df	p
Vocals	2.265	1, 49	n.s.
Vocals x Volume*	0.021	1, 49	n.s.
Vocals x Speed*	0.547	1, 49	n.s.
Vocals x Volume* x Speed*	0.512	1, 49	n.s.
Text	0.908	1, 49	n.s.
Text x Speed*	0.292	1, 49	n.s.
Text x Volume*	0.002	1, 49	n.s.
Text x Vocals	4.002	1, 49	0.051**
Text x Volume* x Speed*	0.757	1, 49	n.s.
Text x Vocals x Speed*	0.592	1, 49	n.s.
Text x Vocals x Volume*	0.406	1, 49	n.s.
Text x Vocals x Volume* x Speed*	0.082	1, 49	n.s.
Volume*	1.879	1, 49	n.s.
Speed*	0.258	1, 49	n.s.
Volume* x Speed*	0.244	1, 49	n.s.

Table 6.17: Inferential analysis of music distraction ratings with speed classification
(* between groups factor, ** non-significant result where p <0.1)

The two-way interaction between text and vocals approached significance, $F(1, 49)$ = 4.002, p = 0.051. This interaction has already been explored in the previous analysis without the speed classification, so is not repeated here. There were no other significant effects or interactions with the speed classification applied.

6.2.4 Summary of Results

The analysis presented in the previous sections is detailed and complex, so the key results are summarised in the following paragraphs, starting with typing performance and moving onto typing experience.

6.2.4.1 Typing Performance Summary

When the CPM data was analysed as two groups, without the speed classification applied, the only significant omnibus effect was for text. CPM was significantly higher in the English tasks than when typing in Dutch. There were no significant interactions and volume group was not a significant between groups factor. However, with the speed classification applied the results included numerous significant omnibus effects and interactions.

With the typing speed classification applied, the omnibus effect of vocals on CPM was significant. Overall, typing speed was higher when accompanied by instrumental music and lower when the music contained vocals. The omnibus effect of text was also significant, with higher CPM in the English tasks and lower CPM in the Dutch tasks. Volume was not a significant between groups factor overall. Speed classification was a significant between groups factor, with the fast typists' CPM higher than the slow typists'.

The two-way interaction between vocals and volume was significant. The simple main effect of vocals was significant for high volume music only. CPM was higher when the high volume music was instrumental, and lower when it contained vocals.

The two-way interaction between vocals and speed group was significant. The simple main effect of vocals was significant for fast typists, but not for slow typists. The fast typists' CPM was higher when accompanied by instrumental music and lower when the music contained vocals.

The two-way interaction between volume and speed group approached significance. The simple main effect of volume was significant for fast typists but not for slow typists. The fast typists' CPM was higher with low volume music, and lower with high volume music.

The three-way interaction between vocals, volume and speed group was significant. The simple two-way interaction was significant for fast typists but not for slow typists. The simple simple main effect of vocals was significant with high volume music, but not when the participants were accompanied by low volume music. When accompanied by high volume music, the fast typists' CPM was higher with instrumental music and lower with music that contained vocals. The simple simple main effect of volume was significant for music without vocals, but not for music with vocals. The fast typists' CPM was higher when the without vocals music was played at a low volume, and lower when it was played at a high volume.

The two-way interaction between text and speed group was significant. The simple main effect of text was significant for both slow and fast typists. CPM was higher in the English tasks and lower in the Dutch tasks for both speed groups. The difference in CPM between the English and Dutch tasks for fast typists was larger than the difference for the slow, explaining the significant interaction.

The two-way interaction between text and volume was significant. The simple main effect of volume was significant in the Dutch tasks but not in the English tasks. When typing in Dutch, CPM was higher with low volume music and lower with high volume music.

The three-way interaction between text, volume and speed group was significant. The simple two-way interaction between text and volume was significant for the fast typists, but not for the slow typists. The simple simple main effect of volume was significant in the Dutch tasks, but not in English. When the fast typists were typing in Dutch, CPM was higher when the music was played at the low volume level. The simple simple main effect of text was significant at both volume levels, with the fast typists' CPM higher when typing in English and lower when typing in Dutch.

The two-way interaction between text and vocals approached significance. The simple main effect of vocals was significant in English tasks but not in the Dutch tasks. In the English tasks, CPM was higher with instrumental music and lower when the music contained vocals.

A logarithmic transformation was applied to the error rate data before analysis. When analysed without the speed classification applied, the omnibus effect of vocals was significant. The transformed error rate percentage was higher when the music contained vocals, and lower with instrumental music. The omnibus effect of text was also significant, with higher transformed error rates in the Dutch tasks and lower in the English tasks. Volume was not a significant between groups factor and all other interactions were non-significant.

With the speed classification applied, there was a trend towards a significant effect of vocals, again with higher transformed error rates when the music contained vocals and lower transformed error rates with instrumental music. The omnibus effect of text was significant, with higher transformed error rates in the Dutch tasks than in the English tasks. Speed classification was a significant between groups factor, the slow typists' transformed error rate was higher than the fast typists'. Volume was not a significant between groups factor. This analysis of the omnibus effects and between groups factors aligns with the analysis without the speed classification applied, and also with the previous experiments' results that the fast typists make fewer errors than the slow typists.

The two-way interaction between text and vocals approached significance. The simple main effect of vocals was not significant in the English tasks conditions but was significant in the Dutch tasks. When typing in Dutch, the transformed error rates were lower with instrumental music and higher when the music contained vocals.

The three-way interaction between text, vocals and speed group was significant. The simple two-way interaction between text and vocals was not significant for either slow or fast typists. However, the simple two-way interaction between text and vocals for fast typists had a fairly low p value (p = 0.109) and a large effect size (η_{ρ}^2 = 0.260), so is potentially interesting.

6.2.4.2 Typing Experience Summary

The typing experience measures were first analysed using ANOVAS, with significant effects or interactions followed up with an appropriate non-parametric test.

When the task difficulty ratings data was analysed without the speed classification applied, there was a significant effect of vocals. Task difficulty ratings were lower with instrumental music and higher when the music contained vocals. There was also a significant effect of text, with the English tasks perceived as easier than the Dutch tasks.

With the speed classification applied, the ANOVA resulted in a significant four-way interaction between text, vocals, volume and speed group. This result was investigated further by looking at the effect of vocals for each of the two texts, for the four groups separately (slow typists low volume, slow typists high volume, fast typists low volume, fast typists high volume).

The effect of vocals was significant for the fast typists in the English tasks with both low and high volume music. The fast typists perceived the English tasks as harder when the music contained vocals, both with low and high volume music.

For the slow typists, the effect of vocals on task difficulty ratings was only significant in the Dutch tasks with high volume music. When typing in Dutch, with high volume music, the task was perceived as easier when the music did not contain vocals and harder with music that did contain vocals.

When the music distraction ratings were analysed without the speed classification applied there was a significant effect of volume, with the high volume music perceived as more distracting than the low volume music. There was also a trend towards an interaction between text and vocals which was investigated further with a Wilcoxon signed rank test which revealed a significant effect of vocals in the English tasks but not the Dutch. In the English tasks, music that contained vocals was perceived as more distracting than music without vocals.

With the speed classification applied, the main effect of volume in the mixed ANOVA no longer reached significance, but the interaction between text and vocals was still significant. This result was not explored further, as it had already been identified in the analysis without the speed classification applied.

6.3 Discussion

The analysis presented in the previous sections adds considerable weight to the argument that there are dimensions of music that can be manipulated to affect typing performance and experience. The speed classification was again important, particularly in the analysis of the effect of music on the CPM and task difficulty ratings.

Generally speaking, transcription typing with music that contained vocals was perceived as harder than with instrumental music, and performance was reduced. Though, this was not the case with all of the measures, nor for all of the groups. The results relating to the effect of vocals are explored further in the discussion of Experiment Aim 1.

The music's volume did not have an overall effect on typing performance or task difficulty ratings, but the high volume music was perceived as more distracting than the low volume music. Although volume did not, overall, affect performance there were a number of interesting interactions involving the volume dimension. These interactions show the volume of the music was an important factor affecting performance and experience which has implications for running an experiment using an online methodology as I would be unable to control or easily monitor the playback volume.

In the previous experiments, the slow typists' performance was not affected by the music IVs in the same way as the fast typists' performance. One possible explanation was that the slow typists needed to concentrate on the task more, reducing the impact of the music on their performance. The Dutch transcription typing task was included in this experiment to explore the hypothesis that increasing the level of concentration and attention required on a task reduces the effect of any of the music IVs. To a certain extent, this manipulation has been effective as the fast typists' speed and experience when typing in Dutch does not seem to have been affected. But, in terms of typing accuracy, there was a trend towards a significant interaction between text and vocals when both speed groups were considered together, suggesting that error rates in the Dutch task were increased with vocals in the music but not in the English tasks, albeit inconclusively. The three-way interaction between text, vocals and speed group, was significant and though the simple interactions were both non-significant there were clearly large differences in characteristics between the slow and fast typists' simple interaction.

6.3.1 Discussion of Each Experiment Aim

The initial aims of this experiment were:

1. to gather further evidence that the presence of vocals in a piece of music has a negative impact on typing performance and experience,
2. to investigate how the volume of music affects typing performance and experience,
3. to investigate if the previously identified effect of vocals is consistent across music played at different volume levels,
4. to see whether an increase in task difficulty reduces the effect of the music IVs on typing performance and experience,
5. to gather further evidence that the effect of music on typing performance and experience differs between slow and fast typists,
6. to validate the method is appropriate for a classroom based methodology.

The results presented in the previous section are discussed in more detail in relation to the specific experiment aims in the following sections.

6.3.1.1 Experiment Aim 1

The primary aim of this experiment was to build on the work from Experiment 1 and gather more conclusive evidence that vocals in music reduces performance when typing by making the task more difficult. To explore this aim, both the omnibus effect of vocals and the interaction between vocals and speed classification need to be studied.

In terms of typing speed, the omnibus effect of vocals was significant with performance reduced when the music contained vocals. However, the interaction between vocals and speed group was also significant. The simple main effect of vocals was only significant for the fast typists and not the slow typists. There are a couple of possible explanations for why the omnibus effect of vocals reached significance when the simple main effect was only significant for the fast typists.

One explanation is that the effect for the fast typists was strong, with such a large effect size, that this carried through to an omnibus effect even though only the fast typists actually experienced the effect. However, as only 11 participants out of a sample of 53 were classified as fast typists (i.e. only 20% of the sample) the effect size for the fast typists would need to be extremely large for this to be true. The effect size was large, $\eta_p^2 = 0.670$, with 67% of the variance in typing speed for the fast typists attributed to the presence of vocals in the music. So, it is possible that this large effect size explains why the omnibus effect of vocals was significant when the simple main effect of vocals was only significant for the fast typists and not for the slow typists as well.

However, another possible explanation for this discrepancy is that many of the slow typists were similarly affected by the presence of vocals in the music, but the effect was neither large enough nor consistent enough for the simple main effect to reach significance for the slow typists group when the analysis considered this subset of the data separately. Inspection of the initial histogram of typing speed suggested the distribution might be better approximated as three groups rather than two but there were too many participants whose classification was inconsistent in a three group model, so it was not explored further. Perhaps, there is a group of typists with a skill level that falls between the slow and fast typists. If this middle group of typists does exist in the sample, then it may be that the slowest typists are not affected by the vocals, but the middle group were affected by vocals in the music in a similar way to the fast typists. This would add weight to the omnibus effect of vocals, while still explaining why the effect of vocals for the slow typists group (when it contains potentially both slow and mid speed typists) was not significant in terms of typing speed.

It is not clear which explanation is correct, and there are additional complications due to the task difficulty manipulation and volume groups which are discussed in relation to later experiment aims. However, clearly, another experiment with larger a sample size, particularly of fast typists, is needed to explore this further.

The omnibus effect of vocals on typing accuracy was significant, with higher error rates when the music contained vocals. It is interesting that although the speed classification was significant, with the fast typists making fewer errors than the slow typists, the typing accuracy analysis was not strongly affected by the speed classification. Overall, typing performance in terms of accuracy of the whole sample was improved with instrumental rather than music that contained vocals.

In terms of typing experience, there was an overall significant effect of vocals in task difficulty ratings. The tasks were perceived as harder when the music contained vocals. Though, as the four-way interaction between text, vocals, volume and speed group was significant, it would be premature to claim all the transcription typing tasks without vocals were perceived as easier than the tasks with vocals in the music.

The omnibus effect of vocals on music distraction ratings was not significant so this experiment has not been able to provide evidence that music which contains vocals is perceived as more distracting than music that does not contain vocals overall. However, there was a trend towards a significant interaction between text and vocals which again ought to be considered further in the discussion.

This first aim of this experiment has been partially achieved, as there is clear evidence that vocals in the music affected typing accuracy overall. In terms of typing speed, there is evidence the performance of the fast typists was reduced with vocals in the music, but no evidence the slow typists' performance was similarly reduced. The typing experience analysis led to a significant main effect of vocals on task difficulty ratings with the task perceived as harder with vocals in the music, but an explanation in terms of vocals alone would be somewhat hasty given the significant four-way interaction.

6.3.1.2 Experiment Aim 2

The second aim of the experiment was to explore how the volume of music affects transcription typing performance and experience, in order to establish if an online methodology would be suitable in future experiments. To assess whether the experiment has been successful in achieving this aim, the effect of volume as a between groups factor needs to be explored, as well as considering the interaction between volume and speed group.

In terms of typing speed, the omnibus effect of volume was not significant. But, the interaction between volume and speed group was significant. The fast typists' CPM was higher with quieter music and lower with louder music, but there was no evidence the slow typists' CPM was affected by the volume of music.

The analysis of the effect of volume on typing accuracy was not significant, and there were no significant interactions involving the volume IV. Therefore, this experiment has not been able to provide evidence the volume of music affected typing accuracy overall.

With regard to typing experience, the analysis of the effect of volume on task difficulty ratings was not significant, so again there is no evidence that the volume of music affected task difficulty ratings. However, when the music distraction data was analysed there was a significant main effect of volume, with the high volume music perceived as more distracting than the low volume music. It is interesting that this significant effect, with a moderate effect size, did not lead to a reduction in performance in the high volume conditions for all typists.

This experiment aim has again been partially met, as there is evidence the high volume music was perceived as more distracting than the low volume music overall. But, in terms of an impact on typing performance, volume was only observed to have an effect on the fast typists' CPM and not the slow typists. There was also no indication that the volume of the music made the participants perceive the task as harder or affected typing accuracy.

6.3.1.3 Experiment Aim 3

The third aim of this experiment looked to identify how any effects due to the vocals dimension differ between the two different volume levels. To understand if the experiment was successful with regard to this aim, the interactions between vocals and volume, and between vocals, volume and speed group, needs to be explored.

There is clear evidence that, in terms of typing speed, the effect of vocals on performance differs between the low and high volume groups. The two-way interaction between vocals and volume was significant, with a significant simple effect of vocals only identified in the high volume group, where vocals in the music reduced CPM. The three-way interaction between vocals, volume and speed group was also significant, though the simple two-way interaction was only significant for the fast typists and not the slow typists.

Again, this can be explained in a couple of ways. Either, the slow typists were not affected in the same way as the fast typists, but because the interaction between vocals and volume was strong for the fast typists it led to the higher level interaction being significant as well. Alternatively, some of the slow typists may have been adding to the overall effect, but not all of them, leading to a significant higher level interaction for the fast typists but not the slow typists. Both explanations provide evidence that the effect of vocals on CPM that was previously identified, may actually just be constrained to the high volume group.

In terms of typing accuracy, the interaction between vocals and volume was not significant, and neither was the interaction between vocals, volume and speed group. This result demonstrates the overall effect of vocals that was identified, is not contingent on the volume of the music, nor the speed classification of the participants.

For task difficulty ratings, the two-way interaction between vocals and volume was not significant, but the three-way interaction between vocals, volume and speed group was significant. However, as the four way interaction which also included text as a further variable, was also significant, this three-way interaction was not considered further. But, there is a suggestion that the effect of vocals on task difficulty ratings depends on both the volume and speed group.

The interaction between volume and vocals had no effect on the music distraction ratings, demonstrating that the overall effect of volume was not contingent on the presence of vocals in the music.

The experiment has been successful in terms of this aim as there is clear evidence that in both the typing speed and task difficulty analyses the effect of vocals differs between the volume groups. It seems that the effects of vocals that were previously identified may only be important with high volume music and not with low volume.

6.3.1.4 Experiment Aim 4

The fourth aim of the experiment was to explore whether any of the effects due to music, e.g. any reduction in speed due to the vocals in the music, were consistent across transcription typing tasks with different intended difficulty levels. To inspect this aim further, all analyses involving the task difficulty IV need to be studied.

The effect of text was significant for CPM, transformed error rate percentage and task difficulty ratings, showing clearly that the Dutch transcription tasks reduced performance and made the task harder. These results are unsurprising given that the participants did not speak the Dutch language. The effect of text on music distraction ratings was not significant overall, so the difficulty of the task did not affect how distracting the participants found the music.

In terms of typing speed, the text manipulation was important in relation to the music IVs. The two-way interaction between text and volume was significant, and so was the three-way interaction between text, volume and speed group. Interestingly, I expected that as the Dutch tasks were harder, any effect due to the music IVs would be reduced in the Dutch tasks, but this was not the case. Instead, the significant two-way and three-way interactions revealed that in the Dutch language tasks performance was better with quieter music and worse with louder music. However, the same effect was not identified in the English tasks. Again, the speed classification revealed this effect for fast typists but not for slow.

To explain this, perhaps surprising, outcome it is important to recall that the loud music was perceived as significantly more distracting than the quiet music overall. The Dutch typing tasks were hard and definitely required more concentration by the participants than the English typing tasks. I expect that this result can be explained by the louder music, which was more distracting, penetrating the participants' ability to concentrate on the hard task. The task was hard, and boring, so the louder music distracted the participants to the extent that their speed was reduced. The quieter music was less distracting, so the participants could still concentrate on the hard task, perhaps by managing to block out the music and focus entirely on the task.

It is interesting that the more distracting music, in terms of volume, did not reduce the performance in the English tasks as well. This may be due to a ceiling effect masking the impact of volume in the English tasks. Alternatively, the interaction between text and vocals was significant for the music distraction ratings, with the effect of vocals only being significant in the English tasks. It may be that in the English tasks the vocals are a larger distraction when in the low volume music than when included in the high volume music, meaning that it was not the volume alone that reduces speed, but it is the relationship between the presence of vocals and the volume of the music that leads to these effects. The significant interactions between vocals and volume, and between vocals, volume and speed support this explanation.

In terms of typing accuracy, there was a trend towards a significant interaction between text and vocals, and the interaction between text, vocals and speed group was also significant. The evidence from these interactions was inconclusive. The overall interaction between text and vocals only led to a trend towards significance, the simple main effect of vocals was significant in Dutch and not English. Despite a significant three-way interaction with a large effect size, the simple two-way interaction between text and vocals was not significant for either speed group. So, this result is unclear and requires further exploration in future experiments to establish why, for typing accuracy, the hard typing tasks may be more affected by the presence of vocals in the music, when I had originally hypothesised the opposite.

In the analysis of task difficulty ratings, the language of the typing task was also found to be important as the four-way interaction between text, vocals, volume and speed group was significant. Further exploration of this interaction showed that the first typists perceived the vocals to have a significant effect on task difficulty ratings in the English tasks only, across both volume groups. Contrastingly, the effect of vocals on task difficulty ratings for the slow typists was only significant in the high volume Dutch transcription typing task. These inconsistent results are complicated, but clearly show the effect of the music IVs on task difficulty ratings depends on whether the participants were classified as slow or fast typists. For the fast typists, the result is as I originally hypothesised as vocals in the music made the task harder at both volume levels, in the tasks that involved typing text in English, indicating that the fast typists perceived the task as harder with conflicting verbal stimuli. Arguably, when typing in Dutch, vocals in the music should not conflict in the same manner as the typing task does not require the participants to read words in English. However, the effect of vocals on task difficulty ratings for the slow typists is harder to explain.

Overall, Experiment aim 4 has been achieved as the results show that for typing speed, typing accuracy, task difficulty and music distraction ratings the language of the text, and therefore the difficulty of the task, moderates any effects due to the music IVs.

6.3.1.5 Experiment Aim 5

The fifth aim of the experiment involved considering whether the typing speed classification is still important in the analysis, by gathering further evidence that the effect of music on performance and experience depends on the skill of the typist, as defined by their speed classification. This aim has also been achieved, as there is clear evidence that the performance of fast typists was affected by music in ways that the slow typists were not and vice versa. These effects were explored in more detail with regard to the specific experiment aims presented in previous sections of this discussion.

6.3.1.6 Experiment Aim 6

The final aim of the experiment was to consider whether this classroom based methodology is appropriate, and whether the data collected leads to appropriate outcomes that can be considered experimentally sound. In my opinion, this aim has also been achieved, which can be justified largely through the thorough piloting process that happened prior to running the experiment where a number of problems were identified with the instructions (discussed in Chapter 3, Section 3.4.2). The method itself seems to be robust, leading to experiment outcomes that can be considered valid and safe. The only limitations of using the classroom based methodology relate to the participant demographics as they were constrained to the students who were taking the module. If more data is needed, for example, in this experiment it would have been beneficial to include additional participants in the fast typist group, it would be very difficult to add participants without introducing confounds, e.g. we could not run extra participants using a laboratory based methodology without compromising the integrity of the experiment.

6.3.2 Limitations of This Experiment

There were a few limitations of this experiment, which are addressed in this section. One important limitation is that the participants heard music in all of the experiment conditions, meaning that there was no condition where the typing tasks were completed without exposure to music. This limitation was caused primarily by the setting of the experiment, i.e. as part of a practical class rather than using a laboratory. The classroom is a busy environment, which is typically quiet, but not silent. It would have been hard to control the room sufficiently to eliminate confounds for a without music condition, so one was not included. As there was no without music condition in this experiment, I cannot make any claims that typing with music improves or reduces performance in comparison with no music playing, reducing the potential impact of the results. For example, best performance was achieved with lowest distraction ratings when the participants were exposed to the low volume instrumental music and although typing under this condition was better than when accompanied by high volume that contained vocals, it does not automatically follow that it would also provide a better typing experience than without exposure to any music.

The modest number of fast typists in the experiment is a further limitation. Only 11 of the 53 typists included in the analysis were classified as fast, again limiting the generalisability of the outcomes. Due to the setting of the experiment, it was not possible to select participants with a range of abilities and no further participants could be added without introducing confounds. Experiments with a large group of fast typists are therefore needed to validate the typing speed results in particular.

The results from this experiment suggest that the fast typists were affected by the music in ways that the slow typists were not. Therefore, the comparatively small number of fast typists within the sample limits this experiment. The first two experiments in this thesis, had a minimum of 1/3 participants classified as fast typists from the opportunity sample. But, in this experiment, where the participants were restricted to first year undergraduate Computer Science students, the proportion of participants who were classified as fast typists was reduced to 1/5 of the sample. This reduced proportion of fast typists in the sample is probably due to the difference in education level of the participants. In the previous experiments, a higher proportion of the participants were postgraduate students rather than undergraduates which appears to have affected the numbers of participants who could be classified as fast typists.

6.3.3 Moving Forwards

The next step in this research is to perform a similar study with larger numbers of participants, particularly those that can be classified as fast typists. However, as volume has been identified as an interesting, potentially relevant factor that did affect both typing performance and experience the use of an online methodology would be inappropriate without further investigation. So, the next step, is to undertake another experiment using a similar classroom based methodology over two years to double the number of participants in each group. If the proportion of slow and fast typists is consistent, this should lead to a minimum of 20 participants classified as fast typists.

Chapter 7

Vocals, Volume and Task Difficulty 2

Experiment 4

The results from Experiment 3 provided further evidence that vocals in music reduce typing performance and detract from the experience of typing, as well as suggesting that the volume of the music is an important factor. These results align well with the qualitative comments from Experiment 1, where multiple participants identified vocals and volume as dimensions of music influenced that influenced their perceptions of how distracting the music was.

The classroom based methodology used in Experiment 3 successfully increased the efficiency of the experiment in terms of the amount of data collected with respect to the time spent collecting that data. The experiment was performed during two practical classes lasting two hours each, with only one hour in each class spent on my experiment. During this time, data was collected from 55 participants. Due to the thorough piloting process, the data from this experiment can be considered safe, so running a large scale controlled experiment using this approach was much more efficient than a laboratory based experiment. However, the sample of participants was restricted only to the students who were taking the module. All of the participants reported they were in the 18 to 24 age group, but it is highly likely the majority of participants were at the lower end of the age range. The participants were undergraduate students so were likely to have spent fewer years in education than the participants in Experiments 1 and 2, many of whom were postgraduate students or researchers. The practical classes when the experiment was performed was scheduled for mid-way through the students' first term at University, so the amount of typing experience, the number of essays, the number of hours spent programming at their computers, etc. is likely to be lower than for postgraduate students and probably explains why only 11 of the 55 participants were able to be classified as fast typists. Despite increasing the size of the overall sample, the number of participants classified as fast typists was actually smaller than in either Experiments 1 or 2 which had 12 and 20 fast typists, respectively. The impact of the relatively small number of participants able to be classified as fast typists was not helped by the additional between participants volume independent variable (IV) which sub-divided the group further. However, the fact that there were still omnibus effects for vocals despite the small groups of fast typists adds weight to my hypothesis that there is an effect for all typists, but the effect is more pronounced for the fast typists than the slow. The omnibus effect of vocals would have been unlikely to reach significance otherwise.

The interaction between text and vocals was also an interesting result as although it did not reach the required p value for significance for either typing speed or typing accuracy measures, the post hoc simple effects analysis after the trends were identified, resulted in a significant simple effect of vocals for both performance dependent variables (DVs). Intriguingly, for typing speed the significant effect of vocals was in the English transcription tasks only and not the Dutch transcription tasks, while for typing accuracy the significant effect of vocals was in the Dutch tasks with only a trend towards significance in the English tasks. In the English tasks, speed was higher when accompanied by music that did not contain vocals and in the Dutch tasks, there were fewer errors with music without vocals. These results mean that my hypothesis that performance in the hard tasks would not be affected by the music as they required high levels of concentration was incorrect.

One area to consider is whether the Dutch transcription task is too hard, to the extent that the participants stop correcting the errors they are making in their input text because they do not notice them as they are typing, because the orthographic conventions and lexical structures of the Dutch language are so vastly different to the English language. This led me to question whether it would be possible to make the task harder whilst still retaining the orthographic features of English. In this experiment the Dutch transcription task was replaced with a hard task that is in English but includes randomly chosen and inserted pseudowords. Pseudowords are non-English words that have no meaning but can be read as they conform to the English rules of word construction. Essentially, the objective of this task manipulation is to increase the difficulty of the transcription typing task while still allowing participants to read the text, even if they cannot understand its meaning.

7.1 METHOD

7.1.1 AIMS

The aims of this experiment were:

1. to gather further evidence that the presence of vocals in a piece of music has a negative effect on typing performance and experience,
2. to see if task difficulty affects the results,
3. to see if any identified effects of music are the same at different volume levels,
4. to gather further evidence that the effect of music on typing performance and experience differs between slow and fast typists.

7.1.2 PARTICIPANTS

Both the participants and experimenters were first year undergraduates studying Computer Science at the University of York, taking the module entitled *Human Aspects of Computer Science*, in 2014 and 2015. The experiment was performed across two years to increase the sample size. The total number of participants from both cohorts together was 116 (17 female, 96 male). The age and gender of the participants separated into the two cohorts is shown in Table 7.1.

Year	**Age**					**Gender**	
	< 18	**18 – 24**	**25 – 34**	**35 – 44**	**> 45**	**Male**	**Female**
2014	1	53	1	0	0	46	9
2015	0	60	0	1	0	53	8

Table 7.1: Participant ages and genders

Out of the 116 participants, 113 were in the 18 to 24 age group. Again, it is likely that the majority of the participants would be towards the lower end of this range as they were first year undergraduate students. In the 2015 cohort, 11 of the participants were non-native speakers of English, and in the 2014 group, 4 participants were non-native English speakers. All participants had demonstrated competency in English by achieving International English Language Test (IELT) scores in excess of 6.5 which was a requirement for their acceptance onto the degree programme. The participants were not asked to report whether they had a hearing disability or dyslexia as this would have been inappropriate given the classroom context. After the experiment, five participants from each year group were randomly selected to win £10 Amazon gift vouchers. The three fastest typists in each year received £30, £20 and £10 Amazon gift vouchers. This approach was taken to encourage the students to participate fully in the experiment and to try to achieve strong typing performance.

7.1.3 DESIGN

The experiment was performed over two years. The first group of students performed the tasks accompanied by low volume music and the second cohort were accompanied by the same music but played at a high volume level. This approach, by separating the experiment into two volume groups, means the analysis could be performed on the data from the two sessions separately. So, even with the speed classification included as a between participants IV (which is expected given the results from previous experiments) the analysis would only have one between groups IV rather than two, which would increase the complexity of the analysis considerably.

This approach leads to a fully factorial 2 by 2 repeated measures design, before inclusion of the post hoc speed classification, which is performed twice. The repeated measures IVs were vocals (two levels: without and with vocals) and intended task difficulty (two levels: medium and hard difficulty). The medium difficulty tasks used the advanced English text passages with the hard difficulty tasks using passages where the advanced English text had pseudowords randomly inserted throughout the text, as described in the section on Presented Text Independent Variables in Chapter 3, Section 3.3.1.1. Each participant completed four transcription typing tasks:

- with music that contained vocals using the advanced English text (vocals: with; intended task difficulty: medium),
- with music that did not contain vocals using the advanced English text (vocals: without; intended task difficulty: medium),

- with music that contained vocals using the advanced English with pseudowords text (vocals: with; intended task difficulty: hard),
- with music that did not contain vocals using the advanced English with pseudowords text (vocals: without; intended task difficulty: hard).

Figure 7.1 shows the arrangement of IVs used in this experiment. As this experiment used the classroom based methodology there was no without music condition because the environment could not be controlled sufficiently to eliminate possible confounds. The alt rock style music was used again to enable a direct comparison with the results from Experiment 3.

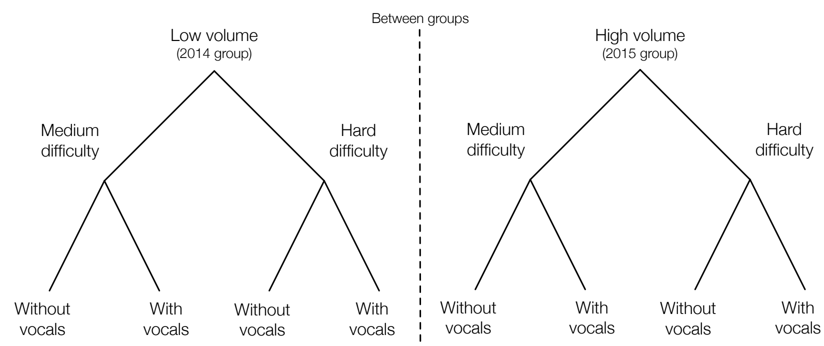


Figure 7.1: Structure of IVs for experiment

To avoid practice and fatigue effects, the order of presentation of tasks was counterbalanced.

The DVs used in this experiment were typing performance in terms of both typing speed (CPM) and accuracy (error rate percentage). Typing experience was measured using 7-point Likert items for perceived task difficulty and music distraction. More explanation of the DVs can be found in Chapter 3, Section 3.3.1.2.

7.1.4 Equipment and Materials

The experiment was performed using the classroom based methodology with bespoke websites created to collect the data and to guide and control the experiment. The experiment took place in an undergraduate laboratory in the Department of Computer Science at the University of York.

Participants used a desktop PC running Windows 7 with Firefox version 42. Inexpensive headphones were used (Astro Tools ATA 1144). The alt rock style music both with and without vocals was used as music stimuli. The transcription typing tasks used the advanced English and advanced English with pseudowords texts (see Chapter 3, Section 3.3.1.1 for more information about the music and text IVs).

Short paper based questionnaires were used to collect subjective ratings of perceived task difficulty and music distraction (see Appendix B.4). The participants also completed a demographics questionnaire (see Appendix B.5).

7.1.5 Procedure

Participants in the 2014 cohort were assigned to the low volume group. Participants in the 2015 sessions were assigned to the high volume group. The procedure for this experiment exactly followed that described in Chapter 3, Section 3.3.3.2. All participants completed four 4.5 minute transcription typing tasks, so were typing for a total of 18 minutes during the experiment.

7.2 Results

The first stage in the analysis of the data involved defining an appropriate speed classification and assigning participants to the slow or fast typists group. CPM data was taken from the advanced English (medium difficulty) task conditions across both volume sessions to achieve a consistent approach to the classification and allow direct comparisons between the two volume groups. The outcomes from the speed classification process are included in Section 7.2.1.

The inferential analysis of the data from the low volume group is presented first (Section 7.2.2), followed by the analysis of the high volume group's data (Section 7.2.3). Finally, a comparison of the results from both of the volume groups is provided (Section 7.2.4). This approach to presenting the analysis of the data collected in this experiment was taken for a few reasons. First, the separation of the analysis into the two different volume groups helps to simplify the process and improves clarity of the results which helps to support more easier interpretation of the implications of the results from this experiment. Further, as the data was collected with a year between the cohorts an unintended confound may have been introduced as I had taught the 2014 class as their lecturer but had not been involved in teaching the 2015 cohort. Finally, any issues of heteroscedasticity in the inferential analysis were fewer using this approach, and it was therefore easier to interpret the implications of these violations to the ANOVA assumptions when the data was analysed as two sets of two groups, rather than four groups within a single analysis.

The CPM data was analysed with inclusion of speed classification as a between groups factor as this has been shown to be very important in the other experiments. The error rate analysis was performed both with and without inclusion of the speed factor, but it was not expected to be an important factor affecting the results. The analysis of the experience ratings was performed both with and without the speed classification applied as the importance of this classification on experience ratings had been inconsistent in the previous experiments.

7.2.1 Speed Classification

The first important step in the analysis is to perform a speed classification. In this experiment, data was discarded from three participants due to strong indications they had not engaged fully in the experiment. Participant 1 from the low volume group did not complete two of the four transcription tasks. In the two tasks they did complete the participant achieved less than 30 CPM, i.e. less than 6 words per minute (WPM). Participant 5, again from the low volume group, only completed three of the four tasks, all with less than 2 WPM which is again extremely low. As both these participants did not complete all tasks and achieved very poor performance in the tasks that were submitted their data was discarded from the sample. The data from participant 9 from the high volume group was also discarded as they made no attempt to transcribe the presented text, instead their task submissions for all conditions included frequent use of profanities and insults. The data from these three participants was discarded from the sample prior to any of the descriptive or inferential statistics presented in the following sections.

The speed classification was performed by creating a histogram of CPM data from all the tasks using the advanced English text from the low and high volume groups combined, (Figure 7.2).

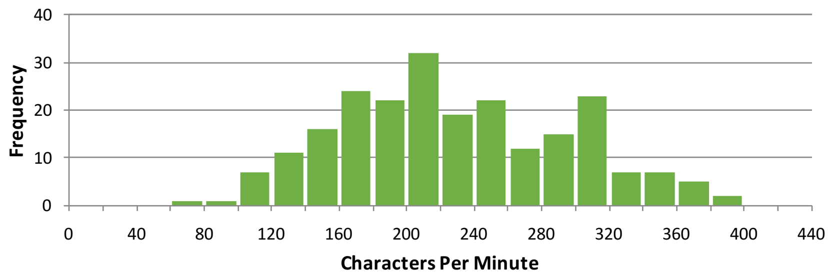


Figure 7.2: Histogram of CPM for the medium difficulty texts
n = 226, M = 228.81, SD = 68.20

The first thing to note from this distribution is that the highest speed achieved by any participant was less than 400 CPM, which contrasts with the laboratory based experiments where multiple participants achieved in excess of 400 CPM. But, aligns with the data from Experiment 3, where only a single participant exceeded 400 CPM.

The bimodal characteristic is less obvious in this distribution than those from the other experiments. If there is a bimodal distribution, then the cut-off point between the two distributions appears to be at approximately 275 CPM, which is lower than the 330 CPM threshold used in Experiments 1, 2 and 3.

A scatterplot of CPM was created (Figure 7.3) to explore whether grouping the typists into slow and fast groups based on a threshold of 275 CPM would be suitable.

The scatterplot verifies that a 275 CPM threshold is appropriate. Other values were explored as possible threshold levels, e.g. 250 CPM and 340 CPM, but these values did not align well with the cut-off point identified in the histogram and also led to a number of participants being difficult to classify as their CPM values for the different conditions sat either side of the threshold.

With the 275 CPM threshold, participant 37 from the low volume group (P37L in the scatterplot) was classified as a slow typist despite achieving 285.33 CPM in the medium task with vocals condition which is above the 275 CPM threshold. This decision was taken as their CPM in the medium task without vocals condition was 261.33 CPM which is low enough that it would have been inappropriate to classify them as a fast typist.

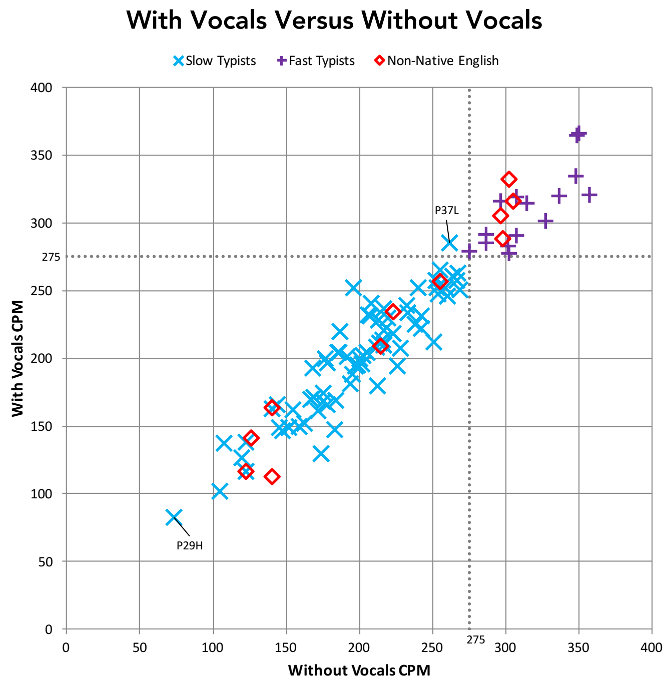


Figure 7.3: Scatterplot of CPM for the medium difficulty texts

Participant 29 from the high volume group (identified as P29H in the scatterplot) achieved less than 90 CPM in both conditions. This CPM equates to 18 WPM which is again low but given the large variance in the CPM distribution may not be an outlier. This participant was a native English speaker and inspection of their submitted text files indicates that these are genuine data points that should be retained in the analysis.

The analysis of the data collected in this experiment proceeded using the classification shown in Figure 7.3. This classification resulted in 82 participants designated as slow typists and 31 as fast typists. The analysis of the data collected from the low volume group is presented first, followed by the data collected from the cohort that were exposed to high volume music. Finally, a comparative analysis using data from both groups combined is included.

7.2.2 Analysis of the Low Volume Conditions

The analysis began by exploring the effect of vocals and task difficulty manipulation on typing performance and experience for the low volume group.

7.2.2.1 Typing Performance

Typing performance was measured using CPM and error rate. The CPM data was analysed only with the speed classification applied, as previous experiments have shown the classification to be very important in the analysis of CPM. The error rate data was analysed both with and without the speed classification used. Box plots were created to identify outliers. Shapiro-Wilk's tests and skewness and kurtosis characteristics were analysed before inferential analysis to assess the distributions for normality. The results from the normality tests and box plots can be found in Appendix G.1. Outliers which affected the results of the analysis were capped following the procedure in Chapter 3, Section 3.5.1.3.

Characters Per Minute

None of the eight distributions of CPM with the speed classification applied deviated significantly from normality and there were no outliers. Table 7.2 shows descriptive statistics for all typists together, and with separation of participants into fast and slow typist groups.

Text Condition	Vocals Condition	All Typists (N = 53)		Slow Typists (n = 41)		Fast Typists (n = 12)	
		M (CPM)	SD (CPM)	M (CPM)	SD (CPM)	M (CPM)	SD (CPM)
Medium Difficulty	Without	229.01	62.14	202.52	39.60	319.52	31.59
	With	231.92	67.76	202.56	41.07	332.24	37.17
Hard Difficulty	Without	198.61	62.70	172.73	41.06	287.04	37.80
	With	200.68	59.76	174.81	34.42	289.07	39.45

Table 7.2: Descriptive statistics of CPM

A fully factorial 2 by 2 by 2 mixed ANOVA was performed. The between participants IV was speed (slow or fast) with two within participants IVs, the presence of vocals in the music (with and without) and intended task difficulty (medium and hard). All of the assumptions for mixed ANOVAs were met.

Table 7.3 shows the results from the mixed ANOVA.

Test	F	df	p	η_p^2	Figure
Vocals	4.024	1, 51	0.050**	0.073	-
Vocals x Speed*	2.253	1, 51	n.s.	-	-
Text	177.824	1, 51	**< 0.0005**	0.777	-
Text x Speed*	3.283	1, 51	0.076**	0.060	7.4
Text x Vocals	1.303	1, 51	n.s.	-	-
Text x Vocals x Speed*	2.821	1, 51	0.087**	0.052	7.5
Speed*	96.495	1, 51	**<0.0005**	0.654	-

Table 7.3: Inferential analysis of CPM
(* between groups factor, ** non-significant result where $p < 0.1$)

The omnibus effect of vocals was extremely close to significance, $F(1, 51) = 4.024$, $p = 0.050$, $\eta_p^2 = 0.073$. Contrary to the hypothesis that vocals would reduce performance, CPM without vocals was lower than CPM with vocals (without vocals: $M = 213.81$, $SD = 63.97$; with vocals: $M = 216.30$, $SD = 65.49$).

The omnibus effect of text was significant, $F(1, 51) = 177.824$, $p < 0.0005$, $\eta_p^2 = 0.777$. CPM was higher in the medium difficulty tasks and lower in the hard tasks (medium: $M = 230.47$, $SD = 64.72$; hard: $M = 199.64$, $SD = 64.72$).

Speed classification was a significant between groups factor, $F(1, 51) = 96.495$, $p < 0.0005$, $\eta_p^2 = 0.654$. The fast typists' CPM was higher than the slow typists' (fast: $M = 306.97$, $SD = 40.52$; slow: $M = 188.15$, $SD = 41.38$).

There was a trend towards a significant interaction between text and speed group, $F(1, 51) = 3.283$, $p = 0.076$, $\eta_p^2 = 0.060$, (Figure 7.4).

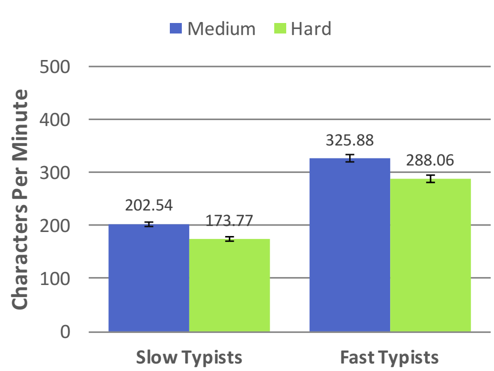


Figure 7.4: Trend towards a significant interaction between text and speed group ($p = 0.076$)

The simple main effect of text was significant for both slow, $F(1, 40) = 152.933$, $p < 0.0005$, $\eta_p^2 = 0.793$, and fast typists, $F(1, 11) = 64.461$, $p < 0.0005$, $\eta_p^2 = 0.854$. For both speed groups, CPM was higher in the medium difficulty tasks and lower in the hard tasks (slow medium: M = 202.54, SD = 40.09; slow hard: M = 173.77, SD = 37.67; fast medium: M = 325.88, SD = 34.35; fast hard: M = 288.06, SD = 37.80). As both the simple main effects are significant, the significant interaction is probably due to the differences in the effect size between the two groups as the difference in CPM between the tasks is larger for the fast typists than for the slow typists (fast: Δ = 37.82; slow: Δ = 28.77); .

The three-way interaction between text, vocals and speed group approached significance, $F(1, 51) = 2.821$, $p = 0.087$, $\eta_\rho^2 = 0.052$, (Figure 7.5).

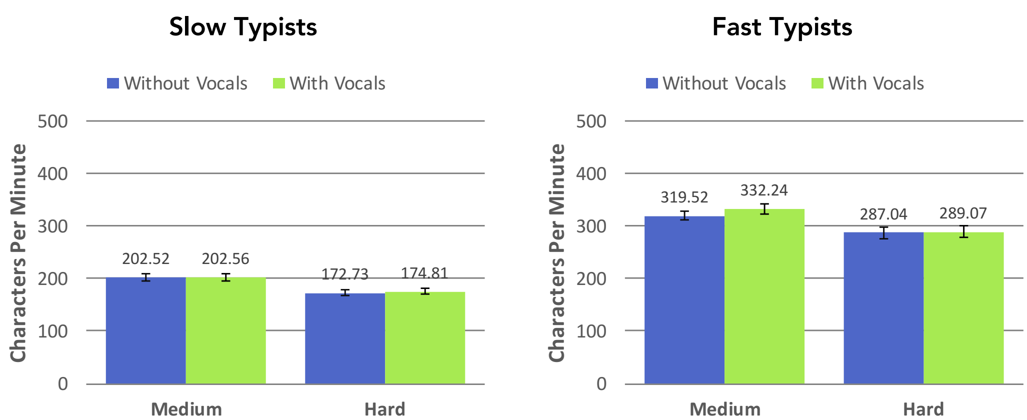


Figure 7.5: Trend towards a significant interaction between text, vocals and speed group (p = 0.087)

The simple interaction between vocals and text was not significant for either the slow, $F(1, 40) = 0.339$, *n.s.*, or fast typists, $F(1, 11) = 0.173$, *n.s.* However, visual inspection of Figure 7.5 suggests that, in this case, conceptualising this three-way interaction in terms of differences across the two speed groups may not have been the most appropriate approach. So, to further explore why the three-way interaction between text, vocals and speed group approached significance it was reconceptualised. Instead of exploring how the interaction between text and vocals differs across the two speed groups, the simple effects analysis was repeated with a focus on whether the interaction between vocals and speed group differs between the two texts.

The simple interaction between vocals and speed group was not significant in the hard text, $F(1, 51) = 0.000$, *n.s.*, but the simple interaction between vocals and speed group was significant for the medium difficulty tasks, $F(1, 51) = 5.953$, $p = 0.018$, $\eta_\rho^2 = 0.105$. The simple simple main effect of vocals in the medium difficulty tasks was significant for fast typists, $F(1, 11) = 11.543$, $p = 0.006$, $\eta_\rho^2 = 0.512$, but not for slow typists, $F(1, 40) = 0.000$, *n.s.* The fast typists' CPM was higher with vocals than without vocals in the medium difficulty tasks (without vocals: M = 332.24, SD = 37.17; with vocals: M = 319.52, SD = 31.59).

Error Rate

The previous experiments showed the error rate percentage data had a strong positive skew and a logarithmic transformation should be applied before analysis. The experiments also showed that speed classification is not always important when it comes to the interpretation of results from the analysis of the transformed error rate data. So, the analysis of the typing accuracy data began by inspecting logarithmic transformations of the error rate data without the speed classification applied with an analysis including the speed classification performed later.

The logarithmically transformed error rate data included a high number of outlier data points across all the conditions. These outliers affected the results of the inferential analysis, so the values were capped, following the procedure described in Chapter 3, Section 3.5.1.3. Two of the four distributions (50%) had strong deviations from normality after capping. However, the sample size is large with 53 participants in a single group, so these deviations to normality should not affect the results of the analysis.

Table 7.4 shows descriptive statistics of transformed error rate percentage.

Low Volume							
Text Condition	**Vocals Condition**	**All Typists** (N = 53)		**Slow Typists** (n = 41)		**Fast Typists** (n = 12)	
		M (%)	SD (%)	M (%)	SD (%)	M (%)	SD (%)
Medium Difficulty	Without	-0.113	0.37	-0.094	0.37	-0.175	0.36
	With	-0.046	0.34	-0.038	0.33	-0.074	0.38
Hard Difficulty	Without	0.032	0.46	0.057	0.52	-0.051	0.18
	With	-0.100	0.32	-0.083	0.34	-0.157	0.26

Table 7.4: Descriptive statistics of transformed error rate percentage

The inferential analysis was performed using a 2 by 2 repeated measures ANOVA, with vocals and text condition as the IVs. Table 7.5 shows the results from the repeated measures ANOVA. Aside from the normality deviations discussed earlier, there were no other assumption violations.

Test	F	df	p	η_p^2	Figure
Vocals	0.615	1, 52	n.s.	-	-
Text	1.334	1, 52	n.s.	-	-
Text x Vocals	6.037	1, 52	**0.017**	0.104	7.6

Table 7.5: Inferential analysis of transformed error rate

The omnibus effect of vocals was not significant, $F(1, 52) = 0.615$, *n.s.*, and neither was the omnibus effect of text, $F(1, 52) = 1.334$, *n.s.* However, the two-way interaction between text and vocals was significant, $F(1, 52) = 6.037$, $p = 0.017$, $\eta_\rho^2 = 0.104$, (Figure 7.6).

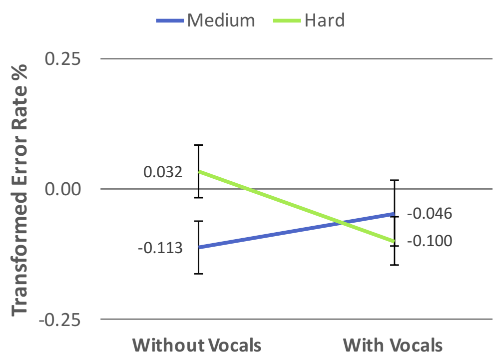


Figure 7.6: Significant interaction between text and vocals ($p = 0.017$)

The simple main effect of vocals was not significant in the medium difficulty tasks, $F(1, 52) = 1.237$, *n.s.*, but was significant in the hard tasks, $F(1, 52) = 5.394$, $p = 0.024$, $\eta_\rho^2 = 0.094$. In the hard tasks, the transformed error rate was higher without vocals in the music and lower with vocals (without vocals: $M = 0.032$, $SD = 0.46$; with vocals: $M = -0.100$, $SD = 0.32$).

The simple main effect of text was not significant in tasks accompanied by music with vocals, $F(1, 52) = 1.328$, *n.s.*, but was significant without vocals in the music, $F(1, 52) = 4.919$, $p = 0.031$, $\eta_\rho^2 = 0.086$. The transformed error rate was higher in the hard tasks and lower in the medium difficulty tasks when accompanied by music without vocals (hard: $M = 0.032$, $SD = 0.46$; medium: $M = -0.113$, $SD = 0.37$).

The analysis was then performed with the speed classification applied. There were significant violations of the assumption of homogeneity of variance for both of the hard tasks and additional outliers were introduced, which were again capped for the analysis. The results of the mixed ANOVA are included in Appendix G.2.1. This analysis did not add any significant effects or interactions involving the speed classification. Speed classification itself was also not a significant between groups factor, $F(1, 51) = 0.672$, *n.s.*

7.2.2.2 Typing Experience

Typing experience was measured in terms of perceived task difficulty and music distraction ratings. The previous experiments had shown inconsistencies in how the speed classification affected the ratings of perceived task difficulty and music distraction. So, both analyses were performed with and without the typing classification applied.

The analysis of the Likert items data was performed following the procedure described in Chapter 3, Section 3.5.3, by first using ANOVAs despite meaningful assumption violations and following up significant omnibus effects or interactions with an appropriate non-parametric test.

Task Difficulty

Participants rated how difficult they found each task using 7-point Likert items ranging from 1 = extremely difficulty to 7 = extremely easy. The mapping of rating to description was reversed for analysis so that easier tasks had a lower numeric value.

Table 7.6 shows descriptive statistics of task difficulty ratings for all typists, including upper and lower quartiles, median and mean, labelled *Q1*, *Q3*, *Mdn* and *M* respectively, Table 7.7 separates the descriptive statistics into the slow and fast typist groups.

Low Volume					
Text Condition	**Vocals Condition**	**All Typists** (N = 53)			
		Q1	*Mdn*	*Q3*	*M*
Medium Difficulty	Without	2.00	3.00	4.00	3.30
	With	2.00	3.00	4.00	3.28
Hard Difficulty	Without	4.00	5.00	6.00	4.81
	With	4.00	5.00	6.00	4.77

Table 7.6: Descriptive statistics of task difficulty ratings

Low Volume									
Text Condition	**Vocals Condition**	**Slow Typists** (n = 41)				**Fast Typists** (n = 12)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
Medium Difficulty	Without	2.00	3.00	4.00	3.34	2.00	3.00	4.00	3.17
	With	2.50	3.00	4.00	3.41	2.00	2.50	3.75	2.83
Hard Difficulty	Without	4.00	5.00	6.00	4.85	4.00	5.00	6.00	4.67
	With	4.00	5.00	6.00	4.95	3.25	4.00	5.00	4.17

Table 7.7: Descriptive statistics of task difficulty ratings

The data was first analysed as if coming from a single group using a repeated measures ANOVA. Table 7.8 shows the results from this analysis.

Test	**F**	**df**	**p**
Vocals	0.037	1, 52	n.s.
Text	99.216	1, 52	**< 0.0005**
Text x Vocals	0.011	1, 52	n.s.

Table 7.8: Inferential analysis of task difficulty ratings

The omnibus effect of text was significant, $F(1, 52) = 99.216$, $p < 0.0005$. This result was verified by restructuring the data so that rather than four columns representing the two vocals conditions and two text conditions. The vocals condition was collapsed for this analysis so that there were two columns for the two task conditions. A Wilcoxon signed rank test was then performed to compare the effect of text on the task difficulty ratings. The text had a significant effect on task difficulty rating, $z = -7.683$, $p < 0.0005$. The medium difficulty tasks were perceived as easier than the hard tasks (medium: *Mdn* = 3.00; hard: *Mdn* = 5.00).

The omnibus effect of vocals from the repeated measures ANOVA was not significant, $F(1, 52) = 0.037$, *n.s.*, and neither was the interaction between text and vocals, $F(1, 52) = 0.011$, *n.s.* So, neither of these aspects of the analysis were taken further.

The speed classification was then included as a between groups factor in a mixed ANOVA. The results from this test are included in Appendix G.2.1. There were no additional significant effects or interactions involving the speed classification, and it was not a significant between groups factor, $F(1, 51) = 2.148$, *n.s.*

Music Distraction

Music distraction ratings were taken using Likert items ranging from 1 = extremely distracting to 7 = not at all distracting. The mapping was reversed for analysis so that lower ratings indicated less distracting music. Tables 7.9 and 7.10 show descriptive statistics of music distraction ratings for all typists and with speed classification applied.

Low Volume					
Text Condition	**Vocals Condition**	**All Typists** (N = 53)			
		Q1	*Mdn*	*Q3*	*M*
Medium Difficulty	Without	1.50	2.00	4.00	2.81
	With	2.00	3.00	4.50	3.11
Hard Difficulty	Without	2.00	3.00	4.00	3.17
	With	2.00	3.00	5.00	3.49

Table 7.9: Descriptive statistics of music distraction ratings

Low Volume									
Text Condition	**Vocals Condition**	**Slow Typists** (n = 41)				**Fast Typists** (n = 12)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
Medium Difficulty	Without	2.00	3.00	4.00	2.88	1.00	2.00	4.50	2.58
	With	2.00	3.00	5.00	3.37	1.00	2.00	3.00	2.25
Hard Difficulty	Without	2.00	3.00	4.00	3.20	2.00	2.00	4.00	3.08
	With	2.50	4.00	5.00	3.80	2.00	2.00	3.75	2.42

Table 7.10: Descriptive statistics of music distraction ratings

A 2 by 2 repeated measures ANOVA was performed, with any significant effects followed up with the appropriate non-parametric test. Table 7.11 shows the results from the repeated measures ANOVA.

Test	F	df	p
Vocals	2.629	1, 52	n.s.
Text	15.825	1, 52	**< 0.0005**
Text x Vocals	0.010	1, 52	n.s.

Table 7.11: Inferential analysis of music distraction ratings

The omnibus effect of text was significant, $F(1, 52) = 15.825$, $p < 0.0005$. This result was followed up with a non-parametric comparison of the music distraction ratings in the medium and hard difficulty tasks using a Wilcoxon signed rank tests, which was significant, $z = -3.911$, $p < 0.0005$. The median and lower quartiles for distraction ratings in the medium and hard difficulty tasks were the same, but there were differences in both the Q3 and mean values which indicate the music was perceived as more distracting in the hard tasks, than in the medium difficulty tasks (hard: *Q1* = 2.00, *Mdn* = 3.00, *Q3* = 5.00, *M* = 3.33; medium: *Q1* = 2.00, *Mdn* = 3.00, *Q3* = 4.00, *M* = 2.96).

In the repeated measures ANOVA, the omnibus effect of vocals was not significant, *F*(1, 52) = 2.629, *n.s.*, and neither was the interaction between text and vocals, *F*(1, 52) = 0.010, *n.s.* Again, neither of these analyses were taken further.

An analysis was then performed including the speed classification as a between groups factor. Table 7.12 presents the results from this analysis.

Test	F	df	p
Vocals	0.012	1, 51	n.s.
Vocals x Speed*	5.697	1, 51	**0.021**
Text	10.171	1, 51	**0.002**
Text x Speed*	0.040	1, 51	n.s.
Text x Vocals	0.209	1, 51	n.s.
Text x Vocals x Speed*	0.971	1, 51	n.s.
Speed*	2.652	1, 51	n.s.

Table 7.12: Inferential analysis of music distraction ratings with speed classification
(* between groups factor)

The interaction between vocals and speed was significant, *F*(1, 51) = 5.697, *p* = 0.021. This result was explored further by performing Wilcoxon signed rank tests comparing the effect of vocals on music distraction ratings for the slow and fast typists separately.

The effect of vocals was significant for the slow typists, *z* = -3.751, *p* < 0.0005, *r* = 0.73, but not for the fast typists, *z* = -1.354, *n.s.* The slow typists perceived the music with vocals to be more distracting than the music without vocals (with vocals: *Mdn* = 4.00; without vocals: *Mdn* = 3.00).

The significant omnibus effect of text, *F*(1, 51) = 10.171, *p* = 0.002, was stronger in the analysis of the data without the speed classification applied, so was not taken further. All other effects and interactions were non-significant.

7.2.2.3 Summary of Results for Low Volume Group

The omnibus effect of vocals on typing speed was extremely close to reaching significance (*p* = 0.050). CPM was higher when the typing tasks were accompanied by music that contained vocals and lower with instrumental music.

Typing speed was significantly lower in the tasks that were intended to be harder and higher in the medium difficulty tasks. The typists classified as fast typists achieved significantly higher speeds than the slow typists.

The interaction between text and vocals was significant for transformed error rate. Using the hard text, the effect of vocals was significant with more errors when accompanied by music without vocals and fewer errors when the music contained vocals. The effect of vocals was not significant in the medium difficulty tasks. The effect of text was significant in the without vocals conditions, with higher error rates in the hard text conditions and lower error rates in the medium difficulty tasks.

The transformed error rate analysis was not affected by the typing speed classification and, in this case, there was no evidence of a difference in the effects on transformed error rate between the two typing speed groups.

The transcription typing tasks that were intended to be harder were also perceived as more difficult. There was no evidence that the presence of vocals in the music affected the task difficulty ratings, or that the speed classification affected the results.

The music distraction ratings were also affected by the difficulty of the task. The music was perceived as more distracting in the hard tasks than in the medium difficulty tasks. The speed classification did affect the results of the analysis of music distraction ratings as there was a significant interaction between vocals and speed group. The effect of vocals on music distraction ratings for the slow typists was significant, with the music without vocals perceived as more distracting than the music with vocals. The effect of vocals on music distraction ratings was not significant for the fast typists.

7.2.3 Analysis of the High Volume Group

The following sections present inferential analysis of the data collected from participants completing the transcription typing tasks accompanied by high volume music.

7.2.3.1 Typing Performance

Typing performance was again measured in CPM and error rate percentage. The CPM data was only analysed with the speed classification applied, but the error rate percentage data was analysed both with and without the speed classification. Shapiro-Wilk's tests, skewness and kurtosis were inspected before inferential analysis to assess the normality of the distribution. Box plots were created to identify outliers. The results from these tests and box plots are included in Appendix G.1.

Characters Per Minute

None of the distributions of CPM deviated significantly from normality and there were no outliers. Table 7.13 presents descriptive statistics of CPM.

High Volume							
Text Condition	**Vocals Condition**	**All Typists** (N = 53)		**Slow Typists** (n = 41)		**Fast Typists** (n = 19)	
		M (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)	*M* (CPM)	*SD* (CPM)
Medium Difficulty	Without	226.96	72.59	187.20	187.20	312.75	24.55
	With	227.72	70.73	189.25	189.25	310.74	26.11
Hard Difficulty	Without	196.46	67.45	158.57	158.57	278.21	24.14
	With	193.66	61.63	160.24	160.24	265.79	22.91

Table 7.13: Descriptive statistics of CPM

A fully factorial mixed ANOVA was performed. The between participants IV was typing speed (slow and fast) with two within participants IVs, the presence of vocals in the music (with and without) and intended task difficulty (medium and hard).

The assumption of homogeneity of variance was violated in all task conditions. Inspection of the *SD*s in Table 7.13 show that smaller variance is associated with the smaller group of fast typists which would lead to a conservative rather than liberal F statistic. So, the inferential analysis can proceed using mixed ANOVA despite this heteroscedasticity. Table 7.14 shows the results from the mixed ANOVA.

Test	F	df	p	η_p^2	Figure
Vocals	2.511	1, 58	n.s.	0.041	-
Vocals x Speed*	7.205	1, 58	**0.009**	0.110	7.7
Text	264.362	1, 58	**< 0.0005**	0.820	-
Text x Speed*	6.704	1, 58	**0.012**	0.104	7.8
Text x Vocals	3.556	1, 58	0.064**	0.058	7.90
Text x Vocals x Speed*	3.073	1, 58	0.085**	0.050	7.10
Speed*	122.163	1, 58	**< 0.005**	0.678	-

Table 7.14: Inferential analysis of CPM
(* between groups factor, ** non-significant result where $p < 0.1$)

The omnibus effect of vocals was not significant, $F(1, 58) = 2.511$, *n.s.* But, the omnibus effect of text was significant, $F(1, 58) = 264.362$, $p < 0.0005$, $\eta_\rho^2 = 0.820$. CPM was higher in the medium difficulty tasks and lower in the hard tasks (medium: $M = 227.34$, $SD = 71.37$; hard: $M = 195.06$, $SD = 64.35$).

Speed classification was a significant between groups factor, $F(1, 58) = 122.163$, $p < 0.0005$, $\eta_\rho^2 = 0.678$. The fast typists' CPM was higher than the slow typists' (fast: $M = 291.87$, $SD = 31.52$; slow: $M = 173.82$, $SD = 47.19$).

The two-way interaction between vocals and speed group was significant, $F(1, 58) = 7.205$, $p = 0.009$, $\eta_\rho^2 = 0.110$, (Figure 7.7).

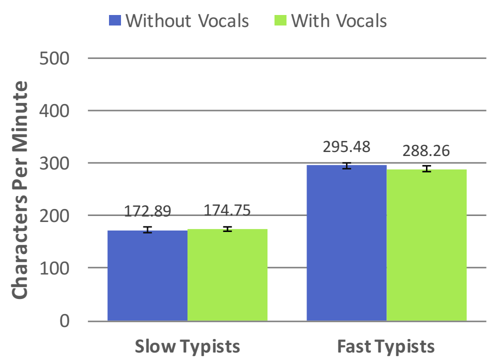


Figure 7.7: Significant interaction between vocals and speed group ($p = 0.009$)

The simple main effect of vocals was significant for fast typists, $F(1, 18) = 8.157$, $p = 0.010$, $\eta_\rho^2 = 0.312$, but not for slow typists, $F(1, 40) = 0.882$, *n.s.* The fast typists' CPM was higher without vocals in the accompanying music and lower with vocals, (without vocals: $M = 295.48$, $SD = 29.71$; with vocals: $M = 288.26$, $SD = 33.25$).

The two-way interaction between text and speed group was significant, $F(1, 58) = 6.704$, $p = 0.012$, $\eta_\rho^2 = 0.104$, (Figure 7.8).

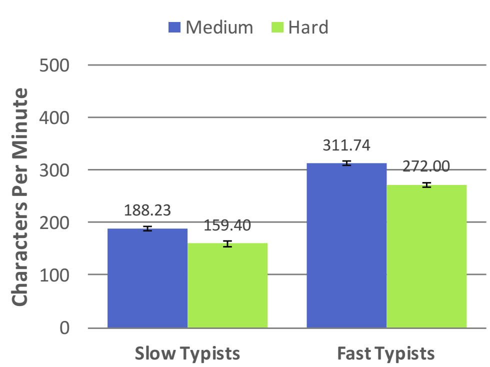


Figure 7.8: Significant interaction between text and speed group ($p = 0.012$)

The simple main effect of text was significant for both slow, $F(1, 40) = 153.519$, $p < 0.0005$, $\eta_\rho^2 = 0.793$, and fast typists, $F(1, 18) = 119.603$, $p < 0.0005$, $\eta_\rho^2 = 0.869$. In both groups, CPM was higher in the medium difficulty tasks than the hard tasks (slow medium: $M = 188.23$, $SD = 48.02$; slow hard: $M = 159.40$, $SD = 41.89$; fast medium: $M = 311.74$, $SD = 25.02$; fast hard: $M = 272.00$, $SD = 24.05$). The significant interaction is because the difference in CPM between the two vocals conditions is considerably larger for the fast typists than for the slow typists, (fast: ΔCPM = 39.74; slow: ΔCPM = 28.83).

The two-way interaction between text and vocals approached significance, $F(1, 58) = 3.556$, $p = 0.064$, $\eta_\rho^2 = 0.058$, (Figure 7.9).

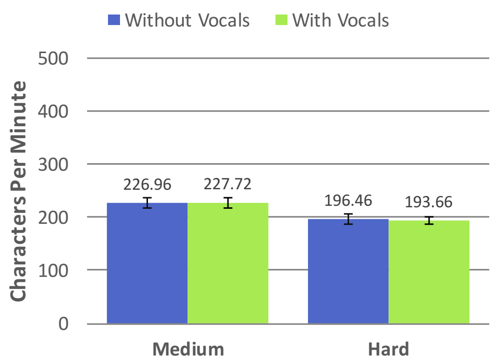


Figure 7.9: Trend towards a significant interaction between text and vocals ($p = 0.064$)

The simple main effect of vocals was significant in the hard tasks, $F(1, 58) = 8.476$, $p = 0.005$, $\eta_\rho^2 = 0.128$, but not the medium difficulty tasks, $F(1, 58) = 0.000$, *n.s.*[28] In the hard tasks, CPM was higher without vocals than with vocals (without vocals: $M = 196.46$, $SD = 67.45$; with vocals: $M = 193.66$, $SD = 61.63$).

Finally, the three-way interaction between text, vocals and speed group approached significance, $F(1, 58) = 3.073$, $p = 0.085$, $\eta_\rho^2 = 0.050$, (Figure 7.10).

The simple interaction between vocals and text was not significant for slow typists, $F(1, 40) = 0.017$, *n.s.*, but approached significance for fast typists, $F(1, 18) = 3.522$, $p = 0.077$, $\eta_\rho^2 = 0.164$. The simple simple main effect of vocals was significant in the hard tasks for fast typists, $F(1, 18) = 14.361$, $p = 0.001$, $\eta_\rho^2 = 0.444$, but not in the medium difficulty tasks, $F(1, 18) = 0.232$, *n.s.* CPM was higher without vocals and lower with vocals for the fast typists in the hard tasks (without vocals: $M = 278.21$, $SD = 24.14$; with vocals: $M = 265.79$, $SD = 22.91$).

28 In both these simple effects analyses, there were significant violations of the homogeneity of variance assumption for all groups. However, as discussed previously, these would lead to conservative rather than liberal *F* statistics.

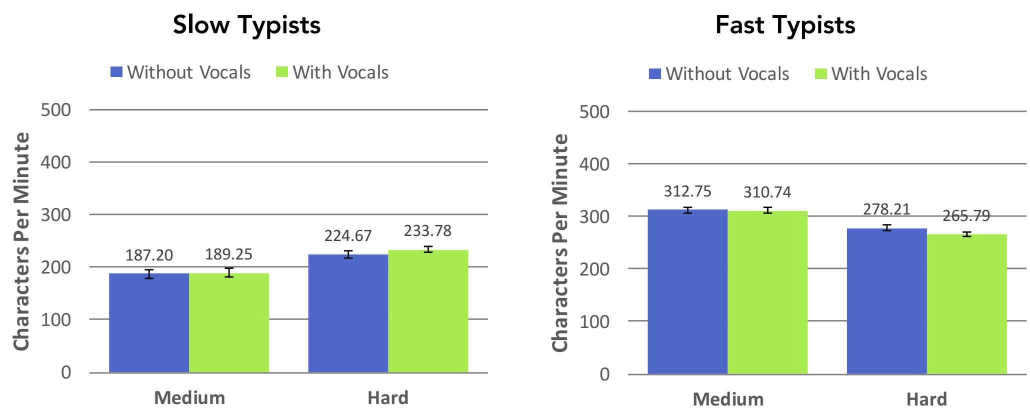


Figure 7.10: Trend towards a significant interaction between text, vocals and speed group (p = 0.085)

Visual inspection of Figure 7.10 suggests it may also be worth conceptualising the three-way interaction between text, vocals and speed group by looking at how the two-way interaction between vocals and speed group differs between the two texts, as well as considering how the interaction between text and vocals differs between the slow and fast typists.

The simple interaction between vocals and speed group was not significant in the medium difficulty tasks, $F(1, 58) = 0.644$, *n.s.*, but was significant in the hard tasks, $F(1, 58) = 14.558$, $p < 0.0005$, $\eta_\rho^2 = 0.201$[29]. The simple simple main effect of vocals was not significant for the slow typists, $F(1, 40) = 0.693$, *n.s.*, but was significant for the fast typists, $F(1, 18) = 14.361$, $p = 0.001$, $\eta_\rho^2 = 0.444$. The fast typists' CPM was significantly higher without vocals and lower with vocals in the music (without vocals: $M = 278.21$, $SD = 24.14$; with vocals: $M = 265.79$, $SD = 22.91$).

Error Rate

The error rate data was again transformed logarithmically. There were a number of outlying data points which affected the outcomes from the inferential analysis so were capped. After capping the two with vocals conditions deviated significantly from normality, but with a sample size of 60 the analysis should be robust to this assumption violation. Table 7.15 shows the descriptive statistics of transformed error rate (including outlier capping) for the transformed error rate percentage data from the high volume session.

---

29 In both these simple effects analyses, there were significant violations of the homogeneity of variance assumption for all groups. However, as discussed previously, these would lead to conservative rather than liberal F statistics.

High Volume								
Text Condition	Vocals Condition	All Typists (N = 60)		Slow Typists (n = 41)		Fast Typists (n = 19)		
		M (%)	SD (%)	M (%)	SD (%)	M (%)	SD (%)
Medium Difficulty	Without	0.089	0.45	0.199	0.47	-0.148	0.30
	With	0.178	0.44	0.227	0.46	0.074	0.38
Hard Difficulty	Without	0.151	0.43	0.215	0.47	0.011	0.26
	With	0.091	0.45	0.179	0.50	-0.099	0.19

Table 7.15: Descriptive statistics of transformed error rate percentage

The transformed error rate data was analysed using a 2 by 2 repeated measures ANOVA. After outlier capping, aside from the normality violations described earlier there were no assumption violations. Table 7.16 shows the results from the repeated measures ANOVA. There were no significant effects or interactions.

Test	F	df	p	η_p^2
Vocals	0.130	1, 59	n.s.	-
Text	0.719	1, 59	n.s.	-
Text x Vocals	0.113	1, 59	n.s.	-

Table 7.16: Inferential analysis of transformed error rate
(** non-significant result where p < 0.1)

The analysis was then performed with the speed classification applied. The results from this analysis are included in Appendix G.2.2. There were no additional significant effects and no significant interactions involving the speed classification. However, speed classification, was a significant between groups factor, $F(1, 58) = 7.869$, $p = 0.007$, $\eta_\rho^2 = 0.119$. The fast typists made fewer errors than the slow typists (fast: M = -0.041, SD = 0.30; slow: M = 0.205, SD = 0.47).

7.2.3.2 Typing Experience

Typing experience was established using measures of perceived task difficulty and music distraction. The inferential analysis was performed following the procedure described in Chapter 3, Section 3.5.3.

Task Difficulty

The participants rated how difficult they found each of the four transcription typing tasks using Likert items ranging from 1 = extremely difficult to 7 = extremely easy. The mapping was reversed for analysis so that lower ratings indicate tasks that are perceived as easier.

Table 7.17 presents descriptive statistics, including lower and upper quartiles, median and mean as *Q1*, *Q3*, *Mdn* and *M* respectively for all the typists combined. Table 7.18 includes separation into the two speed groups.

High Volume					
Text Condition	**Vocals Condition**	**All Typists** (N = 60)			
		Q1	*Mdn*	*Q3*	*M*
Medium Difficulty	Without	2.00	3.00	4.00	2.97
	With	2.00	3.00	4.00	3.03
Hard Difficulty	Without	3.00	5.00	5.75	4.48
	With	4.00	5.00	5.00	4.65

Table 7.17: Descriptive statistics of task difficulty ratings

High Volume									
Text Condition	**Vocals Condition**	**Slow Typists** (n = 41)				**Fast Typists** (n = 19)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
Medium Difficulty	Without	2.00	3.00	4.00	3.10	2.00	3.00	3.00	2.68
	With	2.00	3.00	4.00	3.02	2.00	3.00	4.00	3.05
Hard Difficulty	Without	3.50	5.00	5.50	4.59	3.00	4.00	6.00	4.26
	With	4.00	5.00	6.00	4.73	4.00	5.00	5.00	4.47

Table 7.18: Descriptive statistics of task difficulty ratings

Table 7.19 shows the results from a 2 by 2 repeated measures ANOVA.

Test	**F**	**df**	**p**
Vocals	0.970	1, 59	n.s.
Text	133.195	1, 59	**< 0.0005**
Text x Vocals	0.292	1, 59	n.s.

Table 7.19: Inferential analysis of task difficulty ratings

The omnibus effect of vocals was not significant, $F(1, 59) = 0.970$, *n.s.*, and neither was the interaction between text and vocals, $F(1\ 59) = 0.292$, *n.s.*

The omnibus effect of text was significant, $F(1, 59) = 133.195$, $p < 0.0005$. This result was verified with a non-parametric Wilcoxon signed rank test comparing the effect of text on the task difficulty ratings. The effect of text was significant, $z = -8.326$, $p < 0.0005$. The hard difficulty tasks were perceived as more difficult than the medium difficulty tasks (hard: *Mdn* = 5.00; medium: *Mdn* = 3.00).

The analysis was then performed with the speed classification applied, the results of this mixed ANOVA are included in Appendix G.2.2. There were no additional significant effects or interactions involving the speed classification.

Music Distraction

The participants also rated how distracting they found the music in each condition, using Likert items ranging from 1 = extremely distracting to 7 = not at all distracting. The mapping of rating to meaning was reversed for analysis so that low values indicate music that was perceived as less distracting.

Table 7.20 presents descriptive statistics of the music distraction ratings for all typists while Table 7.21 incorporates the typing speed classification.

High Volume					
Text Condition	**Vocals Condition**	**All Typists** (N = 60)			
		Q1	*Mdn*	*Q3*	*M*
Medium Difficulty	Without	1.00	2.00	4.00	3.02
	With	2.00	3.00	5.00	3.37
Hard Difficulty	Without	1.25	3.00	5.00	3.33
	With	2.00	3.50	5.00	3.43

Table 7.20: Descriptive statistics of music distraction ratings

High Volume									
Text Condition	**Vocals Condition**	**Slow Typists** (n = 41)				**Fast Typists** (n =19)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
Medium Difficulty	Without	1.00	3.00	4.00	3.02	1.00	2.00	5.00	3.00
	With	1.50	2.00	4.00	3.07	2.00	4.00	5.00	4.00
Hard Difficulty	Without	1.00	3.00	5.00	3.39	2.00	3.00	5.00	3.21
	With	2.00	3.00	5.00	3.32	2.00	4.00	5.00	3.68

Table 7.21: Descriptive statistics of music distraction ratings

Table 7.22 presents the results from the repeated measures ANOVA.

Test	F	df	p
Vocals	1.437	1, 59	n.s.
Text	2.735	1, 59	n.s.
Text x Vocals	1.545	1, 59	n.s.

Table 7.22: Inferential analysis of music distraction ratings

The omnibus effect of vocals was not significant, $F(1, 59) = 1.437$, *n.s.*, and neither was the omnibus effect of text, $F(1, 59) = 2.735$, *n.s.* The interaction between text and vocals was also not significant, $F(1, 59) = 1.545$, *n.s.*

As none of the omnibus effects, or the interaction were significant the analysis of music distraction ratings was not taken any further as a single group.

The results from the analysis of music distraction ratings with the speed classification applied are included in Table 7.23.

Test	F	df	p
Vocals	3.367	1, 58	0.072**
Vocals x Speed*	3.598	1, 58	0.063**
Text	1.044	1, 58	n.s.
Text x Speed*	2.098	1, 58	n.s.
Text x Vocals	2.242	1, 58	n.s.
Text x Vocals x Speed*	0.872	1, 58	n.s.
Speed*	0.325	1, 58	n.s.

Table 7.23: Inferential analysis of music distraction ratings with speed classification
(* between groups factor)

The main effect of vocals approached significance, $F(1, 58) = 3.367$, $p = 0.072$, but as the interaction between vocals and speed group also approached significance, $F(1, 58) = 3.598$, $p = 0.063$, it is worth further exploring the effect of vocals on music distraction ratings with the speed classification applied.

Wilcoxon signed rank tests revealed a significant effect of vocals for the fast typists, $z = -2.544$, $p = 0.011$, $r = 0.41$, but not for slow typists, $z = -0.74$, *n.s.* The fast typists perceived the music with vocals and more distracting than the music without vocals (with vocals: $Mdn = 4.00$, without vocals: $Mdn = 2.00$).

7.2.3.3 Summary of Results for High Volume Group

The omnibus effect of vocals was not significant for CPM, but the interaction between vocals and speed classification was significant. The simple main effect of vocals was significant for the fast typists, with higher CPM when accompanied by music that did not contain vocals and lower CPM when the music did contain vocals. The simple main effect of vocals was not significant for the slow typists.

In the analysis of typing speed there was a trend towards a significant interaction between text and vocals. The simple main effect of vocals was significant in the hard tasks but not the medium difficulty tasks. In the hard tasks, CPM was higher when accompanied by music that did not contain vocals and lower when the music did contain vocals.

The three-way interaction between text, vocals and speed classification also approached significance. The simple interaction between vocals and speed group was significant for the hard tasks but not the medium difficulty tasks. In the hard tasks, the fast typists' CPM was significantly slower with music that contained vocals, but there was no evidence of the same effect for the slow typists.

The omnibus effect of text on typing speed was significant, with higher CPM in the medium difficulty tasks than the hard tasks. The interaction between text and speed classification was also significant. The differences in CPM between the two difficulty levels was larger for the fast typists and smaller for the slow typists.

Speed classification was significant between groups factor for CPM. The fast typists' CPM was higher than the slow typists'.

In terms of typing accuracy, there were no significant effects or interactions for the high volume group. Although the speed classification did not alter the effects of music on typing accuracy, it was a significant between groups factor. The slow typists' error rate was higher than the fast typists'.

In terms of the experience of typing, there was a significant effect of text with the perception that the tasks that were intended to be harder were more difficult than the medium difficulty tasks. The speed classification had no effect on the analysis of task difficulty ratings.

The ratings of music distraction were not affected by either the text or the presence of vocals in the music. There was a trend towards a significant interaction between vocals and speed classification, so the effect of vocals was explored between the two speed groups. The effect of vocals was not significant for the slow typists, but was significant for the fast typists. The fast typists perceived the music with vocals as more distracting than the music without vocals.

7.2.4 Analysis of the Two Groups Together

The analysis presented shows that the effects of music on typing performance and experience differs quite considerably depending on whether the participants were exposed to low or high volume music. These contrasting results show that the volume of music is important in regulating how music affects typing performance and experience. In essence, by executing the analysis on the two volume groups separately a simple effects analysis has already been performed with a conceptual break between the two volume groups.

In this section, the results of the statistical analyses from the two volume groups are directly compared. Where there were differences between the two groups or an indication that by combining the data statistical significance may be reached (e.g. if both groups separately showed trends towards a significant result), the results from a statistical test which incorporates volume as a between participants factor are also reported.

7.2.4.1 Typing Performance

The comparison of results from the analysis of typing performance first considers the CPM analyses followed by the transformed error rate analyses.

Characters Per Minute

A mixed 2 by 2 by 2 by 2 ANOVA was performed on the CPM data combined. This analysis included two between groups IVs, volume (low or high) and typing speed classification (slow or fast), and two within participants IVs, with vocals (with and without) and text (medium and hard). The full results of this mixed ANOVA are included in Table 7.24.

Table 7.25 presents a comparison of results from the two separate volume group analyses, and includes the corresponding result from the mixed 2 by 2 by 2 by 2 ANOVA when the data from the two volume groups was combined.

The omnibus effect of vocals was not significant in the high volume condition but was extremely close to significance for the low volume group. So, I would expect the interaction between vocals and volume to be significant when the analysis is performed on the combined data. This prediction was true as the interaction between vocals and volume was indeed significant, $F(1, 109) = 6.611$, $p = 0.011$, $\eta_\rho^2 = 0.057$.

Test	F	df	p	η_p^2
Vocals	0.330	1, 109	n.s.	-
Vocals x Volume*	6.611	1, 109	**0.011**	0.057
Vocals x Speed*	0.264	1, 109	n.s.	-
Vocals x Volume* x Speed*	8.226	1, 109	**0.005**	0.070
Text	427.734	1, 109	**< 0.0005**	0.797
Text x Speed*	9.335	1, 109	**0.003**	0.079
Text x Volume*	0.091	1, 109	n.s.	-
Text x Vocals	4.301	1, 109	**0.040**	0.038
Text x Volume* x Speed*	0.082	1, 109	n.s.	-
Text x Vocals x Speed*	5.894	1, 109	**0.017**	0.051
Text x Vocals x Volume*	6.206	1, 109	n.s.	-
Text x Vocals x Volume* x Speed*	0.083	1, 109	n.s.	-
Volume*	3.294	1, 109	0.072**	0.029
Speed*	213.322	1, 109	**< 0.0005**	0.662
Volume* x Speed*	0.004	1, 109	n.s.	-

Table 7.24: Inferential analysis of CPM
(*between groups factor, ** non-significant result where p < 0.1)

Test	p value		
	Low Volume	High Volume	Combined
Vocals	0.050**	n.s.	n.s.
Vocals x Speed*	n.s.	**0.009**	n.s.
Text	**< 0.0005**	**< 0.0005**	**< 0.0005**
Text x Speed*	0.076**	**0.012**	**0.003**
Text x Vocals	n.s.	0.064**	**0.040**
Text x Vocals x Speed*	0.087**	0.085**	**0.017**
Speed*	**< 0.0005**	**< 0.0005**	**< 0.0005**

Table 7.25: Comparison of CPM inferential analysis
(* between groups factor, ** non-significant result where p < 0.1)

The omnibus effect of text was highly significant in both volume groups, so unsurprisingly the same result is identified when the analysis is performed on the data combined. The interaction between text and volume group was also not significant, $F(1, 109) = 0.091$, *n.s.*

Speed classification was a significant between groups factor at both volume levels. Again, unsurprisingly, when both groups were considered in the mixed ANOVA the speed classification was significant, $F(1, 109) = 213.322$, $p < 0.0005$, $\eta_p^2 = 0.662$, and the interaction between speed and volume group was not significant, $F(1, 109) = 0.0004$, *n.s.*

The two-way interaction between vocals and speed classification was significant in the high volume group, but not the low volume group. This should lead to a significant three-way interaction between vocals, speed classification and volume group from the combined mixed ANOVA. Again, this prediction is correct as the three-way interaction was significant, $F(1, 109) = 8.226$, $p = 0.005$, $\eta_p^2 = 0.070$, (Figure 7.11).

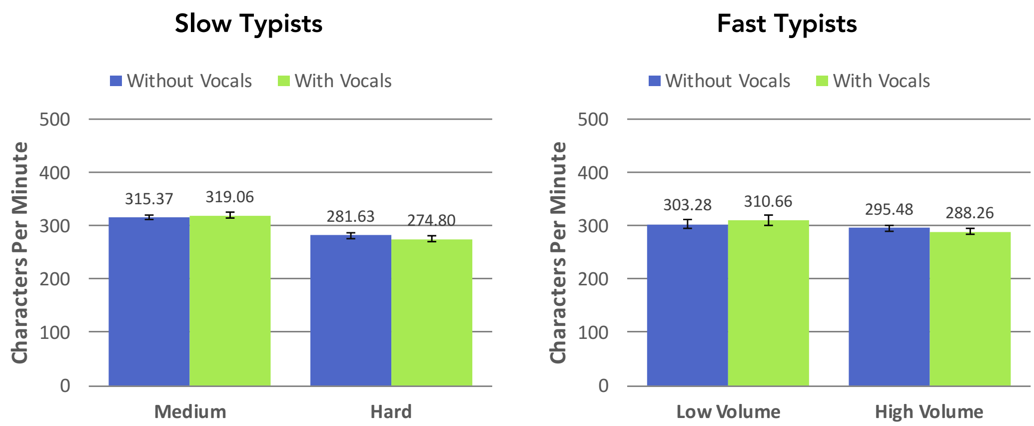


Figure 7.11: Significant interaction between vocals, volume and speed group ($p = 0.005$)

The previous analyses explored how the simple two-way interaction between vocals and speed differs between the two volume groups. As the speed classification has been important in the CPM data analyses throughout this thesis, the significant three-way interaction between vocals, volume and speed group from this mixed ANOVA on the combined data was reconceptualised to explore how the simple interaction between vocals and volume differs between the two speed groups.

The simple interaction between vocals and volume was not significant for the slow typists, $F(1, 80) = 0.784$, *n.s.*, but was significant for the fast typists, $F(1, 29) = 14.083$, $p = 0.001$, $\eta_p^2 = 0.327$. The simple simple main effect of vocals was significant for the fast typists in both the low, $F(1, 11) = 6.891$, $p = 0.024$, $\eta_p^2 = 0.385$, and high, $F(1, 18) = 8.157$, $p = 0.010$, $\eta_p^2 = 0.312$, volume groups. The low volume groups' CPM was higher with vocals and lower without vocals in the music (with vocals: $M = 310.66$, $SD = 43.49$; without vocals: $M = 303.28$, $SD = 24.05$). The high volume groups' CPM was lower with vocals and higher without vocals in the music (without vocals: $M = 295.48$ $SD = 29.71$; with vocals: $M = 288.26$, $SD = 33.25$). The simple simple main effect of volume was not significant without vocals, $F(1, 29) = 0.597$, *n.s.*, but was significant with vocals, $F(1, 29) = 4.577$, $p = 0.041$, $\eta_p^2 = 0.136$. The fast typists' CPM with vocals in the music was higher when played at a low volume than high volume (low volume: $M = 310.66$; high volume: $M = 288.26$).

Returning to Table 7.25, the two-way interaction between text and speed classification approached significance in the low volume group and achieved significance in the high volume group. These results mean that when a mixed ANOVA is performed on the combined data there are two possible outcomes. Either, the three-way interaction between text, speed classification and volume group is significant, indicating that the simple two-way interaction between text and speed classification differs between the two volume levels. Or, the three-way interaction is non-significant but the two-way interaction between text and speed classification strengthens. Here, it was the latter as in the combined mixed ANOVA the two-way interaction between text and speed classification became highly significant, $F(1, 109) = 9.335$, $p = 0.003$, $\eta_p^2 = 0.079$, (Figure 7.12).

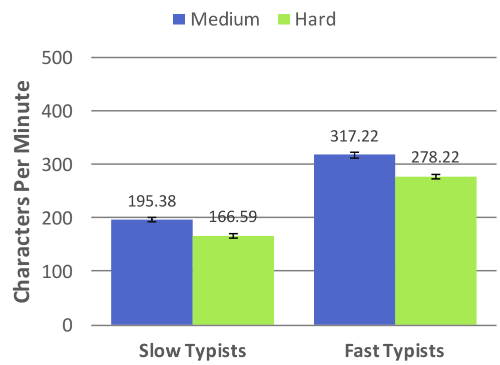


Figure 7.12: Significant interaction between text and speed group ($p = 0.003$)

The simple main effect of text was significant for both slow, $F(1, 80) = 306.452$, $p < 0.0005$, $\eta_p^2 = 0.793$, and fast typists, $F(1, 29) = 172.344$, $p < 0.0005$, $\eta_p^2 = 0.856$. Both groups of typists achieved higher CPMs in the medium difficulty tasks and lower CPMs with the hard tasks (slow medium: $M = 195.38$, $SD = 44.68$; slow hard: $M = 166.59$, $SD = 40.36$; fast medium: $M = 317.22$, $SD = 29.54$; fast hard: $M = 278.22$, $SD = 30.85$).

Again, returning to the comparison of results in Table 7.25, the two-way interaction between text and vocals did not reach significance for either volume group, but a trend towards significance was identified with high volume music. So, I would expect the three-way interaction between text, vocals and volume from the combined mixed ANOVA to be significant. However, this was not the case, as instead the interaction between text and vocals strengthened, $F(1, 109) = 4.301$, $p = 0.040$ $\eta_p^2 = 0.038$, (Figure 7.13).

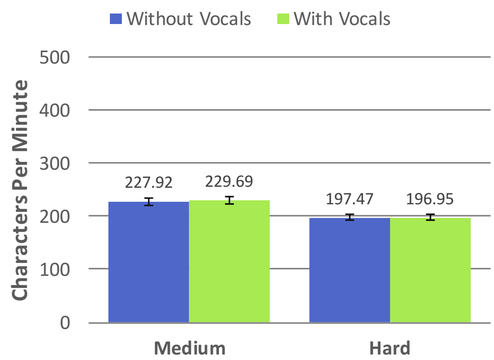


Figure 7.13: Significant interaction between text and vocals (p = 0.040)

The simple main effect of vocals approached significance in the medium difficulty texts, $F(1, 109) = 3.017$, $p = 0.085$, $\eta_p^2 = 0.027$[30]. CPM was lower in the without vocals conditions than the with vocals conditions (without vocals: $M = 227.92$, $SD = 67.90$; with vocals: $M = 229.69$, $SD = 69.08$). The simple main effect of vocals was not significant with the hard texts, $F(1, 109) = 0.932$, *n.s.*

Again, returning to Table 7.25, the three-way interactions between text, vocals and speed group approached significance but did not reach significance for either volume group alone. However, with the combined mixed ANOVA the power was increased and a significant three-way interaction between text, vocals and speed group reached significance, $F(1, 109) = 5.894$, $p = 0.017$, $\eta_p^2 = 0.051$ (Figure 7.14).

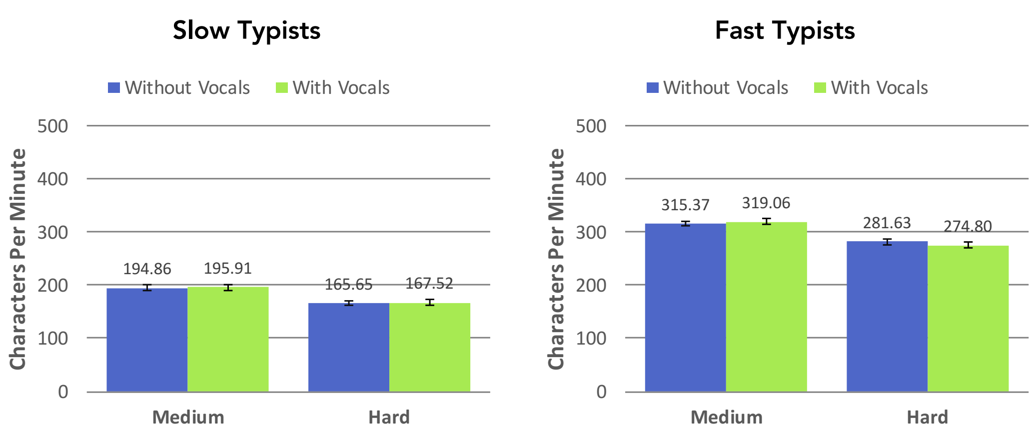


Figure 7.14: Significant interaction between text, vocals and speed group (p = 0.017)

The simple interaction between text and vocals was significant for the fast typists, $F(1, 29) = 5.391$, $p = 0.017$, $\eta_p^2 = 0.157$, but not for the slow typists, $F(1, 80) = 0.132$, *n.s.* For the

30 There was a violation of Levene's test for homogeneity of variance for both conditions which would lead to a conservative result. For medium text without vocals condition, $F(3, 109) = 4.093$, $p = 0.009$ and for the medium text with vocals, $F(3, 109) = 3.083$, $p = 0.030$.

fast typists, the simple simple main effect of vocals approached significance in both the medium, $F(1, 29) = 3.135$, $p = 0.087$, $\eta_\rho^2 = 0.098$, and hard, $F(1, 29) = 3.087$, $p = 0.089$, $\eta_\rho^2 = 0.096$, tasks. In the medium difficulty tasks, CPM was higher with vocals and lower without vocals (with vocals: $M = 319.06$, $SD = 32.08$; without vocals: $M = 315.37$, $SD = 27.18$). With the hard tasks, CPM was higher without vocals and lower with vocals in the music (without vocals: $M = 281.63$, $SD = 29.88$; with vocals: $M = 274.80$, $SD = 31.91$).

7.2.4.2 Error Rate

A 2 by 2 by 2 mixed ANOVA was performed with volume (low and high) as the single between groups factor and vocals (with and without) and text (medium and hard) as repeated measures. The full results from this mixed ANOVA are shown in Table 7.26.

Test	F	df	p	η_p^2
Vocals	0.100	1, 111	n.s.	-
Vocals x Volume*	0.665	1, 111	n.s.	-
Text	0.321	1, 111	n.s.	-
Text x Volume*	1.004	1, 111	n.s.	-
Text x Vocals	7.820	1, 111	**0.006**	0.066
Text x Vocals x Volume*	0.162	1, 111	n.s.	-
Volume*	10.032	1, 111	**0.002**	0.083

Table 7.26: Inferential analysis of transformed error rate percentage
(* between groups factor)

Table 7.27 shows a comparison of the results of the analyses of transformed error rate percentage between the two volume groups.

Test	p value		
	Low Volume	High Volume	Combined
Vocals	n.s.	n.s.	n.s.
Text	n.s.	n.s.	n.s.
Text x Vocals	**0.017**	n.s.	**0.006**

Table 7.27: Comparison of transformed error rate percentage inferential analysis
(* between groups factor, ** non-significant result where $p < 0.1$)

The two-way interaction between text and vocals was significant for the low volume group but was not significant for the high volume group. I would therefore probably expect that the interaction between text and vocals would get weaker in the combined analysis, but this was not the case as instead the p value from the two-way interaction between text and vocals was lower, $F(1, 111) = 7.820$, $p = 0.006$, $\eta_\rho^2 = 0.066$, (Figure 7.15).

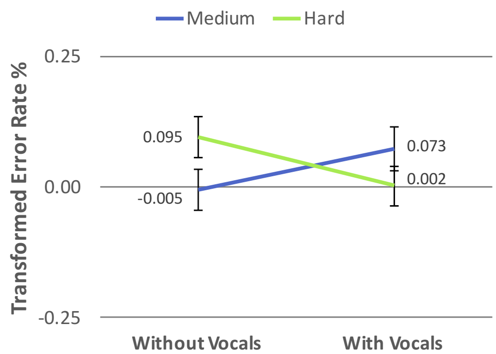


Figure 7.15: Significant interaction between text and vocals (p = 0.006)

The simple main effect of vocals approached significance in the medium difficulty tasks, $F(1, 109) = 3.142$, $p = 0.079$, $\eta_\rho^2 = 0.028$. The transformed error rate was lower without vocals in the music and higher when the music did contain vocals (without vocals: M = -0.005, SD = 0.43; with vocals: M = 0.073, SD = 0.41). This result contrasts with the result when the low volume group was considered separately as the simple main effect of vocals in the medium difficulty tasks was not significant.

In the hard tasks, with both volume groups analysed together, the simple main effect of vocals was significant, $F(1, 109) = 5.398$, $p = 0.022$, $\eta_\rho^2 = 0.046$. The transformed error rate without vocals was higher than with vocals (without vocals: M = 0.095; with vocals: M = 0.002).

Given the differences in the interaction between text and vocals between the two volume groups, I would expect that the three-way interaction between text, vocals and volume from the combined analysis would be significant. However, this was not the case, $F(1, 111) = 0.162$, *n.s.*

Volume itself was a significant between groups factor, $F(1, 111) = 8.511$, $p = 0.004$, η_ρ^2 = 0.071. The transformed error rate was higher when the transcription typing tasks were accompanied by high volume music, and lower when the music was played at the lower volume level (low volume: M = -0.132, SD = 0.44; high volume: M = 1.11, SD = 0.49).

There were no other significant effects or interactions identified in the combined analysis of the transformed error rate data.

7.2.4.3 Typing Experience

The analyses of the typing experience data, in terms of perceived task difficulty and music distraction ratings, are compared in this section across the two volume groups. There is no non-parametric test that can accommodate a between participants factor, so no additional statistical tests were performed and there is no comparison with a combined analysis.

Task Difficulty

The analysis of task difficulty ratings was not affected by the speed classification. Table 7.28 compares the results from the analyses of task difficulty ratings between the two volume groups.

Test	**p value**	
	Low Volume	High Volume
Vocals	n.s.	n.s.
Text	**< 0.0005**	**< 0.0005**
Text x Vocals	n.s.	n.s.

Table 7.28: Comparison of perceived task difficulty inferential analysis

The omnibus effect of text was highly significant for both the low and high volume group, with the hard tasks perceived as more difficult than the medium difficulty tasks at both volume levels.

The omnibus effect of vocals was not significant for either volume group, and neither was the interaction between text and vocals.

Music Distraction

The music distraction ratings were analysed both with and without the speed classification applied. Table 7.29 presents a comparison of the analysis of music distraction ratings without the speed classification applied.

The omnibus effect of text was highly significant in the low volume group, but non-significant in the high volume group. The low volume group found the music to be more distracting in the hard tasks than the medium difficulty tasks, but there is no evidence from the high volume group that distraction ratings were affected by the task difficulty.

Test	p value	
	Low Volume	High Volume
Vocals	n.s.	n.s.
Text	**< 0.0005**	n.s.
Text x Vocals	n.s.	n.s.

Table 7.29: Comparison of music distraction ratings inferential analysis

The omnibus effect of vocals was not significant for either volume group, and neither was the interaction between text and vocals.

The analysis of music distraction ratings was affected by the speed classification. Table 7.30 shows a comparison of the analysis of distraction ratings with speed classification applied.

Test	p value	
	Low Volume	High Volume
Vocals x Speed*	**0.021**	0.063**
Text x Speed*	n.s.	n.s.
Text x Vocals x Speed*	n.s.	n.s.
Speed*	n.s.	n.s.

Table 7.30: Comparison of music distraction ratings inferential analysis

The interaction between vocals and speed group was significant in the low volume group and approached significance in the high volume group. These results were followed up with Wilcoxon signed rank tests looking at how vocals affected music distraction ratings between the two speed groups.

In the low volume group, the effect of vocals was significant for the slow typists who perceived the music with vocals as less distracting than the music without vocals. In the high volume group, the effect of vocals was significant for the fast typists who perceived the music without vocals as less distracting than the music with vocals.

There were no other significant effects or interactions for either volume group.

7.3 Discussion

The main aim of this experiment was to further explore how vocals affected typing performance and experience with a larger sample of data, particularly to see if the slow typists would also be affected, or if the effects of music are still constrained to just the fast typists. In some regards, this experiment has been successful as, in the high volume condition it has provided further evidence than vocals reduce the speed of the fast typists but there was no evidence of speed reductions for the slow typists. However, in terms of typing accuracy, there was no corresponding effect due to vocals for the high volume group.

The important, and particularly interesting outcome from this experiment is that when exposed to low volume music the participants' typing speed was higher when the music contained vocals, and lower when it was instrumental music. Further, in the hard tasks error rate was also reduced when accompanied by music that contained vocals. These results contrast strongly with those from the analysis of the effects due to the high volume music, and my initial hypotheses that vocals would reduce typing performance.

Experiment 3 showed that the effect of vocals differed between the two volume groups, with a reduction in performance with loud volume music due to vocals, but no effects due to vocals were identified when participants were exposed to the low volume music. By repeating the experiment with a larger sample, I was anticipating that either the same interaction would emerge with a significant reduction in performance due to the vocals in the high volume music but not the low volume music. Or, that I would identify significant reductions due to vocals at both volume levels, but with a larger effect size identified in the high volume group. One outcome I had not predicted was that vocals would improve performance in the low volume music.

The unexpected result from the low volume group needs to be explored and considered carefully. It would be one thing if there was no discernible difference between the with and without vocals conditions at the low volume, but to actually improve performance starkly contradicts my initial hypothesis. After checking this result was, indeed, correct many times and that there had been no mistakes in how the experiment was delivered or any problems with the data, I had to accept this unforeseen result was valid. So, the next step was to reflect on why the with vocals music version improved typing speed and reduced errors in the hard tasks when this music was played at a lower volume.

After some deliberation, I found myself considering three different scenarios. The first scenario I considered was that the without vocals version of the quiet music reduced performance but the with vocals version had no effect. The second scenario involves the with vocals version of the music improving performance, but the instrumental version having no effect. Or finally, a combination of both ideas, where the without vocals version reduces performance while the with vocals version improves performance.

In considering these options, the two that involve the without vocals version of the music actually reducing performance seemed somewhat unlikely because, when played at a quiet volume, this music is fairly unobtrusive, and I suspect it could easily be blocked out or ignored by the participants as it is not particularly attention grabbing. The without vocals alt rock music does not have a particularly strong melody, is very repetitive in terms of the rhythm, and when played at a low volume could probably easily be ignored.

Further developing my theory that the without vocals version of the music has no effect on performance at a quiet volume, the only scenario which still aligns involves the with vocals version of the music improving performance. This explanation seemed unlikely and contradicted my initial hypothesis and the evidence from my previous experiments that vocals reduced performance. However, after more reflection, I stopped thinking about the vocals as a separate dimension of music and began to consider the effect of the music as a whole, by asking what does the with vocals music have that the without vocals version does not? The answer, other than the obvious vocals, is that the with vocals version contains features which may bring the music closer to the forefront of the typist's attention. So, if the rest of the music was having a positive influence on performance, perhaps the positive effects dwarfed any negative effects due to the verbal content of the music, cumulatively resulting in an overall improvement to performance.

Essentially, my argument here is that the instrumental version of the music contains very few features which would grab the typists' attention when the music was played at a low volume. When the participant is focused on the task as there is no strong melody, and no one is singing, the music can be more easily ignored, and therefore it has no effect on performance. People are attuned to vocals and catchy melodies, so at a low volume, when the music contains vocals and has a strong melody, it is possible that the music is brought into the attention of the participants somewhat, meaning that any effect caused by other dimensions of the music dominate (e.g. maybe the tempo, or rhythm) and, in this case, the cumulative effect improved performance.

This line of thinking also took me back to consider the original mixes of the music I created to identify exactly how the without and with vocals alt rock pieces differ in terms of dimensions beyond vocals and melody. I asked myself, are there any other dimensions of music where these two versions differ? One other potentially important difference in dimensions soon became apparent – although I had been controlling playback volume through the website used to deliver the experiment, the average loudness of the with and without vocals versions of the music were not the same between the tracks. The with vocals track is louder than the without vocals version because the with vocals version contains additional audio tracks, which adds audio energy to the mix that is not included in the without vocals version of the music. So, not only does the without vocals version of the music contain no melody, or vocals, it also has a lower average loudness (i.e. it is quieter) than the with vocals version.

Together, these differences between the with and without vocals versions of the music may explain why it seems that, at a low volume the without vocals music may be ignored and not attended to at all, whereas the with vocals version has an overall positive effect on performance. Of course, when played at a high volume both pieces of music are noticeable so are in the attention of the typists, so neither version of the music can be ignored. The result for the high volume condition is the expected reduction in performance due to the presence of vocals in the music.

7.3.1 Discussion of Each Experiment Aim

The initial aims of this experiment were:

1. to gather further evidence that the presence of vocals in a piece of music has a negative effect on typing performance and experience,
2. to see if task difficulty affects the results,
3. to see if any identified effects of music are the same at different volume levels,
4. to gather further evidence that the effect of music on typing performance and experience differs between slow and fast typists.

The results presented in the previous section are discussed in more detail in relation to the specific experiment aims in the following sections.

7.3.1.1 Experiment Aim 1

The first aim of this experiment was to gather, hopefully conclusive, evidence that vocals reduce performance when typing. The previous discussion section has shown this aim to be partially achieved, as when accompanied by high volume music the fast typists' speed was reduced when the music contained vocals but there was no evidence their typing accuracy was similarly affected. However, when exposed to the music played at a low volume the presence of vocals improved performance with higher speeds across both speed groups, and lower error rates achieved in the hard tasks. A possible explanation of these results has already been discussed, but it may relate to how the music affects performance as a whole entity, so rather than by looking at the individual effect of dimensions it may be important to consider the cumulative effect due to the sum of its parts, so to speak.

7.3.1.2 Experiment Aim 2

The second aim of this experiment focused on task difficulty and identifying whether the effect of music differs between the medium and hard tasks, when both tasks were notionally English like. This aim has been successful as, in the hard tasks, the error rate was reduced with low volume with vocals music when compared to high volume music. This result means that when the participants were concentrating hard on the task, their performance was better when accompanied by music with vocals. The result shows that, even when working on something difficult, music can infiltrate to the extent of affecting performance. Again, it is not entirely clear if the with vocals music improved performance, or if it was the without vocals music reducing performance, or both. But, my hypothesis that higher concentration on a task would lead to smaller effects due to music has been contradicted in an interesting manner, suggesting the difference in effect of music on the participants when classified as slow or fast typists ought to be considered from a different position.

7.3.1.3 Experiment Aim 3

The third aim of this experiment was to explore whether the effect of vocals differed between the different volume levels. In this case, the experiment has been successful as there is strong evidence that the effect of vocals differs considerably when the music was played at the two different volume levels.

7.3.1.4 Experiment Aim 4

The final aim of this experiment focused on exploring how the effect of music differs between the slow and fast typists. Again, this experiment has been somewhat successful with regard to this aim as, particularly in the high volume group, the effect of music on performance differed between the slow and fast typists. The difference in scale of effect on the performance measures between the slow and fast typists seems to be constrained to the typing speed measure, with the accuracy analysis not influenced or improved with the speed classification.

7.3.2 Limitations of This Experiment

The sample size in this experiment was good, with 116 participants across the two volume groups. However, again, some of the post hoc analyses did not reach significance which is disappointing, especially as the interactions tended to have large effect sizes. Further, similarly to Experiment 3, because the data was collected in a classroom based methodology the participants were constrained to those students taking the module. Although a large sample of data was gathered, with interesting results, as all the participants were first year undergraduate Computer Science students the results should not be over stated in terms of generalising the identified effects.

Also, as the experiment was performed using a classroom based methodology a without music condition could not be accommodated meaning statements such as “low volume music with vocals improves performance when typing” would be premature as it is only the direct comparison to the without vocals music where the improvement was seen. Further, given my new theory that the unexpected result occurred due to effects in the rest of the music dominating and overall improving performance, more work is needed to before any generalisations can authoritatively be made as these results may simply be constrained to this particular piece of music.

7.3.3 Moving Forwards

The results from this experiment are interesting as they contradict my initial hypothesis but confirm that volume is a very important factor in research of this nature. As previously discussed, it would be extremely difficult to conduct a large scale online study while controlling the volume of music so the next stage in taking the work forwards investigating the effect of vocals on transcription typing performance would be to research and establish a suitable online methodology for these sorts of studies where volume cannot be controlled. Alternatively, experiments should be performed using a laboratory or classroom based methodology despite the limitations to sample size and demographics this leads to because volume can be controlled in these environments.

Chapter 8

TIME SIGNATURE AND TEMPO

Experiment 5

Experiment 4 provided clear evidence that the volume of music has an important and large effect on transcription typing performance and experience. When played at a low volume, the music with vocals improved typing speed, a result which contradicted my initial hypothesis that vocals in music would reduce performance. The low volume music with vocals may have improved rather than reduced typing speed for a couple of reasons. First, the inclusion of vocal tracks in the mix may bring the music to the attention of the participants because the vocals are at the audio range humans are most attuned to. Second, although the volume of playback of the music was kept the same in the with and without vocals conditions, the average loudness values differ between the tracks due to the mixing process. The with vocals version of the music track contains more audio energy than the without vocals versions, so although the volume was the same, the loudness was not. This unexpected result suggests that the loudness of the music is an important factor that needs to be carefully controlled. So, an online experiment cannot, at this stage be considered an appropriate approach to perform a valid experiment in this area due to the experimenter's inability to control the loudness of the music when using a remote methodology.

As an online experiment is not feasible at this stage, the objectives of the final experiment in this thesis take a slight change in direction. Instead of further pursuing the volume, vocals and task difficulty line of enquiry, this final experiment follows up on one of the interesting results from Experiment 2 – namely, that the classical and rock music led to similar performance and experience effects despite the two pieces having quite different musical characteristics.

Two important differences between the classical and rock music pieces were the time signature and tempo. The rock music was in a 6/8 time signature at 210 beats per minute (bpm), while the classical music was in 4/4 time at 140 bpm. Interestingly, if you consider that in music with a 6/8 time signature, the first and fourth beats in the bar are typically emphasized, and in a 4/4 time signature, the first beat of the bar is emphasized strongly, at these tempos in 1 minute of playback both pieces of music would result in 35 emphasized beats. Or, in 7 seconds of playback, both would include five emphasized beats, which is shown in Figure 8.1.

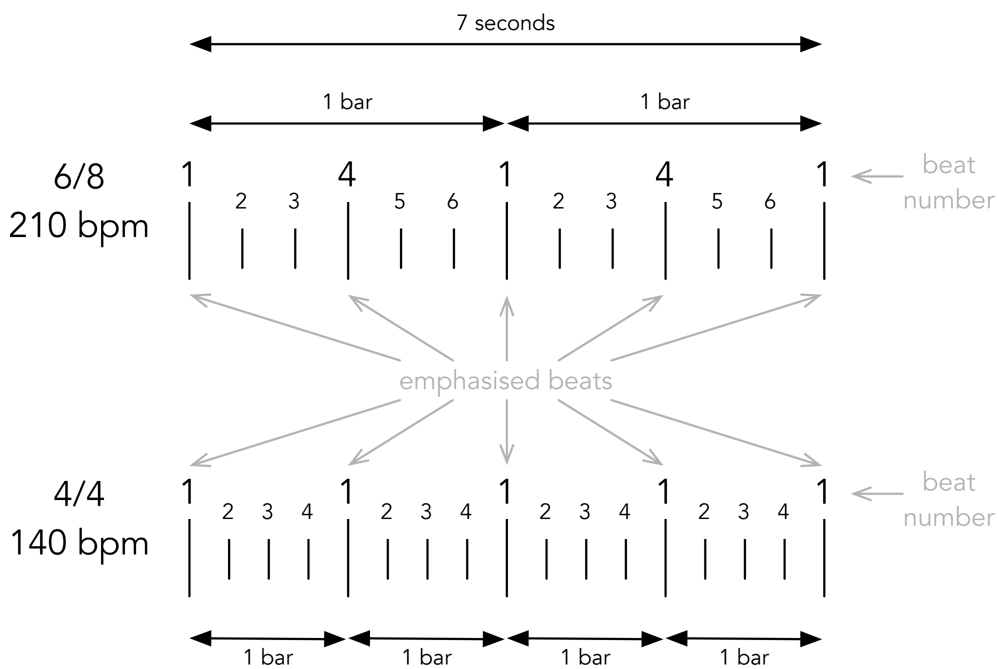


Figure 8.1: Visual representation of matching emphasized beats

So, even though the tempos of the two pieces of music were different, the number and frequency of the emphasized beats is the same. This could, in part, be one justification for why no differences in typing performance where identified between the classical and rock music conditions, despite these pieces of music having vastly different tempos, time signatures and other musicological characteristics. To explore this hypothesis, an experiment investigating the effect of tempo and time signature on transcription typing performance and experience was performed. As an online methodology would be inappropriate without accommodations being made to support the lack of control of the important loudness parameter, this experiment returned to using a laboratory based method.

8.1 METHOD

8.1.1 AIMS

The aims of this experiment were:

1. to investigate whether tempo affects transcription typing performance and experience,
2. to identify if music in 3/4 time signature affects people differently to music in 4/4,
3. to explore whether the frequency of the emphasized beat affects performance and experience,
4. to gather further evidence that the effect of accompanying music on typing performance and experience differs between slow and fast typists.

8.1.2 PARTICIPANTS

The participants in this experiment were recruited by advertising via email lists and university forums. The 36 participants (27 male, 9 female) were all involved in academia in some respect. Sixteen of the participants were undergraduates studying Computer Science, Electronics or Maths, 14 were postgraduates studying Computer Science or Electronics, and 2 were postgraduate English students. There were four non-student participants, two were professional researchers from Computer Science and two university administrators. The ages of the participants separated into groups is shown in Table 8.1.

Age				
< 18	**18 – 24**	**25 – 34**	**35 – 44**	**> 45**
0	21	9	3	3

Table 8.1: Participant ages

All of the participants were native English speakers, and none were dyslexic nor had known hearing impairments. Each participant received a £10 Amazon or Marks and Spencer's gift voucher.

8.1.3 Design

The experiment used a fully factorial repeated measures design. Each participant completed seven transcription typing tasks:

- without music,
- with music in 3/4 at 140 bpm,
- with music in 3/4 at 180 bpm,
- with music in 3/4 at 210 bpm,
- with music in 4/4 at 140 bpm,
- with music in 4/4 at 180 bpm,
- with music in 4/4 at 210 bpm

In this chapter, the 140, 180 and 210 bpm tempos are referred to as slow, mid and fast tempo music respectively.

As this experiment used a laboratory methodology a without music condition could be accommodated so there were two experiment design paradigms. The first paradigm focused on the impact of the pieces of music as a whole with seven levels of a music independent variable (IV). Figure 8.2 shows the arrangement of IVs for experiment design paradigm 1.

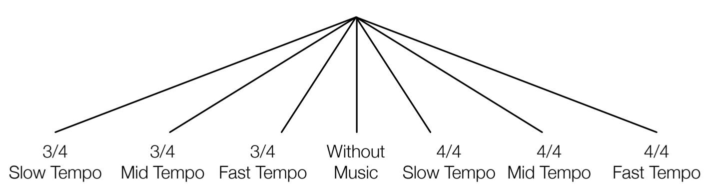


Figure 8.2: Structure of IVs for experiment paradigm 1

The second experiment design paradigm specified two repeated measures IVs, tempo (three levels: slow, mid and fast) and time signature (two levels: 3/4 and 4/4). Figure 8.3 shows the arrangement of IVs for experiment design paradigm 2.

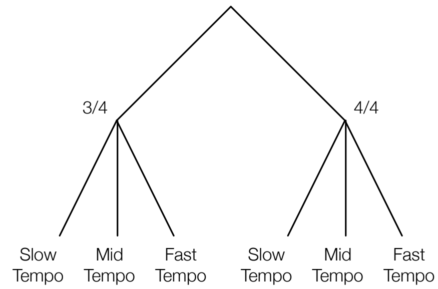


Figure 8.3: Structure of IVs for experiment paradigm 2

To avoid practice and fatigue effects, the order of presentation of tasks was counterbalanced.

In this experiment, I was also interested in the frequency of the emphasized beats in the music, as an indication of the interaction between the music tempo and time signature. The number of emphasized beats in 1 minute of music at each tempo and for each time signature is shown in Figure 8.4.

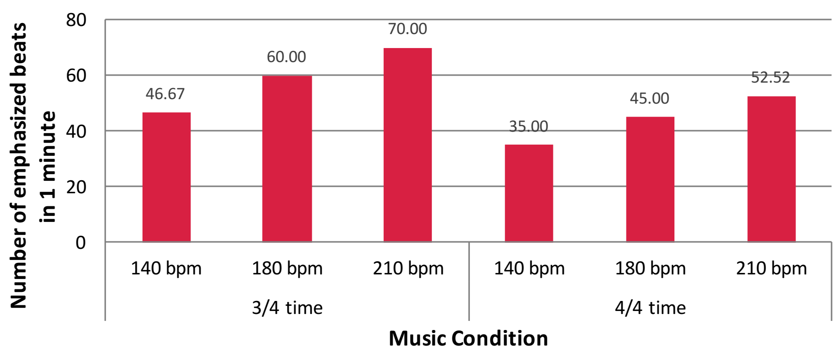


Figure 8.4: Number of emphasized beats in 1 minute of music

The 3/4 time 210 bpm music has the highest number of emphasized beats in 1 minute at 70.00 emphasized beats per minute, with the lowest frequency in the 4/4 time 140 bpm music at just 35.00 emphasized beats per minute. Interestingly, the 3/4 time 140 bpm music has a similar number of emphasized beats in 1 minute to the 180 bpm 4/4 time music at 46.67 and 45.00 respectively. Experiment design paradigm 1 used a 1 by 7 design which encompasses this 1 by 6 design, so it is not necessary to include a third paradigm to accommodate this alternative approach to conceptualising the IV as any significant effects will be revealed in the 1 by 7 analysis.

The dependent variables (DVs) were typing performance measured in both speed (CPM) and accuracy (error rate percentage) and typing experience measured using 7-point Likert items of task difficulty and music distraction. Chapter 3, Section 3.3.1.2 provides more explanation of how these DVs were measured and calculated.

8.1.4 EQUIPMENT AND MATERIALS

The experiment was performed using the laboratory methodology in a quiet usability lab in the Department of Computer Science at the University of York.

The participants used a desktop PC running Windows 7 with Firefox version 28. The music was controlled using Audacity (Open Source, 2012) running on a MacBook Pro laptop, which was connected to a pair of Philips SPA 2210 2.0 laptop speakers. See Section 3.3.2, Chapter 3 for more details of the interface and setup of the experiment.

The *Menuetto* and *Allegro* movements from Mozart's Symphony No. 25 in G Minor were used as the music stimuli. These had been sped up or slowed down to approximately 140, 180 and 210 bpm using Audacity. The advanced English text was used as presented text in this experiment (see Chapter 3, Section 3.3.1.1 for more information regarding the presented text and music IVs).

Short paper based questionnaires were used to collect subjective ratings of perceived task difficulty and music distraction. The participants also completed a demographics questionnaire. Example questionnaires are included in Appendix A.2, A.3 and A.4.

8.1.5 PROCEDURE

The procedure for this experiment exactly followed that described in Chapter 3, Section 3.3.3.1. All participants completed seven 4.5 minute transcription typing tasks, so were typing for a total of 31.5 minutes during the experiment.

8.2 RESULTS

The first stage in the analysis was to identify if a classification into groups based on speed would be appropriate. As in the previous experiments, histograms were created to explore whether the distribution of speed is bimodal with scatterplots created afterwards. Then, inferential analysis of the performance and experience measures was performed.

8.2.1 Speed Classification

A possible speed classification was explored by plotting a histogram of the CPM data combined, Figure 8.5.

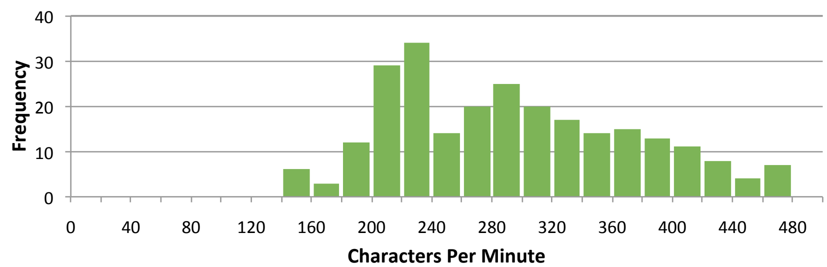


Figure 8.5: Histogram of CPM for all conditions

The bimodal characteristic is less evident in this distribution than in the previous experiments reported in this thesis, but there is a drop at 240 CPM, which is a similar characteristic to that identified in Experiment 3. Again, a possible trimodal characteristic may be shown in this data distribution. To explore this further, distributions of the without music condition, slow, mid and fast tempo conditions were created, looking for evidence of the bimodal or trimodal distributions as well as scatterplots based on numerous thresholds. These distributions are included in Appendix H.1. It was not possible to determine a consistent approach to the speed classification using either bimodal or trimodal thresholds, as many participants would have been placed inconsistently. So, speed classification was not incorporated in the analysis of data collected in this experiment.

8.2.2 Typing Performance

Typing performance was measured using CPM and error rate. The objective data was assessed for outliers using box plots and normality using Shapiro-Wilk's tests and by looking at the distribution's skewness and kurtosis, following the procedure described in in Chapter 3, Sections 3.5.1.2 and 3.5.1.3. The results from the normality tests and box plots can be found in Appendix H.1 and H.2.

8.2.2.1 Characters Per Minute

None of the CPM distributions deviated significantly from normality and there were no outliers. Table 8.2 shows descriptive statistics for CPM.

Time Signature	Tempo	All Typists (N = 36) M (CPM)	SD (CPM)
Without Music		298.80	79.80
3/4	Slow	296.52	79.98
	Mid	293.41	80.20
	Fast	294.44	80.49
4/4	Slow	294.24	77.97
	Mid	293.93	80.55
	Fast	292.19	78.15

Table 8.2: Descriptive statistics of CPM

The CPM data was first analysed using fully factorial 1 by 7 repeated measures ANOVA with music (with seven levels) as the within participants factor. The CPM data was also analysed using a 2 by 3 repeated measures ANOVA with time signature (two levels: 3/4 and 4/4) and tempo (three levels: slow, mid and fast). The assumption of sphericity was violated for music in the 1 by 7 ANOVA, $W(20) = 0.285$, $p = 0.004$, and for tempo, $W(2) = 0.684$, $p = 0.002$, in the 2 by 3 ANOVA, so Greenhouse-Geisser corrections were applied.

Table 8.3 shows the results from the repeated measures ANOVAs. None of the effects or interactions were significant in either experiment design paradigm.

Design	Test	F	df	p	η_p^2
1 x 7	Music	0.925	4.146, 145.119	n.s.	-
2 x 3	Time Sig	0.728	1, 35	n.s.	-
	Tempo	0.349	1.520, 53.207	n.s.	-
	Time Sig x Tempo	0.284	2, 70	n.s.	-

Table 8.3: Inferential analysis of CPM

Error Rate

The error rate data was transformed using logarithms following the procedure described in Chapter 3, Section 3.5.2. Table 8.4 shows descriptive statistics of transformed error rate for each condition.

Time Signature	Tempo	All Typists (N = 36) M (%)	All Typists (N = 36) SD (%)
Without Music		-0.233	0.40
3/4	Slow	-0.241	0.41
	Mid	-0.276	0.46
	Fast	-0.108	0.40
4/4	Slow	-0.286	0.43
	Mid	-0.207	0.39
	Fast	-0.277	0.46

Table 8.4: Descriptive statistics of transformed error rate percentage

The data was analysed using a 1 by 7 repeated measures ANOVA (with seven levels of music) and a 2 by 3 repeated measures ANOVA (with two levels of time signature and three levels of tempo). Table 8.5 shows the results from both analyses.

Design	Test	F	df	p	η_p^2	Figure
1 x 7	Music	1.976	6, 210	0.070**	0.053	8.6
1 x 6	Emphasized Beats	2.278	5, 175	**0.049**	0.061	8.6
2 x 3	Time Sig	1.306	1, 35	n.s.	-	-
	Tempo	1.207	2, 70	n.s.	-	-
	Time Sig x Tempo	4.618	2, 70	**0.013**	0.117	8.7

Table 8.5: Inferential analysis of transformed error rate percentage
(** non-significant result where $p < 0.1$)

In the 1 by 7 analysis, there was a trend towards a significant effect of music, $F(6, 210) = 1.976$, $p = 0.070$, $\eta_p^2 = 0.053$, but in the 1 by 6 analysis the effect of the frequency of the emphasized beat was significant, $F(5, 175) = 2.278$, $p = 0.049$, $\eta_p^2 = 0.061$. Both these results are captured in Figure 8.6.

Post hoc pairwise comparisons were performed. The full set of pairwise comparisons are included in Appendix H.3.1. In the 1 by 7 design, including the without music condition in the pairwise comparison the Bonferroni adjustment sets the required level to achieve significance at $p < 0.0024$, while using the 1 by 6 design the required level for significance is $p < 0.0033$.

Table 8.6 shows a subset of the pairwise comparisons which were significant at a non-adjusted level.

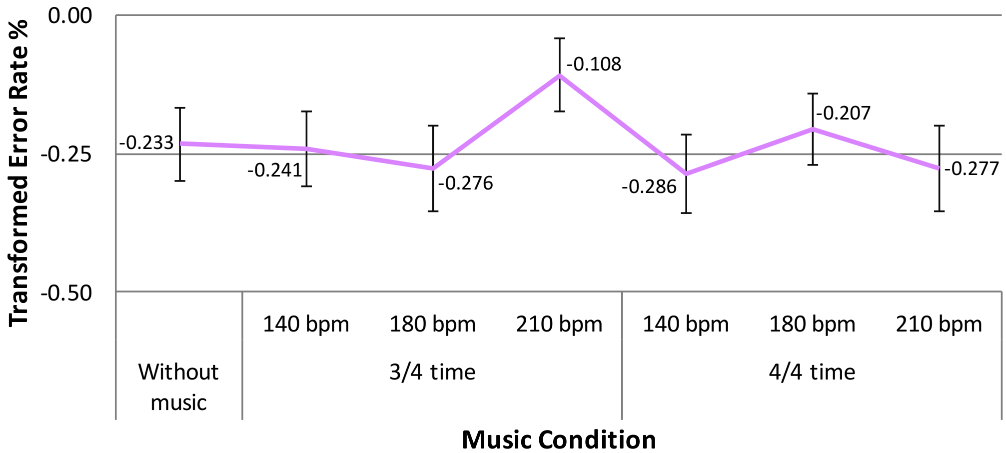


Figure 8.6: Trend towards significant effect of music (*p* = 0.070) **and significant effect of frequency of emphasized beat** (*p* = 0.049)

Condition 1	Condition 2	Mean Difference	p
3/4 Fast	Without music	-0.125	0.037**
3/4 Fast	3/4 Slow	-0.133	0.025**
3/4 Fast	3/4 Mid	-0.168	0.030**
3/4 Fast	4/4 Slow	-0.178	0.005**
3/4 Fast	4/4 Fast	-0.169	0.025**

Table 8.6: Post hoc tests for omnibus effect of music
(significance at $p < 0.0024$ in 1 by 7 design, significance at $p < 0.0033$ in 1 by 6 design, ** significant at the non-adjusted level)

The most striking observation when looking at the comparisons in Table 8.6 is that they all involve the 3/4 time signature, fast tempo condition. That is, the only pairwise comparisons which indicate reduced performance in terms of accuracy involved the 3/4 time signature, fast tempo music which had the highest frequency of emphasized beats, at 70.00 emphasized beats in 1 minute. In fact, the only comparison which did not reach significance at the non-adjusted level involving the 3/4 fast tempo piece, was the comparison to the 4/4 mid tempo music.

Returning to the results presented in Table 8.5, the two-way interaction between time signature and tempo was significant, $F(2, 70) = 4.618$, $p = 0.013$, $\eta_p^2 = 0.117$, (Figure 8.7).

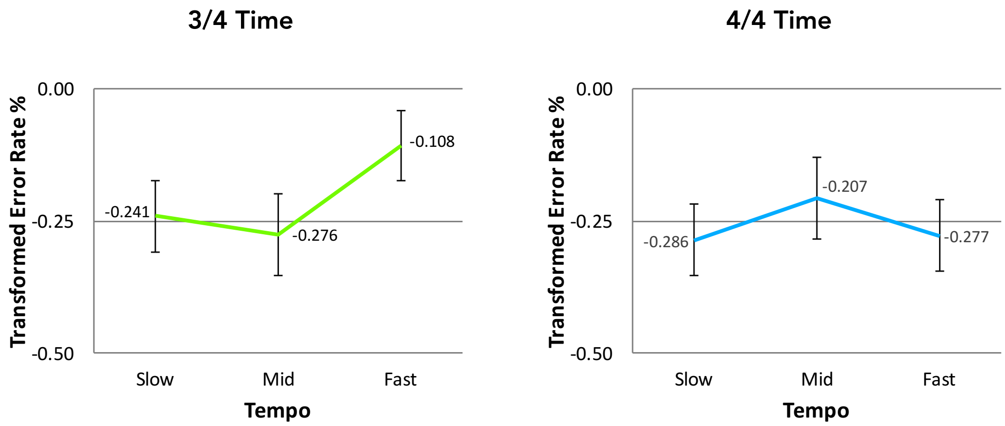


Figure 8.7: Significant interaction between time signature and tempo (p = 0.013)

The simple main effect of tempo was significant with music in ¾ time, $F(2, 70) = 3.662$, $p = 0.031$, $\eta_\rho^2 = 0.095$, but not with music in 4/4 time, $F(2, 70) = 1.206$, *n.s.*

Post hoc pairwise comparisons were performed between the pieces of music with a 3/4 time signature music, (Table 8.7).

3/4 Time Signature			
Condition 1	**Condition 2**	**Mean Difference**	**p**
Slow	Mid	0.036	n.s.
Slow	Fast	-0.133	0.025**
Mid	Fast	-0.168	0.030**

Table 8.7: Post hoc tests for omnibus effect of tempo in 3/4
(significance at p < 0.0166, ** significant at the non-adjusted level)

All pairwise comparisons were non-significant at the Bonferroni adjusted level. But, at the non-adjusted level, the difference in transformed error rate between the slow and fast tempo conditions, and between the mid and fast tempo conditions was significant. The transformed error rate was higher with fast tempo music, than with either slow or mid tempo music (fast: $M = -0.108$, $SD = 0.40$; mid: $M = -0.276$, $SD = 0.46$; slow: $M = -0.241$, $SD = 0.41$).

The simple main effect of time signature was significant for fast tempo music, $F(1, 35) = 5.450$, $p = 0.025$, $\eta_\rho^2 = 0.135$. The transformed error rate was higher with music that was played at a fast tempo using a 3/4 time signature and lower at the fast tempo in 4/4 (3/4: $M = -0.108$, $SD = 0.40$; 4/4: $M = -0.277$, $SD = 0.46$).

The simple main effect of time signature was not significant for either mid, $F(1, 35) = 1.611$, *n.s.*, or slow, $F(1, 35) = 0.628$, *n.s.*, tempo music.

8.2.2.2 Typing Experience

Typing experience was measured quantitatively in terms of perceived task difficulty and music distraction ratings and qualitatively by asking participants to explain why that piece of music was distracting. Following the procedure described in Chapter 3, Section 3.5.3, a parametric analysis was first performed using both experiment design paradigms and any significant effects or interactions followed up by the most suitable non-parametric test.

Task Difficulty

Participants rated how difficult they found each task using 7-point Likert items ranging from 1 = extremely difficulty to 7 = extremely easy. The mapping of rating to description was reversed for analysis so that the easier tasks had a lower numeric value. Table 8.8 shows descriptive statistics of lower and upper quartiles (*Q1* and *Q3* respectively), median (*Mdn*) and mean (*M*).

Time Signature	Tempo	All Typists (N = 36)			
		Q1	*Mdn*	*Q3*	*M*
Without Music		2.00	3.00	4.00	3.11
3/4	Slow	2.00	3.00	4.00	3.22
	Mid	2.00	3.00	4.00	3.11
	Fast	2.00	3.50	4.00	3.33
4/4	Slow	2.00	3.00	4.00	3.06
	Mid	3.00	3.00	4.00	3.33
	Fast	2.00	3.00	4.00	3.28

Table 8.8: Descriptive statistics of task difficulty ratings for all typists

The results from this parametric analysis are included in Table 8.9.

Design	Test	F	df	p
1 x 7	Music	0.759	6, 120	n.s.
2 x 3	Time Sig	0.000	1, 35	n.s.
	Tempo	0.802	2, 70	n.s.
	Time Sig x Tempo	1.409	2, 70	n.s.

Table 8.9: Parametric inferential analysis of task difficulty ratings

None of the effects or interactions were significant in either experiment paradigm.

Music Distraction

Ratings were taken for how distracting the music was perceived as using 7-point Likert items ranging from 1 = extremely distracting to 7 = not at all distracting. The mapping of values was revered for analysis so that lower numbers indicated less distracting music. Table 8.10 presents descriptive statistics of the music distraction ratings.

Time Signature	Tempo	All Typists (N = 36)			
		Q1	*Mdn*	*Q3*	*M*
3/4	**Slow**	2.00	3.00	4.00	3.11
	Mid	2.00	3.00	4.00	2.97
	Fast	2.00	3.00	5.00	3.39
4/4	**Slow**	2.00	2.50	4.00	2.92
	Mid	2.00	3.50	4.00	3.11
	Fast	2.00	4.00	5.00	3.58

Table 8.10: Descriptive statistics of music distraction ratings for all typists

Again, the approach of performing a parametric analysis using both experiment design paradigms and following up significant effects with appropriate non-parametric tests was taken. The results from the parametric analysis are included in Table 8.11.

Design	Test	F	df	p
1 x 6	Music	2.502	5, 175	**0.032**
2 x 3	Time Sig	0.171	1, 35	n.s.
	Tempo	3.743	2, 70	**0.029**
	Time Sig x Tempo	1.184	2, 70	n.s.

Table 8.11: Parametric inferential analysis of music distraction ratings

In the 1 by 6 analysis, the omnibus effect of music was significant, $F(5, 175) = 2.502$, $p = 0.032$, so this was followed with a 1 by 6 Friedman test, which was also significant, $\chi^2(5) = 0.049$.

Post hoc pairwise comparisons using Wilcoxon signed rank tests were performed. With 15 comparisons the p value required for significance is 0.0033 using the Bonferonni adjustment. The full set of results is included in Appendix H.3.2, but pairwise comparisons where p is less than 0.50 are included in Table 8.12.

Condition 1	Condition 2	z	p	r
4/4 Fast	3/4 Slow	-1.958	0.049**	0.23
4/4 Fast	3/4 Mid	-2.306	0.021**	0.27
4/4 Fast	4/4 Slow	-2.450	0.014**	0.29
4/4 Fast	4/4 Mid	-1.966	0.049**	0.23

Table 8.12: Post hoc tests for omnibus effect of music
(significance at $p < 0.0033$, ** significant at the non-adjusted level)

These pairwise comparisons show that the 4/4 fast tempo music was perceived as more distracting than almost all of the other pieces, the one exception being the 3/4 fast tempo piece. Comparisons of the medians show that the 4/4 fast tempo music was perceived as more distracting than the 3/4 and 4/4 slow and mid tempo pieces (4/4 fast: *Mdn* = 4.00; 3/4 slow: *Mdn* = 3.00; 3/4 mid: *Mdn* = 3.00; 4/4 slow: *Mdn* = 2.50; 4/4 mid: *Mdn* = 3.50).

The 2 by 3 parametric analysis also revealed a significant effect of tempo, $F(2, 70) = 3.743$, $p = 0.029$. This was followed up by collapsing the data into the three tempo groups and performing a Friedman test, which was significant, $\chi^2(2) = 11.118$, $p = 0.004$.

Post hoc pairwise comparisons were performed, (Table 8.13).

Condition 1	Condition 2	z	p	r
Slow	Mid	-0.165	n.s.	-
Slow	Fast	-2.471	**0.013**	0.021
Mid	Fast	-2.423	**0.015**	0.021

Table 8.13: Post hoc tests for main effect of tempo
(significance at $p < 0.0166$)

The fast tempo music was perceived as significantly more distracting than both the slow and mid tempo music (slow: *Mdn* = 3.00; mid: *Mdn* = 3.00; fast: *Mdn* = 4.00).

As well as the quantitative music distraction ratings 19 of the 36 participants provided comments to explain why they found a particular piece of music distracting. A light weight approach to content analysis was performed, where each comment (or part of a comment) was assigned to a category that emerged through inspection of the comments. Table 8.14 includes the category name, frequency of recurrence of the comment and some examples.

Category		Frequency	Examples
Dynamics	Too loud	19	"Occasionally it became very loud" "There were bits that got loud and demanded attention"
	Changes to dynamics	11	"When it changed volume it was more noticeable, even when decreasing in volume" "Changes in volume usually rapid, very off putting"
	Not specified	1	"The volume was a little distracting"
Tempo	Too fast	5	"It had been sped up and I found the quickened rhythms very distracting, felt in a way I had to keep up with those" "It seemed too fast, out of rhythm with my typing"
	Changes to tempo	5	"Lots of changes in speed in the music" "The tempo ... changes seemed frequent and would draw my attention away from the typing"
	Too slow	1	"It sounded ... slow"
Rhythm		7	"The staccato rhythm of the music was jarring" "It was very staccato which jarred a bit"
Style		5	"It was ... very lively music" "It was ... quite dramatic music" "The music made me feel tense and I kept losing my place in the text"
Heard in previous task		4	"It sounded like the previous song, so I was consciously checking the differences between the two" "Only briefly distracted as I recognised the music from the previous exercise"
Too repetitive		2	"Very annoyingly repetitive, felt a little nightmarish going round in circles"
Non-specified irregularity		1	"In places the music was ... irregular and this was distracting"

Table 8.14: Qualitative analysis of music distraction justifications

Out of the 61 comments categorised from this study, 31 related to the music's dynamics. Only one participant used the musicological term dynamics, with the others referring to the volume of the music. Inspection of the comments and context suggest that although those in the *too loud* category may be referring to playback volume comments like "occasionally it became very loud" suggest that the participants are really referring to the dynamics in the music rather than the playback volume. There were 19 comments stating the music was too loud in some regard. A further 11 comments highlighted the change in dynamic as being distracting, i.e. it was not just that the music was too loud, it was distracting when the loudness of the music varied.

The next most frequently mentioned category involved comments relating to tempo. There were five comments that the music was too fast, many of which referred to the impact this had on their typing and feeling like they had to maintain speed with the music. The *changes to tempo* category is particularly interesting, as there were five comments by different participants referring to the tempo changing, but the tempo was fairly constant throughout both pieces of music in terms of beats per minute. The rhythm of the melody, for instance, does change with some passages containing more frequent quavers and semiquavers, but that is not tempo *per se*. Other participants commented that the rhythm was staccato, with multiple participants using the term jarring.

8.3 Discussion

The main aim of this experiment was to explore how tempo affects transcription typing performance, together with exploring whether the frequency of emphasized beats are important through incorporating a time signature manipulation. In terms of typing speed, the results do not provide any evidence that the tempo, time signature or frequency of emphasized beats has an effect on CPM. However, in this experiment I was unable to incorporate a speed classification because a clear grouping structure did not emerge through inspection of the CPM data. This classification had been crucial in the analysis of CPM in the previous experiments, so it is possible that, had a classification structure been able to be applied, the music may have affected typing speed. In fact, the qualitative comments indicate a number of participants were consciously trying to keep up with the music which was a distraction in some circumstances. But, was the tempo (or frequency of emphasized beats) affecting typing speed before it got to the point of being a distraction? From this experiment, there is no evidence to suggest so, but perhaps my inability to apply a successful classification was due to the effect of the music. This is the first experiment reported in this thesis to suggest that the approach of applying a classification based on achieved speeds across the experiment conditions may benefit from refinement.

In terms of typing accuracy, there were significant effects, which as other experiments have shown the classification of typing speed to be less important in the accuracy analysis is not altogether surprising. The results indicate there were different effects on typing accuracy caused by the music. There was a trend towards a significant effect of music when the without music condition was included in the 1 by 7 analysis, but in the 1 by 6 analysis focusing on the effect of the frequency of the emphasized beat the effect of music was significant. The music with the highest number of emphasized beats during the experiment was associated with the highest error rate, which was higher than all other conditions except the 4/4 mid tempo music. Although, the pairwise comparisons were significant at the non-adjusted rather than Bonferroni adjusted level, which is mainly due to the high number of pairwise comparisons that were performed.

In terms of typing accuracy, there was also a significant interaction between tempo and time signature. In 3/4 time, the simple effect of tempo was significant, with the fast tempo music leading to more errors than both the slow and mid tempo music. The simple main effect of time signature was also significant in the 4/4 tempo music, with the 3/4 fast tempo music leading to more errors than the 4/4 fast tempo music. Taken together, these results clearly show that tempo alone is not the most important aspect. Instead, the time signature of the music at particular tempos is an important consideration in terms of affecting the accuracy of transcription typing.

When looking at the experience measures, the task difficulty ratings were not affected by the music or any of the music IVs. But, the participants perceived the 4/4 fast tempo music as more distracting than all of the other conditions, except the 3/4 fast tempo condition. This is a particularly interesting result as, in terms of the frequency of the emphasized beats, the 4/4 fast tempo music is similar to the 3/4 mid tempo music. This result indicates, that perhaps it is something other than tempo, time signature or frequency of emphasized beats which is leading to the participants perceiving this particular music as more distracting. Perhaps there is some other musicological difference between the 3/4 and 4/4 music which is emphasized at the faster tempo and led to an increase in distraction causing capabilities of this music.

Finally, overall, in the music distraction ratings the effect of tempo was significant, with the slow and mid tempo music perceived as significantly less distracting than the fast tempo music. In this regard, it is interesting that the effect of time signature and the interaction between tempo and time signature were not significant because, if there are musicological differences between the two pieces which affect distraction ratings, I would expect to see differences due to the 4/4 time signature as well.

8.3.1 Discussion of Each Experiment Aim

The initial aims of this experiment were:

1. to investigate whether tempo affects transcription typing performance and experience,
2. to identify if music in 3/4 time signature affects people differently to music in 4/4,
3. to explore whether the frequency of the emphasized beat affects performance and experience,
4. to gather further evidence that the effect of accompanying music on typing performance and experience differs between slow and fast typists.

The results presented in the previous section are discussed in more detail in relation to the specific experiment aims in the following sections.

8.3.1.1 Experiment Aim 1

The first aim of the experiment was to explore whether tempo affected transcription typing performance or experience. To establish success with regard to this aim, the effect of tempo needs to be considered from the 2 by 3 analysis. The main effect of tempo was significant in the music distraction ratings, with the fast tempo music perceived as more distracting than the both the slow and mid tempo music. However, the overall effect of tempo on the other DVs was not significant.

8.3.1.2 Experiment Aim 2

The second aim of this experiment focused on whether time signature affected performance through comparing the effect of music in 3/4 time to music in 4/4 time. To explore success in regard to this aim the main effect of time signature needs to be considered across all of the DVs. In this regard, the experiment has not been successful as overall, the effect of time signature was not significant.

8.3.1.3 Experiment Aim 3

The third aim of the experiment was to explore whether the frequency of the emphasized beat affected performance and experience where there has been some success from this experiment. Although the post hoc pairwise comparisons were not significant at the Bonferroni adjusted level, the 3/4 time signature fast tempo music led to reduced performance when compared to almost all other music conditions. This particular combination of tempo and time signature had the highest number of emphasized beats in 1 minute, demonstrating that my approach of considering the frequency of emphasized beats is a better method than just focusing on tempo alone. This result makes me question much of the previous literature regarding the effect of tempo and wondering if similar experiments should be performed using the same tempo values, but with music with different time signatures.

8.3.1.4 Experiment Aim 4

The final aim was to gather further evidence that the effect of accompanying music on typing performance and experience differs between slow and fast typists. In this regard, the experiment has been unsuccessful because the participant classification as either a slow or fast typist could not be applied based on the CPM data. That is not to say the effect of the music does not differ between the skill level of the participants, rather that this avenue was unable to be explored in this experiment without introducing potentially strong levels of experimenter bias in the classification itself.

8.3.2 Limitations of This Experiment

The main limitation of this experiment is that the typing speed data could not be analysed with a speed classification applied to the participants, which has been shown to be important in the previous experiments. Clearly another, more formal, approach to classification is needed, perhaps focusing on just performing classification based on the without music CPM data alone or setting the classification prior to the experiment. However, as the threshold value for classification as a slow or fast typist in the previous experiments emerged through analysis of the achieved CPM data across multiple conditions, it did not seem reasonable to artificially constrain this classification to data from the without music condition alone post hoc.

A fully factorial within participants design collected data from 36 participants, which is a good number. Significant effects were identified, but with pairwise comparisons between 7 (or 6) conditions the required *p* value for significance in the post hoc tests could not be reached. In a design like this, many more participants would be needed to achieve significance at the post hoc level, or much stronger effects would need to be identified.

Chapter 9

PRELIMINARY DISCUSSION

The programme of research in this thesis began with a seemingly simple question – can music positively affect people's performance in and experience of mundane work related computing tasks? Five experiments and 275 participants later, the answer to my initial question has, unfortunately, not been quite so simple. Nevertheless, this thesis has moved the field forward by raising a number of important methodological issues and providing some insights into how music affects people when transcription typing.

This chapter begins by briefly revisiting the motivation behind and justification for focusing my research on the effect of music on transcription typing performance and experience. Then, the key outcomes and implications from the research are presented. First, the importance of participant classification as either a slow or fast typist is discussed. The speed classification was found repeatedly to influence the effects of music, so it is important to consider the implications carefully of this unexpected result. Next, the implications for future research and our understanding of how music affects people are presented in terms of the dimensions of music that had most interesting outcomes. The most interesting results involved the presence of vocals at different volume levels, the genre of music and the interaction between tempo and time signature, so these areas are discussed in some depth. The typist's skill level and task difficulty have also been important features of this work, so these concepts are then discussed across all the experiments. The classroom based methodology used in two experiments is somewhat novel but has been effective, so what I have learned about how to run experiments successfully using this sort of methodology is also reported. Finally, the limitations of the research presented in this thesis are considered.

9.1 Research Motivation

The introduction and literature review (in Chapters 1 and 2, respectively) showed clear evidence that people are frequently exposed to music in their day to day lives and that music affects both behaviour and experience. We know that people enjoy listening to music and as music can be accessed easily it should, in theory, be possible to exploit this affective medium. But, to exploit music's potential to affect people in a systematic manner there is a need for a better understanding of why and how music affects people. Although there is a lot of empirical evidence showing the effect of music on people's behaviour, theories explaining why music is such an affective medium are somewhat limited. In Chapter 10 some potentially relevant theories are considered in more depth. But, the literature review showed that much of the previous research investigating the effect of music on people has focused on manipulating the tempo or loudness of the music. Clearly there are more avenues to explore in the area of affective music, which was one of the initial aims of the work presented in this thesis.

Another key finding from the literature review was that sounds are typically incorporated in technological interfaces to explicitly communicate information to the user in the form of auditory messages. Sounds are rarely used to affect users directly with the exception of games, gambling and some previous research that I performed during my MSc. So, I identified an opportunity for further research investigating whether music could be used to directly influence people's behaviour and experiences when using a computer. Much of the previous research into how music affects people has focused on leisure activities, yet people spend much of their time at work, and often use a computer at work. So, the objective of investigating whether music can be used to positively affect performance at work related computing activities was established.

I identified that transcription typing would be a suitable task for investigating how music affects performance in mundane, work related computing tasks. Transcription typing was chosen after finding a paper which compared the effect of jazz and dirge music on the performance of trained typists (Jensen, 1931). Both genres of music affected performance, but in very different ways. The jazz music led to an increase in errors, while the dirge music decreased speed. Jensen's paper did not suggest reasons why these two genres of music had such distinct effects, but they were unmistakeably identified, and the method used in the experiment had strong levels of validity. After reading the paper, I was intrigued by the fact that music could have such a large impact on transcription typing performance, and even more so by the lack of explanation or discussion of why the effects due to jazz and dirge music differed so greatly. Given both music and typing are now such frequent features of many people's daily lives, investigating the effect of music on typing in a modern context seemed like the obvious next step. So, in my research I aimed to investigate how music affects transcription typing performance and experience. I also planned to expand the scope of research into affective music by exploring dimensions of music beyond tempo and loudness.

9.2 Speed Classification

Before exploring how the different dimensions of music affected typing performance and experience in the experiments in this thesis, one important outcome of this work needs to be addressed. That is, the classification of participants into speed groups was crucial when looking at the effects of music on typing speed, but this classification was seemingly less important when considering the effects of music on typing accuracy.

When taken as a single group of participants typing in English language tasks, the effect of music on typing speed did not always emerge in the inferential analysis. But, when the data was split into two groups based on applying a speed classification to each participant, there was evidence from all but Experiment 5 that some dimensions of the music was affecting the speed of the fast typists, in particular. However, when looking at the effect of music on typing accuracy in English, the results tended to be strongest when the data was treated as one speed group, and not split into two. That is, with the exception of Experiment 1, where vocals reduced typing accuracy for the fast typists only. To that end, the results that are discussed in more depth within this chapter will be taken from the CPM analysis including the speed classification, and from the error rate analysis without the speed classification applied.

With regard to the typing experience measures, in some experiments the speed classification mattered in the analysis of task difficulty or music distraction ratings, but in others the grouping did not affect the results. So, the discussion presented in this chapter will consider the strongest results from the experience measures.

9.3 MUSIC DIMENSIONS

Most of the research to date that investigates the effect of music on people has focused on manipulating loudness and tempo, as dimensions of the music, or self-selection of music as a dimension of the listener's relationship with the music. These manipulations are interesting, but it seemed to me that there are other dimensions of music which could be manipulated that may also affect performance and experience. So, in these experiments, a number of different dimensions of music were manipulated, as follows:

- Presence or absence of music (in Experiments 1, 2 and 5)
- vocals (in Experiments 1, 3 and 4)
- music style (in Experiment 1),
- genre (in Experiment 2),
- volume (in Experiments 3 and 4),
- time signature (in Experiment 5)
- tempo (in Experiment 5)

In terms of the scale of the implications from the results, the most important dimensions investigated in this thesis were the presence of vocals in the music and their interaction with volume, the effect of genre, and the interaction between tempo and time signature. The contributions based on these dimensions are discussed in more detail in the following sections.

9.3.1 THE EFFECT OF VOCALS AND VOLUME

In a verbal task like transcription typing, the obvious, at least to me, dimension of music which could potentially affect performance was whether or not the music contained vocals. Shortly before submitting this thesis, Dr Dean Burnett, a Neuroscience lecturer from Cardiff University, made a similar hypothesis in an article for the Guardian newspaper:

> "Something ... to look out for is music with catchy lyrics. Musical pieces without words might be better working companions, as human speech and vocalisation is something our brains pay particular attention to."
>
> (Burnett, 2016)

Although not specifically about typing, Burnett's hypothesis that performance at work might be reduced when accompanied by music with vocals and better with music without vocals, is the same hypothesis that underpins Experiments 1, 3 and 4. This quote demonstrates that I am not the only scientist who believes this is a hypothesis worth studying!

The results from Experiments 1 and 3 supported my hypothesis that vocals in music reduces typing performance. In terms of the negative effect of vocals in music on typing speed, the result was particularly clear for the fast typists but there was no evidence the slow typists experienced the same effect. The results from Experiment 3 also showed a reduction in the fast typists' speed due to vocals, but this effect was identified only for the high volume music, with no evidence of a similar effect when the fast typists were exposed to the music played at a lower volume. When looking at typing accuracy, there was also evidence that vocals in the music led to an increase in error rate. In Experiment 1, this effect was only evident for the fast typists, but in Experiment 3, where the sample size was larger, the effect of vocals on accuracy did not depend on the speed classification.

The results from Experiment 4 again showed that vocals reduced performance, but only in the high volume condition. When the music was played at a lower volume, vocals in the music actually improved performance (albeit not quite significantly, as $p = 0.050$). With low volume music the participants' typed faster with vocals, and in the hard tasks, they also made fewer errors. Contrastingly, when the music was played at a high volume, the fast typists' speed was lower with vocals in the music, and overall accuracy was reduced.

The primary reason for including volume as an independent variable (IV) in Experiments 3 and 4 was to identify whether an online methodology would be suitable without needing to take extra precautions to compensate for being unable to the control playback volume of the music. In this respect, the aims of these experiments have clearly been met, as I have collected evidence that the volume of music is a crucial factor which interacts with the effects caused by other dimensions of the music. Further work is needed to identify an approach to performing research using an online methodology which would mitigate for a researcher's inability to control the music's playback volume.

The reduced performance with vocals seen in the high volume conditions was predicted. So, in some regards, the improvement to performance with vocals in the low volume music is a more interesting result to explore further. I would not have been surprised if there was no difference in performance due to vocals in the low volume music, but the performance improvement is quite striking.

One explanation for these surprising results is that when music is played at a low volume the vocals are bringing the music to the attention of the typist as “our brains pay particular attention to [words]” (Burnett, 2016), meaning that any effect caused by the rest of the music dominates and reduces the effect of the vocals themselves. In essence, the vocals bring the music to the attention of the typist, allowing other dimensions of the music to affect their performance. Without the vocals in the music, the typist was able to block the music out and focus their attention predominantly on the transcription typing task itself. If this explanation is valid, then in this case the rest of the music was improving performance, so despite the expectation that vocals would reduce performance, the overall cumulative effect was an improvement.

Further, although the playback volume of the two versions of the music was the same, the average loudness of the with vocals version was louder than the without vocals track, which also increases the participants’ attention on the music rather than on the task. The vocals version is louder because including vocals in the mix adds more sounds to the track and increases the overall average loudness. Essentially, when manipulating the presence of vocals in the music by mixing down two versions of the song with and without the vocals tracks, I inadvertently manipulated the loudness of the music in these two experiments as well. So, because the with vocals track is slightly louder than the without vocals version, it is brought to the attention of the participants, and any effects due to the rest of the music are again strengthened.

The brain’s increased capacity to attend to vocals, together with the increase in loudness due to the mixing process, leads to improvements to performance when accompanied by the with vocals version of the music because any effects due to the rest of the music dominate. Clearly, this explanation is mere supposition at this stage, as there is no evidence that the rest of the music does indeed improve performance. But, as a theory to explain an unexpected result, it seems to have some merit.

When the fast typists were exposed to high volume with vocals music their performance was reduced, so there also seems to be a point at which the music becomes so loud that the vocals in the music dominates the rest of the music and reduces performance. In this case though, can I be sure that the performance reductions are due to the vocals? Or, could they again be due to the loudness difference? Or due to something else entirely?

With the benefit of hindsight, it is worth observing that the with and without vocals versions of the music differ in other dimensions besides the inclusion of vocals and the average loudness. For example, the without vocals version also has a less prominent melody, and the vocals perhaps mask some other aspects of the music like the rhythmic guitar. When taken together, this draws into question the internal validity of the experiments. Can I be sure that the changes observed in the dependent variables (DVs) are due to my manipulations of the IVs?

On the face of it, the following explanation for the unexpected result identified in Experiment 4 seems plausible. When played at a low volume, vocals bring the music to the attention of the listener sufficiently that the rest of the music affects performance. In this case, that effect was positive, perhaps due to the tempo of the music, or some other musicological dimension that improves performance. However, when the music is played at a higher volume, the vocals are too loud masking any effects due to the rest of the music and reducing performance.

Though, perhaps this explanation is too simplistic given my concerns regarding the manipulations described previously. It may be, that rather than treating the different dimensions of music as discrete, separate entities it is better conceptually to consider music as a whole unit, reminiscent of the term gestalt, defined as "a 'shape', 'configuration', or 'structure' which as an object of perception forms a specific whole or unity incapable of expression simply in terms of its parts (e.g. a melody in distinction from the notes that make it up)" (Gestalt, 2016).

By controlling and changing one dimension of the music, I was in fact controlling and changing multiple dimensions of the music simultaneously. Some dimensions may have larger effects than others, and their effects dominated the results. It is interesting that, to my knowledge, these sorts of inconsistencies and unexpected results have not been identified in other empirical research looking at the effect of music on people's behaviour. However, as stated previously, most of the studies controlled tempo or loudness as the IVs, which are dimensions of music which change the whole music piece when they are manipulated as they are not focused on an internal aspect of the music, such as the melody, rhythm or vocals. Perhaps the only manipulations of music which are appropriate from an internal validity perspective are those which change the piece of music as a whole, and do not involve manipulations of the smaller, more focused, internal dimensions. However, even that approach seems somewhat unsatisfying now as by doubling the tempo you would also double the amount of times the melody is heard, so can you be sure it is the tempo affecting performance, or could it be the repeated melody?

It is interesting to note that, without the unexpected result of an improvement to performance with vocals in the music at a low volume, this concept of treating music as an experiential gestalt in empirical work may not have been revealed. On the face of it, the experiments I have performed seemed sensible and valid and follow a well used, respected scientific approach of isolating and manipulating IVs. But, the unexpected result raises a lot of methodological questions which do require further investigation and perhaps a rethink of how music is manipulated in these types of experiments.

9.3.2 THE EFFECT OF GENRE

One way to avoid the issue of manipulating specific dimensions of music, is to abstract from dimensions of the music that involve properties of the sounds themselves by instead comparing the effect of higher order dimensions of music that are less clearly defined, such as genre. In Experiment 2, the focus was on how ambient, classical and rock music affected transcription typing performance when compared to a without music condition, across three different task difficulty levels. The reason why I used this genre abstraction as a manipulation in this experiment was not due to concerns about the internal validity when manipulating vocals. Instead, it was to mimic the approach taken in the Jensen experiment.

When looking at the tables of results from the pairwise comparisons of the three analysis paradigms, and most of the DVs in this experiment, something quite visually arresting emerged, which is repeated in Table 9.1.

Condition 1	Condition 2	Condition with Lower Mean	p
Without	Ambient	-	n.s.
Without	Classical	Classical	< 0.1
Without	Rock	Rock	< 0.1
Ambient	Classical	Classical	< 0.1
Ambient	Rock	Rock	< 0.1
Classical	Rock	-	n.s.

Table 9.1: Summary of post hoc tests for omnibus effect of music

Although the pairwise comparisons of the effect of music were not always significant at the Bonferroni adjusted level, they were often less than 0.1, and the pattern was almost always that shown in Table 9.1 for CPM with the speed classification applied and for typing accuracy and task difficulty ratings when treating the data as from a single group. The without music and ambient music conditions led to better performance and a perception that the tasks were easier than both the classical and rock music conditions. However, there were no obvious differences revealed between the without music and ambient music conditions, nor between the classical and rock music conditions. This experiment did not provide enough evidence to categorically make these claims, as the pairwise comparisons were not always significant at the Bonferroni adjusted level. But, there is a sufficient body of evidence to say that the without music and ambient music conditions are similar, and so are the rock and classical music conditions. And further, to state that these are interesting results which warrant further investigation in the future.

9.3.3 The Effect of Tempo and Time Signature

The similar effects on typing performance when exposed to classical and rock music conditions identified in Experiment 2 warranted further investigation because, in many regards, the music stimuli used to represent the two genres were quite strikingly different. Closer inspection of the stimuli from a musicological perspective revealed that although the tempo of the two pieces of music differed quite considerably, the difference in time signature may have led to the same number of emphasized beats in the two pieces of music. Previous research has investigated the effect of tempo, so this led to an idea to investigate the effect of both time signature and tempo on transcription typing performance together, in order to simultaneously explore how the frequency of emphasized beats affects performance. An experiment was conducted to specifically control and manipulate both time signature and tempo of the music (Experiment 5).

The time signature manipulation used is an interesting one, as although the two pieces of music were similar in style and composition, they also differed in dimensions other than time signature. So, any significant effects or interactions with the time signature may not simply be due to the difference in time signature, but perhaps also to do with other dimensions which were not controlled.

The results from Experiment 5 have interesting implications for people performing research investigating the effect of tempo on people. The effect of tempo and time signature on speed did not lead to any significant effects, but this may relate to my inability to properly classify participants as either slow or fast typists in this experiment, which reduces the power.

The particularly interesting result was in the analysis of the error rate percentage data, where the interaction between time signature and tempo was significant. Although the pairwise comparisons were not significant at the Bonferroni adjusted level, which is likely to be an issue of power, there were differences identified between the three different tempos in the 3/4 time but not in 4/4 time music. The fast tempo music in 3/4 time led to reduced accuracy over both the slow and mid tempo music in the same time signature. However, difference in accuracy were not constrained to just the 3/4 time music as there were also significant differences between the 3/4 time and 4/4 time fast and slow tempo pieces, but not the 4/4 time mid tempo music.

It is particularly interesting that overall there was no omnibus effect of tempo, as any effects due to the tempo depended on the time signature as well. It may not be the time signature it depended on, it may be some other difference between the 3/4 and 4/4 pieces. Nevertheless, it is interesting that tempo alone was not a significant factor and calls into question the validity of previous studies investigating the effect of tempo specifically and suggest that it is important to also report the time signature of the music stimuli.

9.4 Task Difficulty And Typist Skill

The speed classification was an important factor which affected the results in Experiments 1, 2, 3 and 4. Task difficulty was included as a manipulation in Experiments 2, 3 and 4.

What has clearly emerged through this work, is the idea that the scale of the effect of music, in terms of typing speed and how difficult the task was perceived as, and how distracting the music felt to be, relates to the participants' speed classification and the relationship between skill and difficulty of the task.

When the speed classification emerged clearly in Experiment 1, I considered whether this was the difference between touch typists and hunt and peck typists. But, informal observations of the typing approach taken by the participants in the laboratory Experiments 2 and 5, did not lead to any consistent categorisation of typist approach. In fact, in Experiment 5, one of the fastest typists only used two or three fingers! Typing has become an organic skill which is rarely taught formally, and people tend to develop their own idiosyncratic approaches.

The task difficulty manipulation used in Experiment 2 was somewhat successful. There was no discernible difference in the difficulty of typing between the simple and advanced English conditions, despite the differences in reading level of the material, but the difference between the advanced English and Dutch tasks was notable, while the effect of music depended on both the text and DV. The different texts used in Experiments 3 and 4 worked well to alter the intended difficulty of the tasks, and the results from these experiments clearly show that task difficulty has an important mediating effect on the scale of the impact of music on performance and experience. It is particularly interesting that some of the effects due to music were not constrained to only the easier tasks, with significant interactions and simple main affects in the Dutch transcription tasks as well.

These results have important consequences for the validity of the experiment tasks in relation to the original aim of the work in this thesis to investigate how music affects people at work-related computing tasks. If the significant effects were all constrained to the easier, English language tasks, then perhaps the significance of this work would be somewhat limited. But, as these experiments have demonstrated that, no matter what your skill level, when you are doing a difficult task the music that you are exposed to can affect your performance levels at that task. These results mean that although transcription typing is not necessarily a task that is often involved in people's day to day computing activities, we can extrapolate to consider that music may also affect people when undertaking more complex tasks as well.

9.5 Classroom Based Methodology

As well as the contributions relating to the effect of music dimensions on transcription typing performance and experience, this thesis reported experiments that were performed using a somewhat novel methodology with the classroom approach. The contributions regarding how to run an experiment successfully which has both pedagogic and research value using a classroom based approach are discussed in this section.

The key disadvantage and an area that the researcher must be acutely aware of and sensitive to in their design, is that by allowing students to perform the experiment on their behalf, they are relinquishing the tight control they would normally have when running an experiment in a traditional laboratory setting. As a result, the researcher must prepare the materials which the students use to perform the experiment very carefully, with instructions that are precise, unambiguous and concise. The researcher is trusting the students to collect their data, but as the students are unlikely to be personally invested in the research process, or the outcomes from the research, they cannot be guaranteed to take the same amount of care and attention to proceedings as the researcher themselves would take. This requirement to trust the students to perform the research well can be minimised through careful preparation of the materials and instructions to be used in the experiment. To successfully run an experiment in this environment, the researcher must provide the experimenters with precise instructions, bearing in mind that the students are unlikely to actually read each step carefully, so a simplistic approach is crucial. It is a difficult balance between providing enough information so that the experiment is performed correctly but not too much information that the students ignore or do not read the documentation carefully.

In the experiments performed using the classroom methodology in this thesis, the experimenter instructions were provided on paper. This decision was taken to ensure that the participant was not able to read the instructions or get information about the IVs that were being manipulated, but the experimenter would be provided with this information in an easy to use format. It might have been better to perform the experiment with instructions provided in a step by step fashion via the computer. But, care would have to be taken to make sure the participant did not receive information that could confound the experiment. Further, if instructions are provided in a way that the participant can read them easily (e.g. on the same computer they are undertaking the experiment on), the experimenter may be tempted to let the participant control the experiment themselves. This would reduce the educational and pedagogic merits of the exercise, as the main objective is for the experimenters to get experience of running an experiment. As these students have not been involved in designing the experiment, or creating the materials, it is important that they retain as much control over the running of the experiment as possible.

Additionally, when performing a study in an environment where approximately 30 participants are working simultaneously the data collection system and setup that is used must be robust to the pressures being applied. As with any experiment, a pilot version of the study is advised, particularly using pairs of students working together to run the experiment using the same approach and instruction sets to those that would be used in the real experiment. However, even with a pilot study, some opportunities for errors which can lead to a loss of data, or unsafe quality data can result. The simplest way to ensure that the data collected is safe and there is no risk of cross contamination is by collecting all data on paper with no automation. Though, while this may be appropriate in some experimental contexts, it would have been insufficient to achieve the aims of this research.

In the classroom environment, the experimental design also needs to be simple with as few between groups factors as possible as ensuring adequate counterbalancing of experimental conditions can also be a complex procedure.

The external validity of the data collected through an experiment performed in a classroom setting is limited as the participants are going to be from the same cohort. As long as this is acknowledged by the researcher and claims made from the outcomes are appropriately restricted by this factor then the data can still be used for insights that inform additional research. As my research was being performed in a Computer Science department at the University of York, the cohort of students participating was predominantly male students from the UK, though there were also a few participants (and experimenters) who were non-native English speakers.

A final disadvantage to this classroom based approach is that the environment is likely to be noisy and interruptions are a likely factor. This concern is particularly relevant to experiments where the IVs relate to sounds the participants were exposed to and where participant concentration is a vital requirement to ensure high quality data is collected.

The impact some of these disadvantages to approach can be minimised by taking a meticulous approach to the preparation of the experimenter materials. Though, it is not always possible to foresee all the different ways that participants and experimenter could conduct the experiment wrong. As such, from the researcher's point of view, the set up cost in terms of preparation time using a classroom based methodology approach is considerably larger than required to conduct the equivalent laboratory based experiments.

One other area that can, potentially, cause issues for researcher's performing studies using this methodological approach is with the unpredictability of the participant's capabilities. For example, in a gaming context, it would be impossible to control how experienced the participants are at gaming. In the context of typing studies, such as those presented in this thesis, the limitation of being unable to predict how many of the participants would be classified as slow and as fast typists resulted in data that was overloaded by participants who typed slowly. If this occurred in a laboratory setting, or using online evaluation methods, further participants could simply be recruited until the groups had more suitable numbers of participants within them. However, when running research in the classroom, the researcher has a single attempt to make it work for them.

Nevertheless, by using this classroom approach a large volume of data can be collected over a relatively short period of time. I was able to collect data from around 60 participants during two 2 hour practical classes. Therefore, despite the large cost to the researcher for preparation, the benefits of such a large quality of data over such a short period of time makes the endeavour worth the effort, particularly if the experiment setup can be reused in future studies. However, as the researcher has a single attempt (or in my case, two attempts in very quick succession as the large cohort of students were split into two practical classes) to get the study set up correctly. As such, it is a high risk strategy which requires significant depth of thought and careful planning to gather safe, empirically sound, valid data.

Researchers must also be aware of ethical implications of performing research in this manner. As with any experimental research involving people, informed consent must be sought and the participants should have the opportunity to withdraw their participation at any time, without giving a reason. The latter issue is potentially hard to manage as it is important to ensure that any student who does not wish to take part, either as participant or experimenter, still receives the same level of educational experience and their education is not penalised by not being involved in the experiment.

9.6 Limitations Of the Research

The research presented in this thesis has a few limitations which need to be acknowledged and addressed, most of these have been discussed in previous chapters but the implication of each limitation is discussed in more detail in this section.

One recurring limitation of the experiments in this thesis is the modest sample size of participants who could be classified as fast typists, even in the larger classroom based experiments. Although this sample size issue did not cause big problems with regard to the accuracy data, which were not affected by the speed classification, the typing speed analysis did show evidence of dependency on this classification, so the smaller group of fast typists is a limitation of the work. This limitation was especially evident from the experiments where significant interactions were identified that included the speed classification IV, but the simple post hoc tests did not always reveal statistically significant differences, indicating a lack of power in the experiment.

Another limitation is that all the participants were involved in academia in some respect. Would these results generalise to people working at computers in non-academic careers? Ideally, yes. But, future experiments should aim to expand the participant pool and setting of the work beyond academia, to improve the external validity of the work.

One other issue is that in using the classroom methodology a without music condition could not be used without introducing further possible confounds. So, Experiments 3 and 4 do not have a without music condition. This approach improves the validity of the experiments but means that the results cannot show improvements or reductions in performance or experience over transcription typing without exposure to music.

The transcription typing task itself also has somewhat restricted ecological validity as it was contrived in order to be tightly controlled. However, this is not a particularly large concern in this context as there is already evidence that people exploit music while working because they believe it improves their concentration. So, although the task is somewhat contrived, the context has merit and the task is representative of something that people do in the real world, i.e. listen to music while performing mundane work related computing tasks.

Chapter 10

SYNTHESIS AND CONCLUSIONS

The preliminary discussion presented in Chapter 9 focused on summarising the results and outcomes from the five experiments presented in the thesis and considering some potential, surface-level explanations for the results. This chapter begins by distilling the complex results and outcomes from the experiments into summary tables focused on comparing results across the experiments. Then, these distilled results are explored in relation to the following potentially relevant theoretical and conceptual frameworks:

- models of typing performance,
- working memory, unattended speech and music effects,
- the multiple resource model,
- the relationship between arousal and performance,
- entrainment.

Each of these conceptual frameworks is first considered through presenting a critical analysis of relevant literature. Then, the appropriateness of the framework for explaining the results from the experiments in this thesis is explored by hypothesising what I would expect to see if the framework applied, and how these expected effects differ to the results that were identified.

Following exploration of these conceptual frameworks, suggestions are made to take this body of research work forward in the future as there does seem to be potential to exploit music to affect people's performance at work related computing tasks, both positively and negatively. The key contributions of this thesis are then summarised. Finally, this chapter, and indeed the thesis, ends with a reflection on the success of this programme of research in relation to the original objectives and those that emerged throughout the process.

10.1 DISTILLATION OF RESULTS

The complex results from the experiments presented in this thesis have been condensed into a series of tables which focus on answering whether each experiment has provided evidence to support the key hypotheses investigated in this thesis. Tables 10.1 to 10.10 (pages 328 to 344) summarise the results from each relevant experiment in relation to the following questions:

- Is there evidence of the presence of music affecting typing performance (Table 10.1) and experience (Table 10.2)?
- Is there evidence of vocals affecting typing performance (Table 10.3) and experience (Table 10.4)?
- Is there evidence of different styles or genres of music affecting typing performance (Table 10.5) and experience (Table 10.6)?
- Is there evidence of the volume of music affecting typing performance (Table 10.7) and experience (Table 10.8)?
- Is there evidence of the tempo, time signature or frequency of the emphasized beats affecting typing performance (Table 10.9) and experience (Table 10.10)?

Note that some of the tables extend beyond a single page, and at the bottom of each page a brief summary of the independent variables used in that experiment is included. Key findings from the experiments are highlighted in bold, blue text.

	Is there evidence of the presence of music affecting typing performance? For CPM	For Transformed Error Rate
Experiment 1	**No.**	**No.**
Experiment 2	**Yes.** Evidence of a reduction in CPM with presence of music. When analysing the data from the advanced English task only, with and without typing speed classification applied there was a: • Significant main effect of music genre. **CPM without music significantly higher than with rock music accompaniment.** When analysing the data from the advanced English task only, with typing speed classification applied there was a: • Significant interaction between music and speed group and a significant simple main effect of music for fast typists but not slow. **The fast typists' CPM without music was significantly higher than with both rock and classical music accompaniment.** When analysing the data from the simple versus advanced English tasks group there were: • No significant main effects or interactions.	**Yes.** Evidence of an increase in error rate with presence of music. When analysing the data from the advanced English task only, there were: • no significant main effects or interactions When analysing the data from the simple versus advanced English tasks group, with and without typing speed classification applied there was a: • Significant main effect of music genre. **The error rate was significantly lower without music than when accompanied by classical music (significant at non-adjusted level). There was a trend towards a significant difference in error rate between the without music and rock music conditions, with higher error rates in the latter.**

Experiment 1: Vocals (2 levels: with and without), Style (2 levels: alt rock and pop rock) plus without music condition, used advanced English text only.
Experiment 2: Music (4 levels: without music, ambient, classical and rock), Task Difficulty (3 levels: simple English, advanced English and Dutch)
Experiment 3: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and Dutch)
Experiment 4: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and advanced English with pseudowords)
Experiment 5: Tempo (3 levels: low, mid and high), Time signature (2 levels: 3/4 and 4/4) plus without music condition, used advanced English text only. Also analysed frequency of emphasized beats

	Is there evidence of the presence of music affecting typing performance?	
	For CPM	For Transformed Error Rate
Experiment 2	When analysing the data from the advanced English versus Dutch tasks group, with typing speed classification applied there was a: • Significant interaction between music, text and speed group. The simple interaction between music and text was not significant for either slow or fast typists. The simple interaction between music and speed group was significant in advanced English but not in the Dutch tasks. The simple simple main effect of music was significant for fast typists, but not for slow typists. **The fast typists' CPM, in the advanced English task, without music was higher than with both rock and classical music accompaniments (at non-adjusted significant level).**	When analysing the data from the advanced English versus Dutch tasks group, with and without typing speed classification applied there was a: Significant main effect of music genre. **The error rate was significantly lower without music than when accompanied by classical music (significant at non-adjusted level). There was a trend towards a significant difference in error rate between the without music and rock music conditions, with higher error rates in the latter.**
Experiment 5	**No.**	**Yes.** Some weak evidence of an increase in error rate with presence of music. Without typing speed classification applied there was a: • Trend towards a significant main effect of music. **Pairwise comparisons revealed a significant difference between the without music condition and 3/4 time signature, fast tempo condition at the non-adjusted level.**

Table 10.1: Summary of evidence of the effect of the presence of music on typing performance

	Is there evidence of the presence of music affecting typing experience?	
	For Task Difficulty Ratings	For Music Distraction Ratings
Experiment 1	**No.**	N/A
Experiment 2	**Yes.** Evidence of an increase in task difficulty ratings with presence of music. When analysing the data from the advanced English task only, both with and without typing speed classification applied there was a: • Significant main effect of music genre. **Task difficulty ratings without music were significant higher than when accompanied by both rock and classical music.** When analysing the data from the simple versus advanced English tasks group, both with and without typing speed classification there was a: • Significant effect of music genre. **Task difficulty ratings without music were significant higher than with classical music and higher than with rock music at the non-adjusted level.** When analysing the data from the advanced English versus Dutch tasks group, both with and without typing speed classification applied there was a: • Significant effect of music genre and a significant interaction between music genre and text. In the advanced English tasks the effect of music genre was significant. **Pairwise comparisons revealed the without music condition was rated as lower difficulty than the ambient, music and rock conditions (but at a non-adjusted significant level only) with English tasks.**	N/A

Experiment 1: Vocals (2 levels: with and without), Style (2 levels: alt rock and pop rock) plus without music condition, used advanced English text only.
Experiment 2: Music (4 levels: without music, ambient, classical and rock), Task Difficulty (3 levels: simple English, advanced English and Dutch)
Experiment 3: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and Dutch)
Experiment 4: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and advanced English with pseudowords)
Experiment 5: Tempo (3 levels: low, mid and high), Time signature (2 levels: 3/4 and 4/4) plus without music condition, used advanced English text only. Also analysed frequency of emphasized beats

	Is there evidence of the presence of music affecting typing experience?	
	For Task Difficulty Ratings	For Music Distraction Ratings
Experiment 5	**No.**	N/A

Table 10.2: Summary of evidence of the effect of the presence of music on typing experience

	Is there evidence of vocals affecting typing performance?	
	For CPM	For Transformed Error Rate
Experiment 1	**Yes.** Some weak evidence of a reduction in CPM when the accompanying music contained vocals. With typing speed classification applied there was a: • Trend towards significant interaction between music style and vocals. But, the simple effects analysis did not reveal any significant differences. • Significant interaction between music style, vocals and speed group with a significant simple interaction between music style and vocals for the fast typists but not for slow typists. The simple simple main effect of vocals had a trend towards a significant effect for pop rock but not alt rock style music. **The fast typists' CPM, was lower when the for pop rock music contained vocals.**	**Yes.** Evidence of an increase in error rate when the accompanying music contained vocals. With typing speed classification applied there was a: • Significant interaction between vocals and speed group with a significant simple main effect of vocals for fast typists but not for slow typists. **The fast typists' error rate was higher when the accompanying music contained vocals.**
Experiment 3	**Yes.** Evidence of a reduction in CPM when the accompanying music contained vocals. With typing speed classification applied there was a: • Significant main effect of vocals. **CPM was higher with music that did not contain vocals.** • Significant interaction between vocals and volume with a significant simple main effect of vocals for the high volume music but not low volume. **With high volume music, CPM was higher when the music did not contain vocals.** • Significant interaction between vocals and speed group with a significant simple main effect of vocals for fast typists but not slow typists. **Fast typists' CPM was higher when music did not contain vocals.**	**Yes.** Evidence of an increase in error rate when the accompanying music contained vocals. Without typing speed classification applied there was a: • Significant main effect of vocals. **Error rate was higher with music that contained vocals.** With typing speed classification applied there was a: • Trend towards a significant main effect of vocals. **Error rate was higher with music that contained vocals.** • Trend towards a significant interaction between text and vocals with a significant simple main effect of vocals when the typing task was in Dutch, but not in English. **When typing in Dutch, error rate was higher when the music contained vocals.**

Experiment 1: Vocals (2 levels: with and without), Style (2 levels: alt rock and pop rock) plus without music condition, used advanced English text only.
Experiment 2: Music (4 levels: without music, ambient, classical and rock), Task Difficulty (3 levels: simple English, advanced English and Dutch)
Experiment 3: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and Dutch)
Experiment 4: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and advanced English with pseudowords)
Experiment 5: Tempo (3 levels: low, mid and high), Time signature (2 levels: 3/4 and 4/4) plus without music condition, used advanced English text only. Also analysed frequency of emphasized beats

	Is there evidence of vocals affecting typing performance? For CPM	For Transformed Error Rate
Experiment 3	• Trend towards a significant interaction between text and vocals with a significant simple main effect of vocals when the typing task was in English, but not in Dutch. **With an English typing task, CPM was higher when the music did not contain vocals.** • Significant interaction between vocals, volume and speed group with a significant simple interaction between vocals and volume for fast typists but not slow typists. For fast typists, simple simple main effect of vocals was significant with high volume music, but not low. **With high volume music, the fast typists' CPM was higher when the accompanying music did not contain vocals.**	• Significant interaction between text, vocals and speed group but the simple interaction between vocals and text was not significant for either speed group and the simple interaction between vocals and speed group not significant for either text.
Experiment 4	**Yes.** Evidence of both an increase and decrease in CPM performance when the accompanying music contained vocals. For the low volume group, with typing speed classification applied there was a: • Trend towards a significant main effect of vocals (p= 0.050). **CPM was higher with music that did contain vocals.** • Trend towards a significant interaction between text, vocals and speed group. The simple interaction between vocals and text not significant for either slow or fast typists. But, the simple interaction between vocals and speed group was not significant for the medium difficulty text, but not the hard task. The simple simple main effect of vocals was significant in medium difficulty tasks for fast typists but not slow. **Fast typists' CPM was higher when the accompanying music contained vocals in the medium difficulty tasks.** For the high volume group, with typing speed classification applied there was a: • Significant interaction between vocals and speed group with a significant simple main effect of vocals for fast typists but not slow typists. **Fast typists' CPM was higher without vocals in the music.** • Trend towards a significant interaction between text and vocals with a significant simple main effect of vocals for the hard tasks but not the medium difficulty tasks. **In the hard tasks, CPM was higher without vocals in the music.**	**Yes.** Evidence of a decrease in error rate performance only when the accompanying music contained vocals. For the low volume group, without the typing speed classification applied there was a: • Significant interaction between text and vocals with a significant simple main effect of vocals in the hard tasks but not the medium difficulty tasks. **In the hard tasks, error rate was higher without vocals in the music.** For the high volume group, there were no significant effects or interactions on error rate identified involving the vocals dimension.

	Is there evidence of vocals affecting typing performance?	
	For CPM	For Transformed Error Rate
Experiment 4	• Trend towards a significant interaction between text, vocals and speed group. The simple interaction between vocals and text not significant for either slow or fast typists. But, the simple interaction between vocals and speed group was significant in medium difficulty tasks but not hard difficulty tasks. In the medium difficulty tasks, the simple simple main effect of vocals was significant for fast typists but not slow. **Fast typists' CPM was higher in the medium difficulty tasks when the music did not contain vocals.** In the combined analysis across both volume groups, with typing speed classification applied there was a: • Significant interaction between text and vocals with the simple main effect of vocals approaching significance in the medium difficulty tasks, but not the hard tasks. **In the medium difficulty tasks, CPM was lower without vocals in the music.** • Significant interaction between text, vocals and speed group with a significant simple interaction between text and vocals for fast typists but not for slow typists. For the fast typists, the simple simple main effect of vocals was significant in both the medium and hard difficulty tasks. **Fast typists' CPM was higher with vocals in the medium difficulty tasks. In the hard difficulty tasks, fast typists' CPM was higher without vocals in the music.**	In the combined analysis across both volume groups, with typing speed classification applied there was a: • Significant interaction between text and vocals. The simple main effect of vocals approached significance in the medium difficulty tasks and was significant in the hard tasks. **In the medium difficulty tasks, error rate was lower without vocals in the music but in the hard difficulty tasks, error rate was higher without vocals in the music.**

Table 10.3: Summary of evidence of the effect of vocals on typing performance

Experiment 1: Vocals (2 levels: with and without), Style (2 levels: alt rock and pop rock) plus without music condition, used advanced English text only.
Experiment 2: Music (4 levels: without music, ambient, classical and rock), Task Difficulty (3 levels: simple English, advanced English and Dutch)
Experiment 3: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and Dutch)
Experiment 4: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and advanced English with pseudowords)
Experiment 5: Tempo (3 levels: low, mid and high), Time signature (2 levels: 3/4 and 4/4) plus without music condition, used advanced English text only. Also analysed frequency of emphasized beats

	Is there evidence of vocals affecting typing experience?	
	For Task Difficulty Ratings	For Music Distraction Ratings
Experiment 1	**No.**	**Yes.** Evidence of an increase in distraction ratings when the accompanying music contained vocals. Without typing speed classification applied there was a: • Significant main effect of vocals. **Music without vocals was rated as less distracting than music with vocals.**
Experiment 3	**Yes.** Evidence of an increase in task difficulty ratings when the accompanying music contained vocals. Without typing speed classification applied there was a: • Significant main effect of vocals. **The tasks were rated as more difficult when the music contained vocals.** With typing speed classification applied there was a: • Significant interaction between text, vocals, volume and speed group. Follow up non-parametric tests revealed a number of significant effects. **The slow typists, when typing in Dutch with high volume music, rated the task as easier without vocals in the music. The fast typists, when typing in English with both high and low volume music, rated the task as easier without vocals in the music.**	**Yes.** Evidence of an increase in distraction ratings when the accompanying music contained vocals. Without typing speed classification applied there was a: • Trend towards a significant interaction between text and vocals. Follow-up non-parametric tests revealed a significant effect of vocals in the English tasks but not Dutch. **The music was rated as more distracting when it contained vocals when typing in English.**
Experiment 4	**No.**	**Yes.** Evidence of an increase in distraction ratings when the accompanying music contained vocals. For the low volume group, with the typing speed classification applied there was a: • Significant interaction between vocals and speed group. Follow up non-parametric tests revealed a significant effect of vocals for the slow typists only.

	Is there evidence of vocals affecting typing experience?	
	For Task Difficulty Ratings	For Music Distraction Ratings
Experiment 4		**The slow typists rated the music with vocals as more distracting than the music without vocals.** For the high volume group, with the typing speed classification applied there was a: • Trend towards a significant interaction between vocals and speed group. Follow up non-parametric tests revealed a significant effect of vocals for the fast typists only. **The fast typists rated the music with vocals as more distracting than the music without vocals.**

Table 10.4: Summary of evidence of the effect of vocals on typing experience

Experiment 1: Vocals (2 levels: with and without), Style (2 levels: alt rock and pop rock) plus without music condition, used advanced English text only.
Experiment 2: Music (4 levels: without music, ambient, classical and rock), Task Difficulty (3 levels: simple English, advanced English and Dutch)
Experiment 3: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and Dutch)
Experiment 4: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and advanced English with pseudowords)
Experiment 5: Tempo (3 levels: low, mid and high), Time signature (2 levels: 3/4 and 4/4) plus without music condition, used advanced English text only. Also analysed frequency of emphasized beats

	Is there evidence of different styles or genres of music affecting typing performance?	
	For CPM	For Transformed Error Rate
Experiment 1	**Yes.** Some weak evidence that the effect of vocals is different between the pop rock and alt rock style music and that CPM is lower for alt rock style music than pop rock when the music is instrumental. With typing speed classification applied, there was a: • trend towards a significant interaction between music style and vocals. The simple main effect of vocals was bit significant for either style. • significant interaction between music style, vocals and speed group with a significant simple interaction between music style and vocals for fast typists but not slow. The simple simple main effect of vocals showed a trend towards significance for pop rock music, but not for alt rock music. **Fast typists' CPM was lower when accompanied by pop rock music that contained vocals.** Also, the simple simple main effect of music style approached significance without vocals but not with vocals. **Fast typists' CPM was lower when accompanied by without vocals music that was in the alt rock style than rather than the pop rock style.**	**No.**
Experiment 2	**Yes.** Evidence of an effect due to genre on CPM, with higher CPM when accompanied by ambient music in comparison to both rock and classical music. When analysing the data from the advanced English task only, with typing speed classification applied there was a: • significant interaction between music and speed classification with a significant simple main effect of music for fast typists but not slow typists. Pairwise comparisons revealed a significant difference between ambient and rock music (at Bonferroni-adjusted level) and between ambient and classical music (at non-adjusted level). **Fast typists' CPM was higher with ambient music than with classical or rock music.** When analysing the data from the simple versus advanced English tasks group there were: • No significant main effects or interactions specifically involving the 3 genres.	**Yes.** Some weak evidence that error rates differ due to music genre, with lower error rates when accompanied by ambient music in comparison to both rock and classical music. When analysing the data from the advanced English task only, without typing speed classification applied there was a: • significant main effect of music. Pairwise comparisons revealed a significant difference between ambient and rock music and between ambient and classical music at the non-adjusted level. **Error rates were lower when accompanied by ambient music than with either rock or classical music.** When analysing the data from the simple versus advanced English tasks group, without typing speed classification applied, there was a: • significant main effect of music. Pairwise comparisons revealed a significant

	Is there evidence of different styles or genres of music affecting typing performance?	
	For CPM	For Transformed Error Rate
Experiment 2		difference between ambient and rock music and between ambient and classical music at the non-adjusted level. **Error rates were lower when accompanied by ambient music than with either rock or classical music.**
	When analysing the data from the advanced English versus Dutch tasks group, with typing speed classification applied there were: • No significant main effects or interactions specifically involving the 3 different genres.	When analysing the data from the advanced English versus Dutch tasks group, with and without typing speed classification applied there was a: • significant main effect of music. Pairwise comparisons revealed a significant difference between ambient and rock music (at Bonferroni-adjusted level) and between ambient and classical music (at non-adjusted level). **Error rates were lower when accompanied by ambient music than with either rock or classical music.**

Table 10.5: Summary of evidence of the effects of music style and genre on typing performance

Experiment 1: Vocals (2 levels: with and without), Style (2 levels: alt rock and pop rock) plus without music condition, used advanced English text only.
Experiment 2: Music (4 levels: without music, ambient, classical and rock), Task Difficulty (3 levels: simple English, advanced English and Dutch)
Experiment 3: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and Dutch)
Experiment 4: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and advanced English with pseudowords)
Experiment 5: Tempo (3 levels: low, mid and high), Time signature (2 levels: 3/4 and 4/4) plus without music condition, used advanced English text only. Also analysed frequency of emphasized beats

	Is there evidence of different styles or genres of music affecting typing experience?	
	For Task Difficulty Ratings	For Music Distraction Ratings
Experiment 1	**No.**	**No.**
Experiment 2	**Yes.** Evidence of an effect due to genre on task difficulty ratings When analysing the data from the advanced English task only, without typing speed classification applied there was a: • significant main effect of music. Pairwise comparisons revealed a significant difference between ambient and rock music and between ambient and classical music at the non-adjusted level. **Task difficulty ratings were lower when accompanied by ambient music than with either rock or classical music.** When analysing the data from the simple versus advanced English tasks group, both with and without typing speed classification applied, there was a: • significant main effect of music. Pairwise comparisons revealed a significant difference between ambient and rock music (at non-adjusted level) and between ambient and classical music (at Bonferroni-adjusted level). **Task difficulty ratings when accompanied by ambient music were lower than with either rock or classical music.**	**Yes.** Evidence of an effect due to genre on music distraction ratings. When analysing the data from the advanced English task only, without typing speed classification applied there was a: • significant main effect of music. Pairwise comparisons revealed significant differences between ambient and rock, ambient and classical, and between classical and rock music (all at Bonferroni-adjusted level). **Music distraction ratings were lower with ambient music than both classical and rock music, and distraction ratings with classical music were lower than with rock music.** When analysing the data from the simple versus advanced English tasks group, both with and without typing speed classification applied, there was a: • significant main effect of music. This analysis was not followed up due to the: • significant interaction between music and text. In the simple English tasks, the effect of music was significant. Pairwise comparisons revealed a significant difference in distraction ratings between the ambient and classical music conditions (at Bonferroni-adjusted level) and between the ambient and rock music conditions (at non-adjusted level). **In the simple English tasks, the music distraction ratings were lower with ambient music than both classical and rock music.** In the advanced English task, the effect of music was significant. Pairwise comparisons revealed significant differences between ambient and rock, ambient and classical, and between classical and rock music (all at Bonferroni-adjusted level). **Music distraction ratings were lower with ambient music than both classical and rock music, and distraction ratings with classical music were lower than with rock music.**

	Is there evidence of different styles or genres of music affecting typing experience?	
	For Task Difficulty Ratings	For Music Distraction Ratings
Experiment 2	When analysing the data from the advanced English versus Dutch tasks group, with and without typing speed classification applied there was a: • significant effect of music. This analysis was not followed up due to the: • significant interaction between music and text. In the advanced English tasks, the effect of music was significant, but it was non-significant in the Dutch tasks. Pairwise comparisons revealed a trend towards a significant difference in task difficulty ratings between the ambient and rock music conditions only. **In the advanced English tasks, difficulty ratings when accompanied by ambient music were lower than when accompanied by rock music.**	When analysing the data from the advanced English versus Dutch tasks group, with and without typing speed classification applied there was a: • significant main effect of music. This analysis was not followed up due to the: • trend towards a significant interaction between music and text. In the advanced English tasks, the effect of music was significant. Pairwise comparisons revealed a significant difference between ambient and rock music and between classical and rock music (at Bonferroni-adjusted levels). **In the advanced English tasks, music distraction ratings when accompanied by rock music were higher than both ambient and classical music.** In the Dutch tasks, there was a trend towards a significant effect of music. Pairwise comparisons revealed a significant difference between ambient and classical music and between ambient and rock music (at non-adjusted levels). **In the Dutch tasks, music distraction ratings when accompanied by ambient music were lower than both classical and rock music.**

Table 10.6: Summary of evidence of the effects of music style and genre on typing experience

Experiment 1: Vocals (2 levels: with and without), Style (2 levels: alt rock and pop rock) plus without music condition, used advanced English text only.
Experiment 2: Music (4 levels: without music, ambient, classical and rock), Task Difficulty (3 levels: simple English, advanced English and Dutch)
Experiment 3: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and Dutch)
Experiment 4: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and advanced English with pseudowords)
Experiment 5: Tempo (3 levels: low, mid and high), Time signature (2 levels: 3/4 and 4/4) plus without music condition, used advanced English text only. Also analysed frequency of emphasized beats

	Is there evidence of the volume of music affecting typing performance?	
	For CPM	For Transformed Error Rate
Experiment 3	**Yes.** Evidence that volume has an effect on CPM. With typing speed classification applied there was a: • Significant interaction between vocals and volume. Simple main effect of vocals significant for high volume music but not low. **For high volume music, CPM was higher when music did not contain vocals.** • Significant interaction between vocals, volume and speed group. Simple interaction between vocals and volume was significant for fast typists but not slow. For fast typists, simple simple main effect of vocals was significant with high volume music, but not low. **With high volume music, fast typists' CPM was higher when music did not contain vocals.** • Significant interaction between text and volume. Simple main effect of volume was significant in Dutch but not English tasks. **When typing in Dutch, CPM was higher when accompanied by low volume music.** • Significant interaction between text, volume and speed group. Simple interaction between text and volume was significant for fast typists but not slow. Simple simple main effect of volume was significant in Dutch but not English tasks. **When typing in Dutch, the fast typists' CPM was higher when accompanied by low volume music.** • Trend towards a significant interaction between volume and speed group. The simple main effect of volume was significant for fast typists only. **Fast typists' CPM was higher when accompanied by low volume music.**	**No.**
Experiment 4	**Yes.** Evidence that volume has an effect on CPM. In the combined analysis across both volume groups, with typing speed classification applied there was a: • Significant interaction between vocals and volume. There was a trend towards a significant simple main effect of vocals with the low volume music ($p = 0.050$) but not the high volume music. **With high volume music, CPM was higher when the music contained vocals.**	**Yes.** Evidence that volume has an effect on error rate. In the combined analysis across both volume groups, without typing speed classification applied there was a: • Significant main effect of volume. **Error rates were higher with high volume music.**

	Is there evidence of the volume of music affecting typing performance?	
	For CPM	For Transformed Error Rate
Experiment 4	• Significant interaction between vocals, volume and speed group with a significant simple interaction between vocals and volume for the fast typists but not the slow. The simple simple main effect of volume was significant with vocals but not without vocals. **The fast typists' CPM was higher when the with vocals music was played at low rather than high volume.**	

Table 10.7: Summary of evidence of the effect of the volume of music on typing performance

Experiment 1: Vocals (2 levels: with and without), Style (2 levels: alt rock and pop rock) plus without music condition, used advanced English text only.
Experiment 2: Music (4 levels: without music, ambient, classical and rock), Task Difficulty (3 levels: simple English, advanced English and Dutch)
Experiment 3: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and Dutch)
Experiment 4: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and advanced English with pseudowords)
Experiment 5: Tempo (3 levels: low, mid and high), Time signature (2 levels: 3/4 and 4/4) plus without music condition, used advanced English text only. Also analysed frequency of emphasized beats

	Is there evidence of the volume of music affecting typing experience?	
	For Task Difficulty Ratings	For Music Distraction Ratings
Experiment 3	**Yes.** Evidence that volume has an effect on task difficulty ratings. With typing speed classification applied there was a: • significant interaction between vocals, volume and speed group. This analysis was not followed up due to the: • significant interaction between text, vocals, volume and speed group. The effect of vocals on task difficulty ratings was shown to depend on a combination of speed classification and volume condition. **At both low and high volumes, when typing in English, the fast typists rated the task as significantly harder when the music contained vocals. At a high volume, when typing in Dutch, the slow typists rated the task as significantly harder when the music contained vocals.**	**Yes.** Evidence that volume has an effect on music distraction ratings. Without typing speed classification applied there was a: • significant main effect of volume. **High volume music was perceived as more distracting than low volume music.**
Experiment 4	**No.**	**Yes.** Some evidence that volume has an effect on music distraction ratings. Without typing speed classification applied there was a: • significant effect of text when accompanied by low volume music only**. With low volume music, the music was perceived as more distracting in the hard tasks than the medium difficulty tasks. There was no evidence of differences in the task difficulty ratings with the high volume music.**

Table 10.8: Summary of evidence of the effect of volume of music on typing experience

	Is there evidence of the tempo or time signature affecting typing performance?	
	For CPM	For Transformed Error Rate
Experiment 5	**No.**	**Yes.** Evidence that tempo and time signature affect transformed error rate. Without typing speed classification applied there was a: • Significant interaction between time significant and tempo. Simple main effect of tempo was significant with 3/4 time signature but not 4/4. In the 3/4 time signature music, pairwise comparisons revealed significant differences at non-adjusted level between fast tempo music and both mid and slow tempo music. **In the 3/4 time signature music, error rate was higher with fast tempo music than with either slow or mid tempo music.** The simple simple main effect of time signature was significant for fast tempo music only. **Transformed error rate was higher with music that was played at a fast tempo in 3/4 than in 4/4.**

Table 10.9: Summary of evidence of the effects of tempo and time signature on typing performance

Experiment 1: Vocals (2 levels: with and without), Style (2 levels: alt rock and pop rock) plus without music condition, used advanced English text only.
Experiment 2: Music (4 levels: without music, ambient, classical and rock), Task Difficulty (3 levels: simple English, advanced English and Dutch)
Experiment 3: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and Dutch)
Experiment 4: Vocals (2 levels: with and without), Volume (2 levels: low and high), Task Difficulty (2 levels: advanced English and advanced English with pseudowords)
Experiment 5: Tempo (3 levels: low, mid and high), Time signature (2 levels: 3/4 and 4/4) plus without music condition, used advanced English text only. Also analysed frequency of emphasized beats

	Is there evidence of the tempo or time signature affecting typing experience?	
	For Task Difficulty Ratings	For Music Distraction Ratings
Experiment 5	**No**	**Yes.** Evidence that tempo affects transformed error rate. Without typing speed classification applied there was a: • Significant effect of tempo. Pairwise comparisons revealed significant difference between fast and slow tempo music and between fast and mid tempo music. **Fast tempo music was rated as more distracting than both mid and slow tempo music.**

Table 10.10: Summary of evidence of the effect of tempo and time signature on typing experience

10.2 APPLYING THE FRAMEWORKS

Having condensed the results from across the five experiments into summary tables, it is now possible to consider whether some different theoretical and conceptual frameworks could be used to explain some of these complex results, and if so, how.

10.2.1 TYPING PERFORMANCE

The first theoretical framework that will be explored focuses on understanding what is known about the process of transcription typing, by considering the research that has been conducted around how people type. There has been a considerable amount of research attention given to understanding typing processes coming from a number of different perspectives, including perceptual, cognitive and motoric angles. This body of research literature around transcription typing led to the development of a "four-component heuristic model" (Salthouse, 1986, p. 303) that can be used to explain the processes involved. The four components in the model are *input* (grouping the text to be typed into chunks), *parsing* (decomposing the chunks into characters), *translation* (converting the character to be typed into a movement) and *execution* (pressing the key). The Salthouse model was developed out of a comprehensive analysis of a large sample of empirical research, whereby 29 empirical phenomena were identified. This model of transcription typing has been shown to be robust following considerable scrutiny by numerous researchers over a long period of time (e.g. in Inhoff & Wang, 1992; Bosman, 1993; and Keith & Ericsson, 2007). An exhaustive review of the body of literature regarding transcription typing is outside of the scope the research presented here, but a subset of the empirical phenomena within the model are relevant to my experiments as they can be used to further justify the methodological and/or analytical approaches that were taken, and to explain some of the results.

The 29 phenomena are grouped into different subsets, *basic phenomena, units of typing, errors* and *skill effects*. Of those items classified as basic phenomena, five are worth discussing in relation to the experiments in this thesis and two of the skill effects are also potentially relevant. These phenomena are discussed in the following sections. The units of typing and errors phenomena from the Salthouse model are presented at a level of granularity beyond which I could learn anything through application to the results from the experiments in this thesis, so are not considered further.

From the basic phenomena, Salthouse identifies that typing is faster than choice reaction time (phenomenon 1), in other words, people are able to type faster than they can respond to choices. This led to the understanding that the processes involved in typing overlap, that is when executing a keystroke the typist is also simultaneously in the input, parsing and translation phases for other keystrokes. This is a potentially interesting phenomenon in relation to the research presented in this thesis because it suggests that although the output from transcription typing activities is serial (i.e. each character appears on screen one after another) the perceptual, cognitive and motoric processes involved in achieving that output are parallelised, at least to some extent. This parallelisation may have implications for any limitations of cognitive processing ability and processes involved which may, perhaps, limit how well the typist can perform while exposed to music, particularly if we assume that exposure to music is considered to be a task that requires cognitive processing. Salthouse identifies that although typing is faster than reaction time, it is considerably slower than reading irrespective of the typists' skill (phenomenon 2). This finding suggests that it is not the input processes that limit typing speed as people can read material much faster than they can type it.

Typing skill has also been shown to be independent of comprehension (phenomenon 3) of the material, meaning there is no relationship between the skill of the typist and the level to which the material that is being typed is understood. Further, even when asked to make efforts to comprehend as well as to accurately transcribe presented text, the level of typing performance is generally maintained, again across typists of varying skill levels. These findings are explained through consideration of the different goals of the user when reading and typing versus just typing, with the former intending the user to combine words into longer passages to allow interpretation of meaning while the latter requires decomposition of the individual words into characters to be pressed on the keyboard.

This phenomenon, that typing skill is independent of comprehension, is particularly relevant to the work in this thesis when considering the appropriateness of transcription typing as a task in relation to the overall aim of this thesis to investigate the effect of music on performance in work-related computing tasks. This phenomenon means that we can assume that the typing performance rates achieved would be similar if I had asked the participants to comprehend the material as well as to transcribe it. So, although transcription typing without comprehension is not likely to be a typical work-related computing task (because, usually someone would be able to just copy and paste the text), some people may transcribe text as a means to understand the material (e.g. when taking notes about a topic), but their performance at that task should not differ considerably due to the additional comprehension requirement. It would be interesting to follow up the work in this thesis with further experiments where the participants were asked to make efforts to understand the presented text as well as to accurately transcribe it, and to explore if comprehension of the materials was also affected by the music stimuli. If different pieces of music led to differing levels of comprehension of the materials, and reductions in typing performance, this phenomenon means that we could attribute those differences to the music stimuli and not the requirement to understand the material. That is, we would not expect an additional requirement to comprehend the material to lead to a reduction in overall typing performance when compared to a non-comprehension condition, so if differences were to be observed, these could be explained through the music rather than the task.

Although the participants in my experiments were not asked to focus on understanding the presented text, some of the participants may have chosen to think about the text they were typing while they were typing it, while others may just have been copying the presented text without trying to understand it. The phenomenon that comprehension is independent of performance means that, even though the participants were not consistent in either comprehending or not comprehending the text, we know that it was not a variation in comprehension levels that influenced the overall performance results. For example, this phenomenon means the typing speeds that were achieved by the slow group were not systematically caused by those typists choosing to think about the materials more carefully than those classified as fast typists. Instead, the performance differences were more likely due to skill level of the participants, rather than due to levels of comprehension of the material.

Another of the basic phenomena identified by Salthouse is that typing rate is independent of word order (phenomenon 4) and again, this phenomenon is independent of typist skill. So typing performance when the presented text is meaningless is nearly the same as when the typist is presented with meaningful text for both skilled and unskilled typists. This phenomenon has been used as evidence that the processes involved in reading text and transcribing text are different because it seems that the syntactic and semantic nature of the presented text does not affect typing performance, as long as the word chunks[31] themselves are meaningful at an individual level. However, as the presented text becomes more random at a character level, typing speed deteriorates (phenomenon 5). These two phenomena together justify the choice of Dutch text for the presented text used in the hard transcription task in Experiment ?, rather than, for example, just using meaningless text that contained English words. To non-Dutch speaking participants transcription typing in Dutch is equivalent to typing random characters. Further, this phenomenon suggests that to maintain a fast typing speed the typists must be able to code the material into appropriately sized chunks, and that when the size of the coded chunk is reduced due to an inability to predict the next letter, or lack of familiarity with the combination of characters, the process becomes inefficient and performance is reduced. Therefore, the Advanced English with Pseudo-words text in Experiment 4 was an appropriate choice for a moderate level of task difficulty because some of the words would need to be chunked into smaller components due to their level of non-familiarity.

One final phenomenon that was classified under the basic phenomena group is that concurrent tasks do not affect typing performance (phenomenon 12), as “with highly skilled typists, a concurrent activity can often be performed with little or no effect on speed or accuracy of typing” (Salthouse, 1986, p. 309). This phenomenon has been identified in a number of empirical studies (e.g. refs) and is explained through the idea that tasks which have been extensively practiced become automatic. However, there is acknowledgement that this particular phenomenon needs further exploration as to the limits and circumstances under which tasks become automatic, with evidence from Shaffer (1975) that performance was reduced when a typist was asked to read aloud a different text that was presented visually while transcribing something else, although this same typist could maintain typing speeds while reciting nursery rhymes. Salthouse (1986) encouraged further work to understand the limitations of this phenomenon in terms of what sorts of concurrent activities may lead skilled typists to a reduction in performance. The experiments presented in this thesis provide some evidence that the fast typists performance was, in particular, reduced with both the presence of certain types of music and with variations to different dimensions of the music, while the effects on the less skilled typists were minimal. So, given that it is appropriate to consider the fast typists as

31 In the transcription typing literatre, the term *chunk* is used to describe the different segements of a word. So, for example, the word *dodecahedron* could be chunked into its five component syllables or into four components as do-dec-a-hedron, while the word *hyperdendritic* would probably be chunked as hyper-den-dritic.

skilled typists (albeit, not necessarily having received explicit training as touch typists) that their performance was reduced with music suggests that hearing music may provide the sort of cognitive interference that interacts with transcription typing in such a way to detract from a skilled typists' performance. These results contrast with the phenomenon that concurrent tasks do not affect typing performance but adds weight to the argument that there are some types of concurrent activities that do interfere with typing performance and perhaps suggests that complex aural stimuli, such as music, does interfere with typing processes to reduce overall performance.

There are two further phenomena of potential interest to this work, classified under "skill effects" namely that tapping rate increases with skill (phenomenon 24) and variability decreases with skill (phenomenon 25). In the analysis of the data from my experiments, wherever possible, the participants were classified as either fast or slow typists based on the speeds that they achieved in the tasks. Typically, across the experiments, as well as having higher typing speeds the fast typists also made fewer errors. So, taking these two phenomena together, it suggests that those typists who were classified as fast typists are likely to be more skilled than those classified as slow typists.

Although not all 29 phenomena have been explored through this section, one key theme to emerge is that although more skilled typists type faster with less variability, the other phenomena (e.g. reading comprehension, that typing rate does not depend on the word order) are independent of typing skill. So, the fact that the results from the experiments in this thesis vary depending on the typists' speed classification, which was used as an indicator of typing skill level, suggests that the observed differences are not due to typists of differing abilities when using the different processes to type, but that there is some interference caused by the music stimuli that causes the effects to be distinctive between the groups.

10.2.2 Working Memory, Unattended Speech and Music Effects

As well as trying to understand how the task of transcription typing uses perceptual, cognitive and motoric processes, researchers have used writing and transcription typing tasks in order to better understand the cognitive system of working memory. Much of the research to develop models and theories relating to working memory have adopted empirical approaches that expose people to unattended speech while performing some kind of memory task. At the simplest level, this body of research exposes people to unattended speech while asking them to complete some other task in a dual task paradigm. The research typically questions how material presented aurally can influence their memory task performance even if they are not consciously attending to the aural material, with a particular emphasis on the role of working memory within the tasks. As some of the studies used to inform theories of working memory involved typing tasks, and others involved trying to disrupt working memory through auditorily presented materials (i.e. by using unattended speech, unattended music or both) it is worth considering how the theories developed out of this research may apply to the results from the experiments in this thesis. Some of the studies underpinning the theories relating to unattended speech have also explored the effects of unattended music, both with and without lyrical content (e.g. in Martin, Wogalter, & Forlango, 1988 and Pool, Koolstra & van der Voot, 2006). Given that the participants in my experiments may not have been consciously attending to the accompanying music, it is worth further exploring these theories to see if the effects of unattended speech and/or unattended music applied when transcription typing in my experiments.

The first key concept that needs to be explained in more depth is that of the cognitive system known as working memory. The Working Memory Model (Baddeley & Hitch, 1974) represents a type of short-term memory, which consists of different systems that are used to process and store different types of information. Of particular relevance to the work in this thesis is the *phonological loop* system, which is the part of working memory that concentrates on written and spoken material, and has two parts the *phonological store* and the *articulatory control process*. Speech is stored in the phonological store for 1 to 2 seconds and although written words can be stored in the phonological store, they must first be converted into speech form (known as articulatory code) by the articulatory control process before being transferred to the phonological store. There are other aspects of the Working Memory Model (e.g. the *Central Executive* and the *visuo-spatial scratch pad*) but their relevance to the work in this thesis is limited, so they are not considered further.

In order to better understand working memory, and specifically to explore whether unattended speech affects the phonological store, a number of experiments were performed to focusing on how unattended or irrelevant speech affected serial recall of visually presented numbers (Salamé & Baddeley, 1982). The results from these experiments demonstrated that unattended speech does have a detrimental effect on immediate memory, with reduced performance on the memory task when accompanied by speech. This unattended speech effect is widely considered a phonological (i.e. relating to the sound of the words) rather than semantic (i.e. relating to the meaning of the words) effect, as other experiments have shown the same reductions in performance at working memory tasks when using nonsense words and words from other languages (e.g. in Baddeley & Salamé, 1986).

The research around the unattended speech effect has predominantly focused on understanding the implications of unattended speech on working memory. So, the first thing to explore in relation to whether or not this work could be used to explain the results from the experiments in this thesis, is the question of whether transcription typing is actually considered a task that involves working memory systems. Given the importance of working memory in the research literature around the unattended speech effect, if transcription typing is not a task that involves working memory processes, then the relevance of the unattended speech effect to the experiments in this thesis is questionable.

Hayes and Chenoweth (2016) explored the role of working memory in transcription typing as a follow up to research by Levy and Marek (1999) in which unattended speech was found to have no effect on transcription typing performance. As a result of their studies, Levy and Marek (1999) came to the legitimate conclusion that, as performance was not affected by unattended speech, the processes involved in transcription typing, specifically that of programming and executing motor movements, do not involve the phonological loop component of working memory. However, Hayes and Chenoweth (2016) followed this work by using an articulatory suppression approach to interfere with working memory, which involved repeating the word “tap” aloud in time with a metronome rather than using unattended speech as the mechanism. This approach was taken because articulatory suppression had previously been shown to have a larger interference on the phonological loop in working memory than that of unattended speech (e.g. by Salamé & Baddeley, 1982). When using the articulatory suppression technique, performance in transcription typing tasks was reduced both in terms of typing rate and numbers of uncorrected errors, demonstrating that the phonological loop in working memory is involved at some point in the transcription typing task. The authors hypothesise that articulatory suppression may be affecting the reading component of the transcription typing task specifically (Hayes & Chenoweth, 2016), rather than the programming and executing motor movements processes. As these experiments provide evidence that working memory is involved in transcription typing there is potential benefit in exploring whether unattended speech effects could account for the results from the experiments in this thesis, even though the research by Levy and Marek (1999) indicated that unattended speech does not affect typing performance.

There are a number of ways to explain the difference in results between the effects of unattended speech on transcription typing found by Levy and Marek (1999) and the effects of articulatory suppression on transcription typing found by Hayes and Chenoweth (2016). Of particular interest and relevance to my research, is the idea proposed by Hayes and Chenoweth (2016) that perhaps the skill level of the participants in the experiment involving unattended speech was higher than that of the participants in the experiments using articulatory suppression, and hence why performance was not affected by unattended speech. The justification provided for this explanation was that people that are more skilled at transcription typing would “require less working memory for typing than our participants” (Hayes & Chenoweth, 2016, p. 144). However, after comparing the typing speeds from the participants across the studies and finding that their participants were achieving much higher speeds and lower error rates that the participants exposed to unattended speech, they conclude that it is not skill level that is a factor, instead it is likely to be that the scale of the effect of unattended speech on working memory is less than that of the articulatory suppression task. However, their assumption that skilled typists would be using less working memory in the transcription typing task than less experienced typists is interesting in relation to my experiments because there was significant evidence that type and scale of any effects of music on performance clearly depended on the speed classification of the participants. Typically (although, not always) the fast typists’ performance was affected by exposure to different pieces of music but there was no corresponding evidence that the slow typists’ performance was similarly influenced. If the assumption that skilled typists are using less working memory is valid, as proposed by Hayes and Chenoweth (2016), we would expect that the performance of skilled typists would be less affected by unattended speech (or other methods involving interfering with working memory) than the less skilled typists who require larger working memory resources. However, this argument contrasts with the results from my experiments as we would expect to see the slow typists’ performance affected more than the fast typists, which was not the case. Therefore, either the initial assumption is invalid, or exposure to music does not correspond with the expected unattended speech effects due to interference with working memory.

One further study that is of potential interest and relevance to my experiments investigated the effects of both unattended speech and unattended music on reading comprehension (Martin, Wogalter, & Forlango, 1988). Their experiments found that reading comprehension was impaired significantly by auditorily presented speech, but not by instrumental music, nor by random tones. They compared this result to that of Henderson, Crews and Barlow (1945) who had previously found that music with lyrics had a detrimental effect on reading comprehension, and consider that it was therefore likely to be the lyrical, verbal component of the music that specifically reduced reading comprehension. Martin et al. (1988) followed this work with an experiment that compared performance in reading comprehension tasks with music that contained lyrics versus instrumental music and identified that performance was affected in the lyrical condition only, whether this verbal content was spoken or sung.

If theories relating to unattended speech and unattended music described in the previous paragraphs are applied to the results from my experiments, we would expect to see reduced performance in conditions where the music contains vocals. The effect of vocals was explored in three experiments, Experiment 1, 3 and 4. Experiments 1 and 3 provided evidence that vocals in the music reduced transcription typing performance, particularly for the fast typists. Therefore, it is possible that this reduction in performance can be attributed to unattended speech effects interfering with the verbal component of working memory that may be used in transcription typing tasks. However, the results from Experiment 5 showed an improvement in transcription typing performance when the music contained vocals, which cannot be explained through unattended speech effects.

Although some of the results from the experiments in this thesis might be explained through unattended speech affects on working memory processes, given that the results regarding the presence of vocals in the music do not consistently align with the theory, care should be taken not to overstate the relevance of this theoretical framework.

10.2.3 Multiple Resource Model

Another theoretical framework that is worth exploring in relation to my research, is that of Wickens' *Multiple Resource Theory* (MRT) which provides a useful framework for understanding performance in both dual and multi-tasks (Gopher, Brickner, & Navon, 2005). The MRT is based around two key concepts that must first be defined, that of *multiple* and *resources*. In this context, a resource is defined as "something that is both limited and allocatable (i.e. can be distributed between tasks" Wickens, 2002, p. 160) while the term multiple indicates "parallel, separate or relatively independent processing" (Wickens, 2002, p. 160). The MRT connects these two concepts and can be used to understand the limitations of human performance when multitasking.

The MRT is a multidimensional model, comprising of four dimensions which are defined as follows:

- **Stages:** Perceptual and cognitive activities (e.g. speech recognition) appear to use the same resources. These resources are different to those used when selecting and executing responses (e.g. speech production).
- **Perceptual modalities:** Dividing attention across perceptual modalities (e.g. between auditory and visual channels), known as cross-modal, is better than dividing attention within one perceptual modality (e.g. between two auditory or two visual channels), known as intra-modal.
- **Visual channels:** There are two approaches to processing information visually that use different resources, defined as using focal and ambient vision.
- **Processing codes:** There is a distinction between analogue/spatial processes and categorical/symbolic processes (i.e. verbal processes).

(Wickens, 2002)

Wickens' MRT applies to dual or multi-task processes, therefore in order to use it to explain the results from my thesis, we must assume that my experiments are dual or multi-task. The question of whether or not transcription typing while exposed to music can be considered a dual task is therefore extremely important, and is addressed overleaf.

If, for a moment, we assume that my experiments are indeed dual task experiments then given the definitions above, the dimensions of the MRT that are of particular relevance are the stages, perceptual modalities and processing codes dimensions. The stages dimension is relevant because the transcription typing task involves both perception (i.e. reading the text) and executing a response (i.e. pressing a key) while the hearing music task involves perception alone. The perceptual modalities dimension is relevant because the transcription typing task involves processing visually presented material, while the hearing music task involves auditorily presented material (i.e. the tasks are cross-modal). Finally, the processing codes dimension is relevant because the transcription typing task involves both a manual response that is spatial in nature because the typist has to know the spatial position of their fingers in order to press the keys, and verbal because they are typing words, while the hearing music task sometimes involved verbal components when vocals were included in the music. The visual channels dimension is not relevant because the transcription typing task only involves focal vision, and does not require any ambient vision, and the exposure to music task does not require any form of visual channel.

Focusing initially on the *stages* dimension, the model states that performance will be reduced in tasks that require similar resources because the information must be processed in sequence, but when the tasks require different types of resources information can be processed simultaneously. At the simplest level of abstraction, the model proposes that if one task demands more resources there will be greater interference with other concurrent tasks (assuming the types of resources required are similar) leading to reduced task performance[32]. When a task requires the full allocation of resources to be assigned in order to achieve high performance this state is referred to as being *resource limited* (Wickens, 2002) and indicates that there are no additional resources left available for allocation to other tasks, which will affect performance of those other tasks.

If we assume that exposure to music can be defined as a task, because the individual is either passively hearing or actively listening to the music, then Wicken's MRT model may provide an explanation for some of the results from my experiments. However, this initial assumption is an important aspect to question before trying to use the MRT to explain the more complex results from the experiments in this thesis. It is important to begin by carefully considering whether or not the experiments in this thesis do involve requiring the participants to work concurrently on multiple tasks, by questioning if these experiments do use a dual task paradigm. It is clear that the transcription typing aspect of my experiments is a task, and one for which measures of performance can easily be taken. But, can exposure to music also be defined as a task? For the MRT to apply, we must assume that hearing music is a task in itself, as otherwise a dual task paradigm would not

32 A theory which contradicts phenomon 12 in the transcription typing model stating that performance at concurrent activities is not affected when transcription typing (Salthouse, 1986).

apply. So, it is useful to explore logically whether my experiments have produced any evidence that this is an appropriate assumption.

If working under the assumption that hearing music is a task, and therefore that exposure to music while transcription typing can be considered multitasking, we would expect the limited amount of resources must be shared between the two concurrent tasks. One way to verify that exposure to music can be considered a task that requires resources, is to compare transcription typing performance both with and without exposure to music. If hearing music is a task, we would expect to see a reduction in transcription typing performance with exposure to music when compared to transcription typing without music as this latter condition becomes a single rather than a multi faceted task. In a single faceted task, a full allocation of resources could be given to the transcription typing task which should lead to the highest levels of performance as there are no requirements for resource sharing.

Interestingly, the results summarized in Table 10.1 do not consistently identify reductions in transcription typing performance when exposed to music when compared to without music conditions. There are two possible explanations for this inconsistency in results. One explanation is that hearing music is not a task that uses resources, so transcription typing while exposed to music is simply a single task activity, and hence no differences in performance were identified between the with and without music conditions. Alternatively, the results may indicate that the transcription typing task requires a large allocation of resources to achieve high performance, so in a multiple resource model it is would be performance at the hearing music task that is reduced, rather than the performance in the transcription typing tasks. My experiments did not measure performance when hearing music, and it is difficult to define the concept of performance in a hearing music task. However, we could consider performance when hearing music as the amount of attention given to the music, i.e. high performance means to actively listen to the accompanying music, while lower performance means passively hearing the music. Given that, in both Experiments 2 and 5, there was some evidence that exposure to music reduced transcription typing performance, it seems likely that the second explanation applies whereby hearing music *is* a task, but it is the performance in this hearing task that was reduced when under a multitask condition involving exposure to music while transcription typing.

Of further interest, is that in Experiment 2 although CPM was reduced when exposed to music when compared to the without music condition, this performance measure was not reduced in all of the music conditions, and neither was it reduced for both groups of typists. The results from this experiment showed that it was only the fast typists' CPM that was reduced with exposure to music with no evidence that the slow typists' CPM was affected. Further, there was no evidence that even the fast typists' CPM performance was reduced when exposed to the ambient music. This suggests that the fast typists were using more resources to hear the rock and classical music than the ambient music, which led to a reduction in CPM performance in these conditions. And in fact, this may also indicate that exposure to ambient music is either not a task that requires resources or is a task where it was easier to allocate resources away from. In contrast, it seems that exposure to the classical or rock music, required resources which led to interference and a reduction in CPM performance for the fast typists.

Although there was no evidence that the slow typists' performance, in terms of CPM, was reduced, the experiment did provide some weak evidence that performance in terms of typing accuracy was reduced with classical and rock music for both the slow and fast typists when grouped together. Again, these results suggest that hearing the classical and rock music required resources that were then unavailable for the transcription typing task.

The results from Experiment 2 also showed that there was no evidence of a reduction in fast typists' CPM due to exposure to music when typing in Dutch, which would be considered a resource intensive task even for the more highly skilled typists. Therefore, it is likely that the Dutch typing task used all available resources, i.e. this was a resource limited situation, and so performance at the hearing music task would be reduced when typing in Dutch. Although I did not collect any performance data regarding the hearing music task, the fact that the task difficulty ratings were not significantly affected by exposure to music in the Dutch music condition, but were affected in the English tasks, helps to support this explanation. In the more resource intensive task the music did not affect the perception of difficulty of the task, but the comparatively easier task was considered harder with different pieces of music, suggesting that resources had to be allocated to the task of hearing music when typing in English, but not when typing in Dutch.

The idea of resource limitations posed through the MRT could also be used to account for the differences in transcription typing performance between the slow and fast typist groups that were identified fairly consistently throughout my experiments. Typically, the fast typists' performance was affected by exposure to music, and their performance was varied due to differences in dimensions of the music stimuli, while the slow typists' performance was not affected in the same ways. If the MRT is applied, and particularly the idea that tasks that use more resources, affect the performance of concurrent tasks, we could suggest that because transcription typing was easier for the fast typists it did not require much resources to achieve strong performance, and therefore the task of hearing music used more resources thus reducing performance in the transcription typing task. Essentially, for the slow typists to transcription type at a high performance level this may be a resource limited task, so it is the task of hearing music that is largely impacted, meaning the music does not have an effect on the transcription typing performancea. Whereas, as the fast typists are not resource limited when transcription typing, resources are shared between the transcription typing and hearing music tasks, meaning the fast typists' performance is reduced in the transcription typing task.

The *perceptual modalities* dimension of the MRT focuses on how performance is affected by tasks that use visual, audio, tactile and olfactory perceptual modalities, with higher performance when tasks are cross-modal (i.e. one task using visual and one using audio modality) rather than intra-modal (i.e. both tasks using visual modality). In my experiments, the transcription typing task using a visual modality while the hearing music task uses the auditory modality. Thus, we might expect that exposure to music would not interfere with the transcription typing task, due to using different perceptual modalities. However, there is much evidence from the experiments that transcription typing performance was affected by exposure to music. So, it does not seem that the perceptual modalities dimension of the MRT can be used to explain the results of the experiments in this thesis.

The final dimension of the MRT that may apply to my experiments, is the *processing codes* dimension. This dimension refers to the separation of resources between spatial and verbal processes. The transcription typing task is, arguably, both spatial and verbal in that it involves perceiving words and carefully positioning fingers over keys in order to execute the relevant keypress. The hearing music task is verbal when the music contains vocals. As the speakers were positioned in front of the participant, rather than, for example, using surround sound, there is no spatial component of the hearing music task. So, the processing codes dimension may apply to the experiments where the music contained vocals (Experiments 1, 3 and 4), where we would expect performance to be reduced with exposure to music that contained vocals. While performance was reduced for the fast typists when the music contained vocals, there was no evidence of performance reductions in terms of speed for the slow typists. And, in Experiment 4, performance was improved when the music contained vocals when played at a low volume. Therefore, it does not seem that the disruption to performance that would be expected between two verbal tasks is consistently evident from my experiments. One explanation for this inconsistency is that perhaps any effects of disruption due to the processing codes dimension are mitigated by the verbal material being perceived using different perceptual modalities when the music is played at a lower volume.

The Wickens' Multiple Resource Theory has provided some useful insights into how cognitive disruption may be affecting the participants' transcription typing performance in my experiments. However, the use of this theory is dependent on the assumption that both exposure to music and transcription typing are considered tasks. The basis of this assumption has been explored through this section by attempting to apply the model, and in some ways seems to be reasonable, but without corresponding evidence that performance at the exposure to music task was affected, and a clear definition of what that task involves, the appropriateness of the MRT in explaining the results is limited.

10.2.4 Relationship Between Arousal and Performance

Another conceptual framework that is worth exploring in relation to the work in this thesis involves the relationship between the physiological and psychological state of arousal and task performance, as described by the Yerkes-Dodson law (Yerkes & Dodson, 1908). The law states that the relationship between arousal and performance has an inverted-U shape, with both low and high arousal leading to poor performance levels, and an interim range of levels between at which an optimum state of arousal is reached resulting in the high levels of task performance. This relationship between arousal and performance has been explored through a wide variety of studies in numerous contexts, (e.g. Reilly &

Smith, 1986; Teigen, 2005; Wiener, Curry, & Faustina, 1984), and has been shown to be a robust phenomenon.

Various dimensions of music have been shown to affect arousal in a number of studies. For example, the effect of music tempo and intensity (i.e. loudness) on athlete grip strength was investigated while measuring perceived arousal (Karageorghis, Cheek, Simpson, & Bigliassi, 2017) using the Affect Grid (Russell, Weiss, & Mendelsohn, 1989). In this study, high intensity music led to higher arousal, though the tempo of the music did not have an effect on arousal ratings. However, other research has shown a clear relationship between the tempo of the music and reported arousal ratings (e.g. in Khalfar, Roy, Rainville, Dalla Bella & Peretz, 2008), with higher tempo music leading to higher arousal ratings. Arousal has also been considered in relation to the type of music, where music with “fearful and happy melodies were rated more stimulating than those peaceful or sad melodies” (Khalfa, Isabelle, Jean-Pierre & Manon, 2002, p. 146).

If the Yerkes-Dodson law regarding the relationship between arousal and performance is used to explain the results from the experiments in this thesis, we would expect to see improvements in typing performance when accompanied by music with characteristics that increase arousal, unless increases to arousal levels are too extreme in which case typing performance would be reduced. Given that loudness of music has been shown to affect arousal, both perceived and physiological, the effect of volume on transcription typing performance and experience will be considered first. Volume was manipulated in two of the experiments in this thesis, Experiment 3 and Experiment 4.

The results from Experiment 3 showed that the fast typists’ CPM performance was significantly reduced in the high volume condition, but there was no evidence of an effect of volume for the slow typists. As louder music has been shown to increase arousal, in this case where CPM performance is reduced in the condition that is expected to induce high arousal, it may be that the volume of the music increased the fast typists’ arousal levels beyond that of optimum performance and as a result reduced their CPM performance instead. In Experiment 4, which also investigated the effect of volume on performance, there was some evidence that the low volume music improved performance when it contained vocals for all typists. The discussion of this somewhat surprising result concluded that the music with vocals was louder than the music without vocals even though it was played at the same volume, and that this difference in loudness may account for the improvement to typing speed. If Yerkes-Dodson law is considered with regard to this result, it may indicate that the additional loudness in the music with vocals was sufficient to increase arousal in such a way that performance was improved but not so much that the arousal levels were too high leading to reduced performance. However, arousal itself was not measured during these experiments, so to conclude that the results from these two experiments can be explained using the Yerkes-Dodson law alone would be somewhat premature. It is possible that the high volume condition led to too much arousal reducing performance, and that with the low volume music the vocals caused increased arousal over the instrumental condition, but as this was not the initial premise of the experiment, or theoretical underpinning of the research, we must be careful not to cherry-pick and retrofit one explanation to the results, when there is insufficient evidence and alternative explanations that should also be explored.

One approach that would allow us to more make this claim more definitive, would be to conduct a further experiment that measures perceived and/or physiological arousal when exposed to the music with and without vocals played at the low volume level, but without any accompanying transcription typing task. If there was a significant difference in arousal due to vocals in the music, then this could be used as evidence to support the theory that in the low volume condition, vocals in the music increased arousal and resulting in an improvement to typing performance in CPM because the arousal levels had not been increased to the point that the effects were detrimental. We could also measure perceived and/or physiological arousal when exposed to the high volume music as well, where we would expect to see that the high volume music leads to even higher arousal levels than the low volume with vocals music. This result could then be used to corroborate the theory that the high volume music induces too much arousal in the participants, leading to poorer transcription typing performance.

Another dimension of music that has been shown to affect arousal is tempo, with higher tempo music leading to higher arousal levels. The effect of tempo was explored in Experiment 5, but there was no evidence from the experiment that tempo affected typing speed, but there was some evidence of an interaction between tempo and time signature affecting typing accuray. If the tempo of the music was manipulating participant arousal in such a way that performance was increased (assuming, the optimum arousal levels had not been reached) we would expect to see performance improvements with faster tempo music, up to the point of optimum performance where it would reduce. But, there is no evidence of this relationship from the data collected in Experiment 5. However, this inconsistency may again be an issue of using the term tempo as the manipulation when really it is the relationship between tempo and time signature (i.e. the frequency of the emphasised beats) that actually affects arousal. It would be interesting to follow up this work with a study exploring specifically whether it is the notion of tempo alone, or whether it is the frequency of the emphasised beats in the music, that affect arousal.

The Yerkes-Dodson law has clear potential for providing a justification and explanation of some of the more complex results from the experiments in this thesis, in particular why performance was improved when participants were exposed to low volume music which contained vocals. However, further work is needed to verify that the pieces of music do affect arousal before such an explanation can be appropriately adopted.

10.2.5 Entrainment

Entrainment theory describes the interactions between two or more different, independent processes that have rhythmic components which synchronize with each other (Clayton, Sager, & Udo, 2005). As transcription typing involves repeatedly pressing keys this can be considered a tapping style of activity which may therefore have rhythmic components. Most (although not all) of the music stimuli used within the experiments in this thesis also contained clearly perceptible, identifiable beats and rhythmic content. There are numerous studies that show people entrain to the beat and rhythm of music across a variety of different contexts. For example, people have been shown to entrain to music when driving (Brodsky, 2001) and when running (Van Dyck et al., 2015). As both the experiment task and stimuli have potentially rhythmic components, entrainment is another conceptual framework that is interesting to consider in relation to the results from my five experiments. It is possible that some, or all, of the participants were entraining to the beat of the music stimuli when transcription typing which may have affected their performance and experience in a variety of ways.

Before presenting an interpretation of the results from the experiments under the entrainment framework, it should be noted that the data collected within the experiments presented in this thesis cannot be used to definitively state whether or not the participants are entraining to the music as they did not include an analysis of individual keypress data. In these experiments, the typing speed performance measure was characters per minute

(CPM) which was calculated by identifying how many characters are in the final, transcribed text and dividing by the duration of the transcription typing task. A participant's CPM measure will always be considerably lower than their rate of pressing individual keys, i.e. their keypresses per minute (KPM) rate, as they frequently made corrections to their output text during the task, by using the backspace and arrow keys to move the cursor in order to change erroneous input characters.

As CPM data was used rather than KPM data, these experiments will not be able to provide robust evidence of entrainment as CPM does not provide a pure measure of the rate of pressing keys. Initially, I did attempt to capture KPM data as well as CPM data, but I was unable to consistently record an accurate count of individual keypresses due to software issues and differences between how the keypress data was being captured in the laboratory in comparison to the approach used in the classroom environment. Further, when keypress data was captured, the software used in the laboratory setting had not been designed to record keypresses at a sufficient level of granularity and robustness to provide evidence of whether the participants were entraining to the music or not. For example, when using screen capture software in the laboratory setting to record the individual keypresses, if the participant was to press and hold a key, the software would capture a single held keypress as multiple keypresses at a rate defined by the software, rather than recorded as a single keypress that had been held. In contrast, in the classroom environment this software was unavailable, so individual keypresses were recorded using the bespoke website I created to host the task, meaning that if the participant pressed and held, for example, an arrow key, it was captured as a single keypress. I did not identify this inconsistency in approach to capturing the keypress data until the screen capture software failed to record keypresses due to an incompatible update to the software during data capture which took place during Experiment 2. But, the different recording methods renders the keypress data inconsistent between experiments, and even within Experiment 2 itself where a lot of keypress data was lost.

Even though the analysis of the CPM data cannot be used to provide a definitive answer as to whether the participants are entraining to the music stimuli or not, as CPM is a measure of typing speed, it is still interesting to consider whether the differences in CPM caused by the music variations may be attributable to entrainment theory. The final experiment presented in this thesis was the only one that directly manipulated aspects of rhythm as independent variables (IVs). The other experiments in the thesis included exposure to music with different rhythmic characteristics, but rhythm itself was not directly manipulated. So, to explore whether or not entrainment theory applies in this context, the results from Experiment 5 should be analysed within this framework as a useful starting point.

In Experiment 5, the effects of tempo and time signature were explored, together with the associated concept of frequency of the emphasized beats. If the participants were entraining to the tempo of the music, or to the frequency of the emphasized beats, we would expect to see the CPM dependent variable (DV) affected by the experimental manipulation as this experiment explicitly aimed to investigate the effect of tempo and time signature on transcription typing performance and experience. At its most simple interpretation of the theory, if the participants were entraining to the beat, we would expect to see an increase in CPM as tempo increased due to an increase rate of keypressing. However, the simplest interpretation and expectations are possibly too basic for this complex phenomenon.

To explore how entrainment theory might apply with more complex predictions of the expected outputs and differences due to tempo, it is useful to apply some numbers to the rates of keypressing per minute and keypresses per beat. For example, if someone was exposed to music with a tempo of 140 bpm, and they were entraining to the music by making 5 keypresses per beat, we would expect their KPM to be 700 KPM whereas for music at a tempo of 210 bpm a rate of 5 keypresses per beat would lead to 1050 KPM. It is possible that 700 KPM is achievable by skilled typists, but 1050 KPM is too fast. If that is the case instead of entraining at 5 keypresses per beat with the 210 bpm music, the rate of keypresses may be reduced to 4 keypresses per beat, which would lead to a rate of 840 KPM. But, if 840 KPM was still too high to entrain to, they may reduce to 3 keypresses per beat (i.e. 630 KPM) which is less than the 700 KPM achieved with the slower tempo music. So, if the participants were entraining to the beat, we may not always see a relationship in which an increase in tempo directly leads to an increase in KPM. Depending on the music used one could also imagine a situation where the rate of keypresses per beat that is achievable at the different tempos differs to the point that there are no discernible differences in the overall KPM between the various conditions.

In Experiment 5, there were no significant main effects or interactions based on CPM due to tempo, time signature or frequency of emphasized beats. These results suggest that either the CPM measure was not affected by these rhythmic manipulations and therefore that the participants were either not entraining to the rhythmic components of the music, or if they were entraining to the music, it was at such a rate of keypresses per beat that a difference could not be identified. Further, in this experiment I was unable to apply a reliable typing speed classification to the participants' data, so it may be that the skilled typists were affected differently to the less experienced typists, thus leading to overall null effects. If a more reliable method of categorising the participants were identified, then it is possible that entrainment effects would be identified for some, if not all, of the typists.

Although there was no evidence of an impact of these rhythmic components on CPM, with regard to the error rate measure there was a significant main effect of emphasized beats and a significant interaction between time signature and tempo. The pairwise comparisons revealed that error rate performance when accompanied by the 3/4 time signature music at 210 bpm was significantly lower than four of the other five music conditions at a non-adjusted level. That is, with the exception of the mid tempo 4/4 time signature music, error rates were higher when accompanied by the piece of music with the highest frequency of emphasized beats at 70.00 emphasized beats per minute. Given this result, it seems feasible that the participants may have been entraining to the frequency of the emphasized beats within the music, leading them to type faster than they were able to accurately, resulting in higher overall error rates in this condition. However, this explanation does not account for the non-significant difference in error rates between the mid tempo 4/4 condition (which was 180 bpm music in 4/4 time at 45.00 emphasized beats per minute) and the high tempo 3/4 condition. Nevertheless, one possible explanation for the increase to error rate may be that of entrainment but there is insufficient evidence from this experiment to be able to definitively make that claim. In order to do so further data, such as KPM, keypresses per beat and measures relating to the rate of correcting errors, would need to be collected as well. Ideally, software and hardware would be used to capture the exact timings of individual keypresses and relate those to the beats in the music. If the participants were entraining to the music, and typing beyond their natural ability leading to higher error rates in the output text, I would expect to see an increase in keypress rate with the higher tempo music as well as high numbers of corrected errors in the input text.

Although Experiment 5 was the only one in this thesis that explicitly manipulated rhythmic aspects of the music, we may be able to find some evidence of entrainment through interpretation of the results from other experiments where different pieces of music were used as stimuli. For example, if the entrainment theory applies, then we might reasonable expect to see performance differences due to the music style IV used in Experiment 1, which compared alt rock music to pop rock music, because the rhythmic characteristics of these pieces of music were quite different. The alt rock music had a tempo of 210 beats per minute (bpm) in 6/8 time signature while the pop rock music was at 140 bpm in 4/4 time signature. However, there was no main effect of music style on the performance measures in Experiment 1, nor was there any interaction between music style and speed classification, which we would expect to see if it was purely the tempo that participants were entraining to.

That said, there was a significant 3-way interaction between music style, vocals and speed classification on CPM with a significant 2-way interaction between music style and vocals for fast typists, but not for slow. For the fast typists, the simple main effect of music style on CPM was significant when the music did not contain vocals but was not significant for music that did contain vocals. That the fast typists' CPM was higher in the pop rock music than alt rock music when the music did not contain vocals indicates that the style of the music affected CPM for these typists, which could be attributed to the differences in tempo and time signature between the two styles of rock music. However, again there is insufficient evidence from this result alone to clearly make that claim as the alt rock and pop rock music differed in other ways beyond just the rhythmic characteristics[33]. Further, if tempo entrainment was an explanation for this difference, at the simplest interpretation we would expect CPM to be faster with the alt rock music than pop rock music because the tempo is higher, but CPM was reduced when accompanied by alt rock music. However, I have shown earlier that the simplest interpretation of the relationship between tempo and typing speed may be too crude, so this reversal of the expected direction does not necessarily mean entrainment is not a factor in an explanation of this particular result.

We can also consider how entrainment theory would explain the results from Experiment 2, which compared the effects of four different music conditions (without music, ambient, classical and rock) on transcription typing performance and experience. If entrainment theory applied we would expect there to be a difference in typing speed due to the varying rhythmic characteristics between the different genres of music, i.e. between the ambient, classical and rock music conditions. The ambient music can be described as arrhythmic

33 An alternative explanation of this particular result is discussed in relation to the relationship between arousal and performance at the end of Section 0.

with no distinct beat or tempo, while the classical music had a tempo of 140 bpm in a 4/4 time signature (which, coincidentally is the same tempo and time signature combination as the pop rock music used in Experiment 1) and the rock music used was the alt rock without vocals music used in Experiment 1 with a tempo of 210 bpm in a 6/8 time signature.

Given that entrainment theory is predominantly about how people synchronize to the beat and tempo of music, it is first worth considering whether there were any performance differences between the classical and rock music conditions as these pieces of music both had a clear pulse and rhythmic components. If entrainment theory does apply, and was related to tempo specifically, we would expect there to be differences in CPM between the classical and rock music conditions, but this is not the case. The results from Experiment 2, do not provide any evidence of differences to either CPM or error rate between the classical and rock music conditions. Therefore, it does not seem that the participants were entraining purely to the tempo of the rock or classical music, as there were no noticeable differences in typing speed. This result is similar to that in Experiment 1, where there were no clear speed differentials between the alt rock and pop rock music despite the differences in tempo.

However, it is worth noting that the tempo and time signature of the classical and pop rock music conditions were the same, both at 140 bpm in 4/4 time, leading to 35 emphasized beats per minute. And, as shown in Figure 8.4, Chapter 8, that this rate of emphasized beats per minute is the same for the alt rock music even though the tempo and time signature are quite different at 210 bpm in a 6/8 time signature. Therefore, it is possible that instead of entraining to the tempo of the music, participants were entraining to the frequency of the emphasized beats instead. If that is the case, we would not expect to see any speed differences between the alt rock and pop rock styles, or between the classical and rock music, because by chance, the 3 different pieces of music used for these conditions all contained 35 emphasized beats per minute.

Of potential further interest in relation to whether entrainment theory can be used to explain the performance measures in these experiments, is the fact that in Experiment 2, I included comparisons with an arrhythmic piece of ambient music which had no discernible tempo to pieces of classical and rock music and found that typing performance when exposed to ambient music was consistently better when accompanied by ambient music over both the classical and rock pieces. These results mean that CPM was reduced when the participants were exposed to music that had a more rhythmic component which suggests that with the arrhythmic ambient music the participants were able to type at their natural speed, but with the more rhythmic music their speed of typing was reduced, perhaps because they were typing in time with the music. However, again at this stage of the research, this argument is based on somewhat speculative interpretation rather than specific evidence for entrainment, so should not be taken to be a conclusive explanation of this result.

The only clear way to identify if typists entrain to music would be to conduct new experiments focused on exploring typing performance in relation to this theoretical framework. These new experiments would need to accurately measure KPM instead of comparing the overall typing speeds that were achieved by analysing the output text using CPM, when participants were exposed to music at different tempos and with different time signatures and rhythmic characteristics (e.g. some with prominent syncopation and others without). Measures of keypresses per beat would also need to be identified. Ideally, as well as analysing raw KPM data for each condition, this new experiment would benefit from capturing the difference in KPM between non-music conditions and with music conditions, by recording their KPM without music and analysing the relative KPM differences between their natural rate of typing (i.e. without music) and with music at different tempos and time signatures. That is, ideally both absolute measures of KPM and relative KPM measures would be recorded. This relative KPM is an important measure because, if the participants are entraining to the music, some pieces of music may increase speed of typing but others may reduce it. Having an indication of how the rate of keypresssing changes in relation to the participant's natural speed of typing will allow us to account for differences in participant typing ability more easily and to consider if the participant's natural keypress speed is slower than the tempo of the music and it is the music that is increasing their typing speed, or if their natural keypress speed is higher than the tempo of the music and the music is decreasing their speed. For a gold standard experiment, a combination of software and hardware would be developed that would allow us to capture the exact timings of each keypress in relation to the beats in the music, as well.

An experiment could be designed that exposes participants to music at different tempos and time signatures and considers the effect of these on their rates of key pressing, together with the associated concept of frequency of emphasized beats. For example, if someone was exposed to music at 140 bpm, and they made four keypresses per beat we would expect their KPM to be 560 KPM. So, if their natural typing speed was initially higher than 560 KPM we would either expect them to reduce their tapping speed to 560 KPM or to increase it to around 700 KPM if they were entraining with the music. In order to identify this, a robust measure of keypresses per minute needs to be taken, as well as using music that has been carefully chosen so as not to include any changes in tempo. Further, as the participants in my experiments often paused at various points in the transcription tasks, any periods of non-keypressing activity would need to be removed from the data prior to analysis. This would be an interesting experiment to perform, particularly across participants of differing typing skill levels.

This section has considered whether it is possible that the participants in my experiments were entraining with the music by typing in time to the music they were exposed to. But, it is not possible to undercover clear evidence of such entrainment using the data that was captured through my experiments. This section has shown that the experiments reported in this thesis cannot be used to definitively identify whether or not the participants are entraining to the music they were exposed to, although there are some suggestions that this might be the case.

10.3 Thesis Contributions

The research in thesis has moved the field forward in terms of understanding the effect of different dimensions of music on typing performance and experience. The effects are complicated, somewhat inconclusive and depend on numerous other factors, namely the speed classification of the typist and the difficulty or language of the text. But, some take home messages have emerged from the work, which are summarised briefly:

- vocals in music affects transcription typing performance and experience, but the exact effects are dependent on the typist's speed classification, the difficulty or language of the task and the volume of music,
- the volume of music interacts with vocals to affect typing performance and experience in different ways, depending on the typist's speed classification and the difficulty or language of the task,
- transcription typing in Dutch is different to transcription typing in English, and the effect of music on performance and experience in Dutch is therefore, different,

- transcription typing English with inserted pseudowords is harder than just in English, and seems to be a successful way to manipulate task difficulty while retaining the wordiness of the text,
- the understanding that some dimensions of music should not be isolated and treated using an empirical approach to experimental investigation,
- a set of researcher guidelines for conducting experiments using a classroom based methodology that has pedagogic value for the students, whilst resulting in empirically valid data for the researcher.

The key outcome from this thesis is that although a strict and controlled scientific method has been followed, with experiments designed that, on a conceptual level should have led to a meaningful outcome, the complexity of the stimulus (music), the complexity of the tasks (transcription typing and working at a computer) and the complexity of people mean that this approach to answering the research question was inconclusive. However, although on paper this is seemingly a disappointing result, it should be noted that this thesis presents very early-stage work in attempt to this investigate the effect of music in mundane work related computing tasks. And, although the picture at the end of the thesis is not entirely clear, and in many ways this research has posed more questions than it answers… but, in scientific communities, this is not always a bad thing. In fact, as an early foray into a complex area with clear potential for significant impact, posing more questions is definitely not a bad thing. Attempts have been made to apply different theories to explain the results of the experiments in this thesis, but again, these explanations are not always consistent and further research would be needed to verify and validate their appropriateness as explanations of these results.

10.4 Future Research Proposal

In the previous sections, and within the discussion presented in Chapter 9, I have hinted at a number of possible future research ideas. These proposals are formalised in this section, in terms of both the next steps to take this work forwards and some more long term ideas.

Assuming research in this area should continue to focus on transcription typing as a task, which is a fairly large assumption, a different approach to classifying participants into speed groups should be explored. The method I used for this classification, by inspecting the histograms for suitable crossing points, and then checking suitability in scatterplots worked in four of the five experiments in this thesis. In Experiments 1 to 4, the crossing points between distributions emerged fairly clearly, and the scatterplots reinforced the thresholds were suitable. However, in Experiment 5 a suitable classification structure could not be identified, and I tried setting a few different thresholds before deciding I had crossed into manipulating the data to suit my needs by adding experimenter bias. This post hoc approach based on achieved speeds is clearly open to the experimenter selecting a threshold that suits their narrative and results, so a less easily influenced approach would be beneficial. This approach could simply involve setting thresholds before the data is collected, perhaps by asking participants to complete short typing tests before the session and put all participants achieving below 40 CPM as slow typists, between 40 and 60 CPM as mid speed typists, between 60 CPM and 80 CPM as fast typists, those above 80 CPM as super-fast typists. This approach is less open to experimenter bias, as the classification is done before the data has even been collected. But, the values chosen for the thresholds are completely arbitrary. Why four groups? Why not use three? Or just two? Another approach would be to use eye tracking software and classify the participants after the experiment based on how much time they spent looking at the screen while typing without any music playing. But again, the thresholds for classification would need to be set arbitrarily which is a little unsatisfying and requires access to eye tracking equipment which may not be easily available. Neither of these proposed approaches seems ideal, and potential experimenter bias is a problem with the experiments performed using this post hoc classification, so perhaps this is a key indication that actually transcription typing tasks may not be suitable for this work, especially if speed is considered an important aspect of typing performance.

In terms of the dimensions of music affecting performance, further work needs to be done to explore the unexpected result from Experiment 4 with the vocals improving performance in low volume music. The ideal approach to this would be to perform a large scale study with multiple different volume levels with and without vocals in the music. But, to perform such a study would probably require an online methodology as even using my classroom based methodology over two years only 113 participants were involved. And, as previously discussed volume cannot be controlled using an online methodology. So, before attempting something of this nature, the next step is actually to explore possible approaches to mitigate for a researcher’s inability to control playback volume in an online experiment.

I have a few ideas for how this could be achieved, but careful thought and trials need to be conducted to explore which approach is most successful. One approach would be to play a tone to the participants before the experiment starts which starts off very quietly and gradually increases in volume and ask the participants to indicate the moment they first hear the sound. This would give the researcher an indication of the system volume level required to make the sound audible using the participants equipment. There are problems to this approach though, as some participants may have their equipment set to playback so loudly that all sounds are audible. An alternative method, would be to play a loud tone to the participants, and ask them to reduce the volume of their equipment until the sound becomes inaudible. Then, the experimenter knows that any increase to the system playback volume would make the sound audible. But, there are still possibly problems with scaling factors applied by the equipment which would need to be explored.

Another approach would be to accept that a researcher is not going to be able to match the loudness of music in an online experiment, and instead to attempt to control the perception of loudness. For example, the volume at which I consider a piece of music to be loud is likely to be different to the volume someone else perceives as loud. In the setup phase for an online experiment, the system could play sounds at different volume levels and ask the participants to rate how loud they perceive them to be on a Likert item ranging from very quiet to very loud. The system playback volume can then assign participants conditions based on their perception of how loud the music was. This approach is perhaps the most appropriate solution, but the research question is changed so needs careful consideration.

The research should also be taken forwards to explore this idea of treating music as a multidimensional experiential gestalt in empirical research. I strongly suspect there is a hierarchy of dimensions of music manipulations, some of which can be controlled independently without too much impact on other dimensions (e.g. tempo and loudness as two dimensions of music which control the entire piece), and others which cannot (e.g. vocals). I still think it is interesting to explore how different dimensions of music affect people, as they clearly do in different ways, but a simple one-to-one relationship between the dimension manipulation and the effect does not seem to exist. Instead, it seems to be that manipulating one dimension can affect multiple dimensions with the overall cumulative effect having an impact on something else.

Finally, as I have already hinted at, this work does not need to be constrained to transcription typing. There are many other mundane work related computing tasks which could, and should, be explored in order to better understand how music can be used to affect performance and experience when working at a computer.

10.5 Conclusion

The primary objective of the research in this thesis was to explore whether music can affect transcription typing performance and experience in work related computing tasks. The ideal result would have been a simple set of recommendations of dimensions of music to incorporate in a playlist which should improve performance at work. The starting premise was a simple one, but the results have been complex and sometimes unintuitive. What is clear, at least, is that when it comes to the effect of music on people, the answers are not always simple. And, in fact, I now question some of the results from the previous research regarding tempo – by asking is it only true in 4/4 time music? Nevertheless, however complex a stimulus music is, it does have effects on people which can be identified clearly, even if not always well explained. Given the complexity of the stimulus, it is my recommendation that in future, researchers take heed of the pages of results presented due to my experiment designs and aim for much more simple manipulations.

What has emerged through this work, is mainly a key insight into a methodological problem which was not anticipated and has not routinely been considered within other research of this nature. Namely, that when manipulating one dimension of the music as an IV, there are simultaneous unintended changes to other dimensions which can affect the results of an experiment and lead to a reduced understanding of the overall effects identified. Instead of treating different dimensions of music as separate entities, in the empirical research context, it is better to consider the music as a whole, e.g. as a multidimensional experiential gestalt.

The key contribution from my research presented here is that how music affects people when typing is anything but simple. It is interesting that everyone has an opinion on this subject, I have had many conversations with people over the years where they tell me how music affects them. But, clearly, the literature review showed that people cannot always recognise correctly how music affects them. We know it does. But we do not know how or why. This thesis has taken us closer to refining better, more targeted research questions.

To summarise the outcomes from the research in this thesis in one sentence: music is complicated, people are complicated, and transcription typing is more complicated than you might expect. But, that is not to say it is uninteresting and should not be explored further. In fact, I am beginning to suspect that there may be a wealth of research in this sort of area, or similar, that is lying in many an academic's bottom drawer gathering dust due to similarly complex results. It may be that quantitative empirical work using a reductive approach to the science is simply inappropriate with music as a stimulus, and it might be better to consider music as a multidimensional experiential gestalt. It may just be that more work is needed. I suspect the latter.

Appendix A

Materials for Laboratory Based Experiments

This appendix includes:

- informed consent form
- data collection form for without music conditions
- data collection form for music conditions
- demographics questionnaires

A.1 Informed Consent Form

Informed Consent

This study is being conducted by Anna Bramwell-Dicks, a PhD student from the Human-Computer Interaction research group, within the Department of Computer Science at the University of York. This study forms part of the research towards her PhD investigating Affective Musical Interaction.

The study aims to investigate the effects that background music has on typing performance and user experience. You will be asked to complete a number of short typing tasks, with and without a background music accompaniment. At the end of each typing task, you will fill in a short questionnaire.

Your data will be completely confidential and anonymous and only viewed by Anna Bramwell-Dicks and her supervisors, Professor Helen Petrie and Dr Alistair Edwards.

Before you participate the study, please complete Section A, printing your name in the first space and then signing.

Once the study is over and you have been debriefed, you will be asked to initial the three statements in section B to indicate your agreement.

Section A

I, ______________________________, voluntarily give my consent to participate in this study for the Affective Music Interaction project. I have been informed about, and feel that I understand the basic nature of the study. I understand that I may withdraw from the study at any time without prejudice. I also understand that my information will be treated confidentially and will be anonymised. Only Anna Bramwell-Dicks and her supervisors, Professor Helen Petrie and Dr Alistair Edwards will have access to the data collected in its original format and it will only be shared in a grouped and anonymous format.

______________________ ______________________
Signature of participant Date

Section B

Please initial each of the following statements when the study has been completed and you have been debriefed.

I have been adequately debriefed.	Your initials:
I was not forced to complete the study.	Your initials:
All my questions have been answered.	Your initials:

A.2 Data Collection Form for Without Music Condition

Post Typing Test Questionnaire (Without Music)

Please rate on a scale from 1-7 how easy you found it to complete typing task?

Extremely difficult 1	2	3	Neither difficult nor easy 4	5	6	Extremely easy 7
☐	☐	☐	☐	☐	☐	☐

If a rating of 100% indicated your best typing performance, in terms of both speed and accuracy, how would you rate your performance at typing in this task?

______________ %

How long do you think this typing task lasted?

☐ Less than 1 minute

☐ Between 1 and 2 minutes

☐ Between 2 and 3 minutes

☐ Between 3 and 4 minutes

☐ Between 4 and 5 minutes

☐ Between 5 and 6 minutes

☐ Between 6 and 7 minutes

☐ Between 7 and 8 minutes

☐ Between 8 and 9 minutes

☐ Between 9 and 10 minutes

☐ Over 10 minutes

A.3 Data Collection Form for Music Conditions

Post Typing Task Questionnaire (Music 1)

Please rate on a scale from 1-7 how easy you found it to complete typing task?

Extremely difficult			Neither difficult nor easy			Extremely easy
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

Please rate on a scale from 1-7 how distracting you found the accompanying music?

Extremely distracting						Not at all distracting
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

If you found the music distracting (i.e. selected 1-3 in question above), please describe why it distracted you (e.g. it was too loud)?

____

____

____

____

____

____

____

____

Please rate on a scale from 1-7 how much you liked the accompanying music?

I hated it			I neither liked nor disliked it			I loved it
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

If a rating of 100% indicated your best typing performance, in terms of both speed and accuracy, how would you rate your performance at typing in this task?

______________ %

How long do you think this typing task lasted?

- ☐ Less than 1 minute
- ☐ Between 1 and 2 minutes
- ☐ Between 2 and 3 minutes
- ☐ Between 3 and 4 minutes
- ☐ Between 4 and 5 minutes
- ☐ Between 5 and 6 minutes
- ☐ Between 6 and 7 minutes
- ☐ Between 7 and 8 minutes
- ☐ Between 8 and 9 minutes
- ☐ Between 9 and 10 minutes
- ☐ Over 10 minutes

Had you ever heard this song before?

☐ Yes ☐ No ☐ I'm not sure

If yes, where have you heard it before?

____

____

A.4 Demographics Questionnaires

Post Study Demographics Questionnaire

What is your occupation? ____________________

If you are a student, what degree are you studying for?

____

What is your nationality? ____________________

Are you a native speaker of English? ☐ Yes ☐ No

Do you have any known hearing defects? ☐ Yes ☐ No

How old are you?

☐ Under 18	☐ 25 - 34	☐ 45 - 60
☐ 18 - 24	☐ 35 - 44	☐ Over 60

In an average week, how many hours do you spend using a computer for ***work purposes***?

☐ Less than 5 hours	☐ 5 – 10 hours	☐ 10 – 20 hours
☐ 20 – 30 hours	☐ 30 – 40 hours	☐ More than 40 hours

In an average week, how many hours do you spend using a computer for ***leisure***?

☐ Less than 5 hours	☐ 5 – 10 hours	☐ 10 – 20 hours
☐ 20 – 30 hours	☐ 30 – 40 hours	☐ More than 40 hours

Do you consider yourself a musician?	☐ Yes	☐ No
Do you play any instruments?	☐ Yes	☐ No

If yes, what instruments do you play and how long have you played them?

Post Study Music Use Questionnaire

Do you ever listen to music when you are using a computer for ***work purposes?***

☐ Yes ☐ No

If "yes", as a percentage of time how often do you listen to music when using a computer for ***work purposes?***

Under 10%	10% - 30%	30% - 50%	50% - 70%	70% - 90%	Over 90%
☐	☐	☐	☐	☐	☐

If yes, why do you choose to listen to music when using your computer for ***work purposes?***

____

____

____

____

____

____

____

____

Do you have a specific "work music" playlist?

☐ Yes ☐ No

If yes, please describe the types of songs on your work music playlist and why you have chosen them?

____

____

____

____

Do you ever listen to music when you are using a computer for ***leisure?***

☐ Yes ☐ No

If "yes", as a percentage of time how often do you listen to music when using a computer for ***leisure?***

Under 10%	10% - 30%	30% - 50%	50% - 70%	70% - 90%	Over 90%
☐	☐	☐	☐	☐	☐

If yes, why do you choose to listen to music when using your computer for ***leisure?***

Please list your favourite type(s) of music?

If you have a favourite artist, who is it?

Appendix B

Materials for Classroom Based Experiments

This appendix includes:

- informed consent form
- experimenter instructions
- screenshots of experiment interface
- example data collection form for music conditions
- demographics questionnaire

B.1 Informed Consent Form

Informed Consent

The purpose of this form is to tell you about the study and highlight features of your participation in the study.

Who is running this?

The study is being run by Anna Bramwell-Dicks, who is a PhD student and Teaching Fellow in the department of Computer Science, at the University of York.

Your experimenter is a student taking the Human Aspects of Computer Science module, and is a first year undergraduate in the department.

What is the purpose of the study?

The study aims to investigate how background music affects people when copy typing.

What will I have to do?

You will be asked to copy passages of text on a website while listening to music. After a few minutes the music will stop and you will answer some quick questions about your experience of typing while listening to that piece of music. This process will be repeated a few times. At the end of the experiment, you will complete a longer paper-based questionnaire.

Who will see this data?

Obviously the experimenter with you will see this data. It will also be passed on to Anna BD for compiling. However, once it has been compiled, it will be completely anonymised and you will not be able to be identified with your data. The experiment may be published in an academic journal but the data will only be presented in summary form and you will not be directly identifiable in any way.

Do I have to do this?

Your participation is completely voluntary. You can therefore withdraw from the study at any point and if requested your data can be destroyed.

Unusually for a study, this is a class exercise and Dr Paul Cairns strongly advises that you take part to see how such experiments work from both the participant and experimenter perspective.

10 participants from across the two HACS classes will be randomly selected to win a £10 Amazon gift voucher. In addition, the top 3 scores across the two

classes will win £30, £20 and £10 Amazon gift vouchers respectively. Both speed and accuracy are taken into account when calculating a participant's score.

Can I ask a question?

Do ask the experimenter any questions you may have about the procedure that you are about to follow. However, during the study, please refrain from talking to the experimenter and save any questions you may have until the end of the practical class.

If you have any questions about the purpose or background of the experiment, please wait until the end of the class and you will have an opportunity to ask Anna BD or Paul your questions.

Consent

Please sign below that you agree to take part in the study under the conditions laid out above. This will indicate that you have read and understood the above and that we will be obliged to treat your data as described.

Name:

Signature:

Date

B.2 Experimenter Instructions

Experimenter Instructions

There is a wealth of evidence that music affects task performance in different contexts. This experiment is investigating how music affects copy-typing performance – you will ask your participant to copy-type passages of text while listening to different pieces of music.

In this experiment, we are interested in the affect of a) music volume (between participants), b) presence or absence of vocals (within participants) and c) the language of the text that they are copying (within participants).

What you have to do

1. First, you need to briefly explain what the experiment is about to the participant and thank them for their time. So read them the following blurb:

   "Thank you for taking part in this experiment, which is part of Anna BD's PhD studies, where she is looking at how background music affects users in different situations. This particular experiment is looking at the impact of background music on copy-typing. You will be asked to complete a few typing tasks while listening to different pieces of music.

   Your participation is entirely voluntary, you can withdraw at any point during the experiment and your data will be destroyed. 10 participants will be randomly selected to win £10 Amazon voucher, while the top 3 scores will win £30, £20 and £10 Amazon vouchers respectively. Are you happy to continue?"

2. If they say "yes" give the participant the **Informed Consent** form to read and sign. If they say "no" ask a demonstrator to come over.
3. Give the participant a pair of headphones and ask them to put them on.
4. Check that the volume on the computer is set to **MAXIMUM**
5. Open the webpage: http://www-users.cs.york.ac.uk/annab/HACS in **Firefox**
6. Ask the participant to read the webpage and then enter their york username (e.g. abc100) and participant number (you'll find this at the top of the small **experiment order sheet** which should be attached to this document and has a table on it). Reassure the participant that their email address will only be used to send an Amazon voucher, should they win.

7. When the participant clicks on the "Okay! I'm in. Let's get started." button a webpage will load that contains a practice typing task. Tell the participant that this is a shorter task to practice the process and make sure you both understand what you need to do. You are also going to check that the volume is okay for the participant.

8. **Tell the participant that they must not use the mouse to move the cursor during the typing tasks.**

9. Tell the participant that they are aiming **to accurately copy-type the text as fast as they can**, and that the music will begin playing when they start typing.

10. When they start typing, the music will begin and a timer is triggered. But, timing often goes wrong so **you should definitely time their typing as well**. Let the participant type for **20 seconds** in this practice task.

11. After 20 seconds an alert should pop up asking them to stop. If this alert does not come up, please stop the participant copy-typing.

12. Enter the participant number in the text box at the bottom of the screen (you might need to scroll down) and then click on the button to submit their data. This creates a text file on the server that contains the text that they typed. The website then loads a short form for the participant to completes.

13. That was the practice. Check that the participant understands the process and that they are happy with the volume of the music. If they are not happy, then you need to speak with Anna BD and she will sort something out. **Do not change the volume of the computer without Anna's permission!**

14. Ask the participant if they are ready to try a real typing task. If they are, then you need to select their first text type and music type according to the table on the small **experiment order sheet**.

    It is important that you choose the right options here, as this will preserve the complex counterbalancing designed by Anna.
    (E1=English text 1, E2=English text 2, D1=Dutch text 1, D2=Dutch text 2, **With=With vocals, Without=Without vocals**.)

    Click on the button and a webpage will load with the correct typing task and music options.

15. For the real typing tasks, the participant needs to be left to copy type for **4 minutes and 30 seconds** (i.e. 270 seconds). Again, the music and timer will start when they start to type. But please time them as well.

16. After each **real task** the participant needs to fill in rating scales on one side of their handout containing rating scales.

17. Repeat steps 14 to 16 until the participant has typed in both Dutch and English **twice**. (Four of the "real" typing tasks in addition to the practice).

18. Give the participant the **demographics questionnaire** to complete.

19. Give the participant the **CFQ and its correlates** questionnaire to complete.

    After completing the experiment, the participant should have filled in two single-sided paper-based forms and 1 pack of Ratings. This data now needs to be entered into a spreadsheet in a precise manner ready for analysis.
    You can find the spreadsheet at :
    http://www-users.cs.york.ac.uk/annab/HACS/data.xlsx

B.3 Experiment Interface

B.3.1 Set Up Instructions

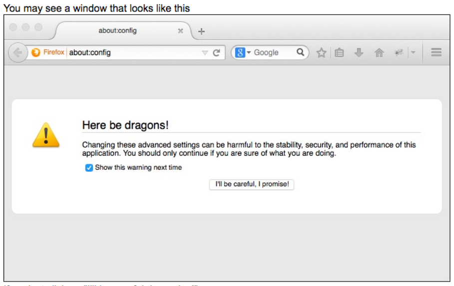
Experimenter set up

For the experiment to work properly, you need to set up the system as follows:

1. First, the Firefox window must be maximised and not in full screen mode. Please click on the maximise icon to maximise the window size now.

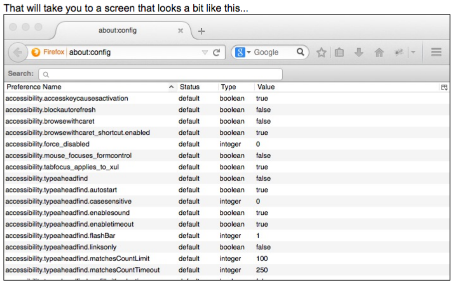
2. To ensure that all of your participant's data is collected by the system, it is very important that when hitting the backspace key on a webpage the browser does not go back a page. To fix this, you must follow this process to temporarily disable this function of the browser:In Firefox's address bar, type about:config and then press enter.

   You may see a window that looks like this

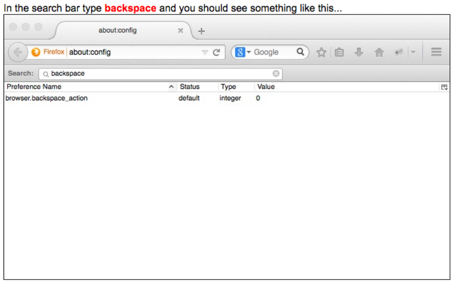


   If so, just click on "I'll be careful, I promise!"

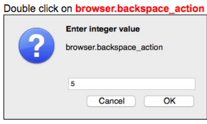
3. That will take you to a screen that looks a bit like this...



4. In the search bar type **backspace** and you should see something like this...

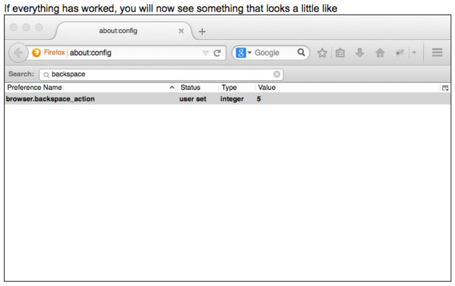


5. Double click on **browser.backspace_action** which will bring up the following pop up where you should **enter the number 5 and click on the OK button**



Experimenter setup instructions – part 1

6. If everything has worked, you will now see something that looks a little like



7. If you **don't see a window like this**, please raise your hand and ask a demonstrator to come across to help you.

8. Finally, you need to maximise the volume of the audio on the computer. Please do this now by clicking on the volume control icon in the bottom right corner of the screen and moving the slider to the top.

Please tick this checkbox to confirm that you have maximised the size of the firefox window. ☐

Please tick this checkbox to confirm that you have followed the above instructions to remove the "backspace means page back" functionality. ☐

Please tick this checkbox to confirm that you have maximised the volume of the computer. ☐

Please also enter your participant's ID number, #: ____

[Okay, let's get this experiment started!]

Experimenter setup instructions – part 2

B.3.2 Music Selection Interfaces

Thank you

As this was the practice you do **not** have to complete any Lickert scale ratings. But after each of the **real** tasks you be asked a few questions about the typing task you just did

Before you participate in a real task, please consider whether you were happy with the volume of the music playing. If you felt it was too loud, then you **must** tell one of the demonstrators.

But, assuming the volume was okay for you...the experimenter should now choose the first task from the following options. Don't ask them to explain the terms yet, they will do so at the end of the experiment:

Text type:

◯ E 1 ◯ E 2 ◯ P 1 ◯ P 2

Music type:

◯ With Vs ◯ Without Vs

[Click to select the chosen text and music options.]

Interface to select first task, loaded after the practice task data has been submitted

Thank you

Please fill in **one side** of the participant rating scales questions.

When you have completed that form....

Text type:

◯ E 1 ◯ E 2 ◯ P 1 ◯ P 2

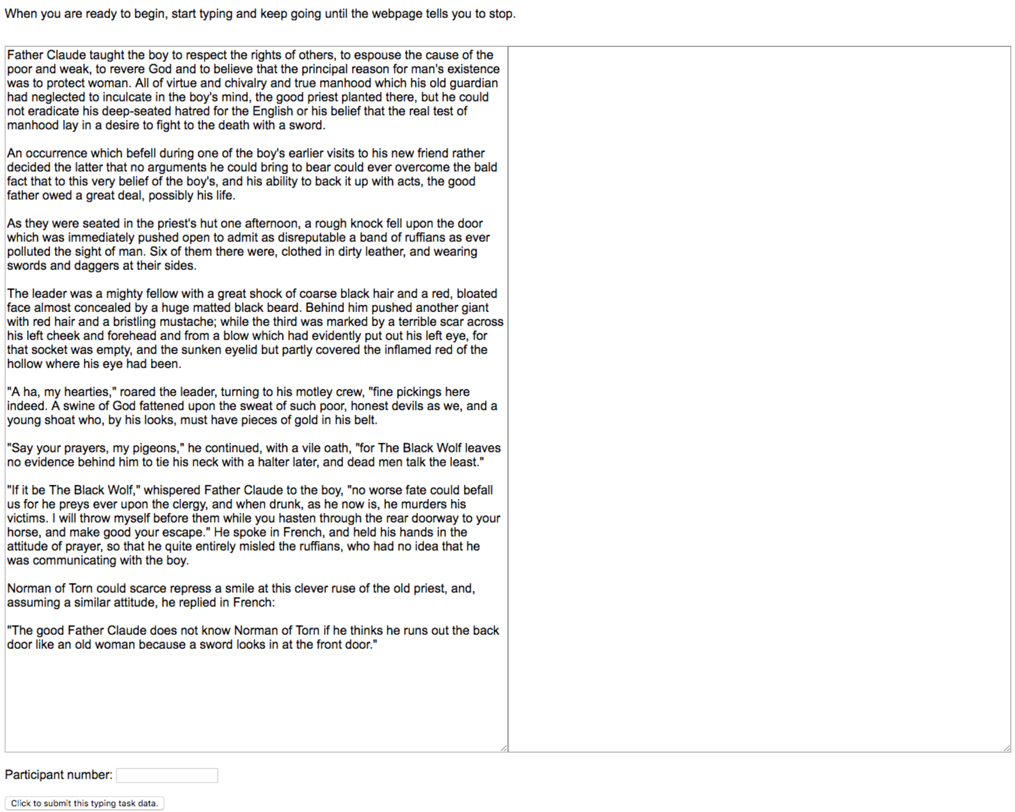
Music type:

◯ With Vs ◯ Without Vs

[Click to select the chosen text and music options.]

Interface to select real transcription typing tasks after data has been submitted

B.3.3 Transcription Task Interface



Transcription typing task interface example

B.4 Example Data Collection Form for Music Conditions

Participant Number

Typing Task 1

Please rate on a scale from 1-7 how easy you found it to complete typing test?

Extremely difficult			Neither difficult nor easy			Extremely easy
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

Please rate on a scale from 1-7 how distracting you found the accompanying music?

Extremely distracting						Not at all distracting
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

Please rate on a scale from 1-7 how much you liked the accompanying music?

I hated it			I neither liked nor disliked it			I loved it
1	2	3	4	5	6	7
☐	☐	☐	☐	☐	☐	☐

Page 1 of 4

B.5 Demographics Questionnaires

Participant Number:

Post Study Demographics Questionnaire

What is your occupation? ______________________________

If you are a student, what degree are you studying for?

____

What is your nationality? ______________________________

Are you a native speaker of English? ☐ Yes ☐ No

Do you speak Dutch? ☐ Yes ☐ No

Are you?

☐ Male ☐ Other

☐ Female ☐ Would rather not say

How old are you?

☐ Under 18 ☐ 25 - 34 ☐ 45 - 60

☐ 18 - 24 ☐ 35 - 44 ☐ Over 60

☐ Would rather not say

In an average week, how many hours do you spend using a computer for ***work purposes***?

☐ Less than 5 hours ☐ 5 – 10 hours ☐ 10 – 20 hours

☐ 20 – 30 hours ☐ 30 – 40 hours ☐ More than 40 hours

In an average week, how many hours do you spend using a computer for ***leisure***?

☐ Less than 5 hours ☐ 5 – 10 hours ☐ 10 – 20 hours

☐ 20 – 30 hours ☐ 30 – 40 hours ☐ More than 40 hours

Appendix C

Flesch Reading Score Analysis of English Texts

This appendix includes:

- descriptive statistics for the Flesch {*Flesch:tgh9mTlM} reading scores for the four simple and advanced English presented text passages,
- the results from Mann Whitney U tests to compare the Flesch {*Flesch:tgh9mTlM} reading scores between the texts passages

C.1 Reading Scores Analysis

C.1.1 Descriptive Statistics

Simple English Text	Flesch Reading Score
1	80.60
2	74.10
3	81.20
4	79.50
Mean	78.85

Flesch Reading Score for simple English presented text passages

Advanced English Text	Flesch Reading Score
1	67.50
2	65.60
3	62.40
4	61.50
Mean	64.25

Flesch Reading Score for simple English presented text passages

C.1.2 Mann Whitney U Test Comparison

There was a significant difference between the reading scores in the simple and advanced text conditions, $U(8) = 0.000$, $z = -2.308$, $p = 0.021$.

Appendix D

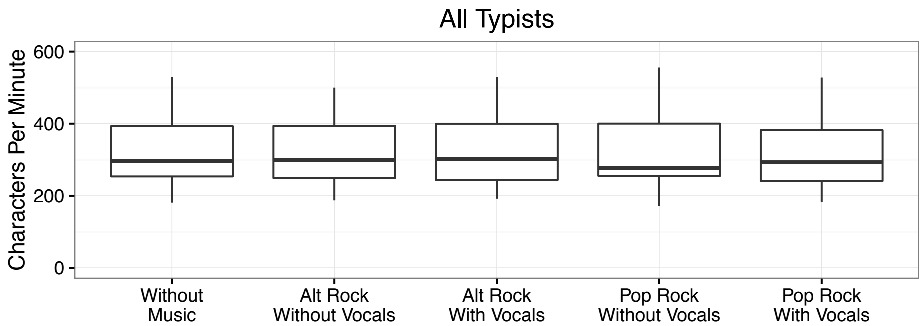
Experiment 1 Supplementary Statistical Analyses

This appendix includes:

- normality tests and box plots for CPM, error rate percentage and the transformed error rate percentage

D.1 Normality Tests and Box Plots

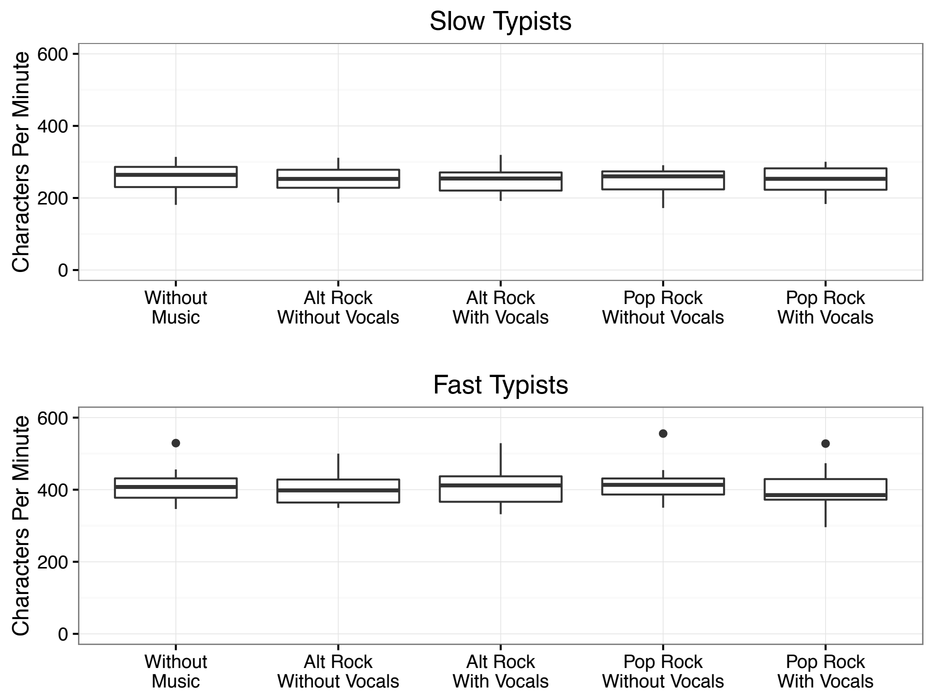
D.1.1 Characters Per Minute



CPM box plot for all typists

All Typists					
Music	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
		s	$\mathbf{z_s}$	**k**	$\mathbf{z_k}$
Without music	$W(28) = 0.963$, *n.s.*	0.39	0.89	-0.62	-0.72
Alt rock without vocals	$W(28) = 0.956$, *n.s.*	0.28	0.63	-0.91	-1.06
Alt rock with vocals	$W(28) = 0.946$, *n.s.*	0.48	1.08	-0.74	-0.87
Pop rock without vocals	$W(28) = 0.937$, *n.s.*	0.50	1.12	-0.46	-0.53
Pop rock with vocals	$W(28) = 0.953$, *n.s.*	0.55	1.25	-0.41	-0.47

CPM Shapiro-Wilk's tests, skewness and kurtosis for all typists

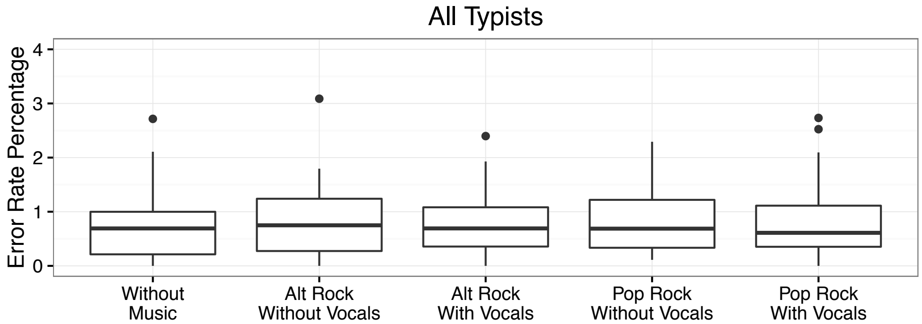


CPM box plots for slow and fast typists

Slow Typists					
Music	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
		s	**z_s**	**k**	**z_k**
Without music	$W(16) = 0.953$, *n.s.*	-0.43	-0.76	-0.86	-0.79
Alt rock without vocals	$W(16) = 0.951$, *n.s.*	-0.25	-0.43	-1.03	-0.94
Alt rock with vocals	$W(16) = 0.952$, *n.s.*	0.09	0.15	-0.84	-0.77
Pop rock without vocals	$W(16) = 0.889$, *n.s.*	-0.97	-1.72	0.01	0.01
Pop rock with vocals	$W(16) = 0.923$, *n.s.*	-0.31	-0.55	-1.28	-1.17
Fast Typists					
Music	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
		s	**z_s**	**k**	**z_k**
Without music	$W(12) = 0.933$, *n.s.*	1.00	1.57	1.57	1.27
Alt rock without vocals	$W(12) = 0.916$, *n.s.*	0.98	1.54	1.20	0.97
Alt rock with vocals	$W(12) = 0.963$, *n.s.*	0.51	0.81	0.25	0.20
Pop rock without vocals	$W(12) = 0.869$, *n.s.*	1.62	2.54	3.98	3.23
Pop rock with vocals	$W(12) = 0.971$, *n.s.*	0.51	0.80	0.62	0.50

CPM Shapiro-Wilk's tests, skewness and kurtosis

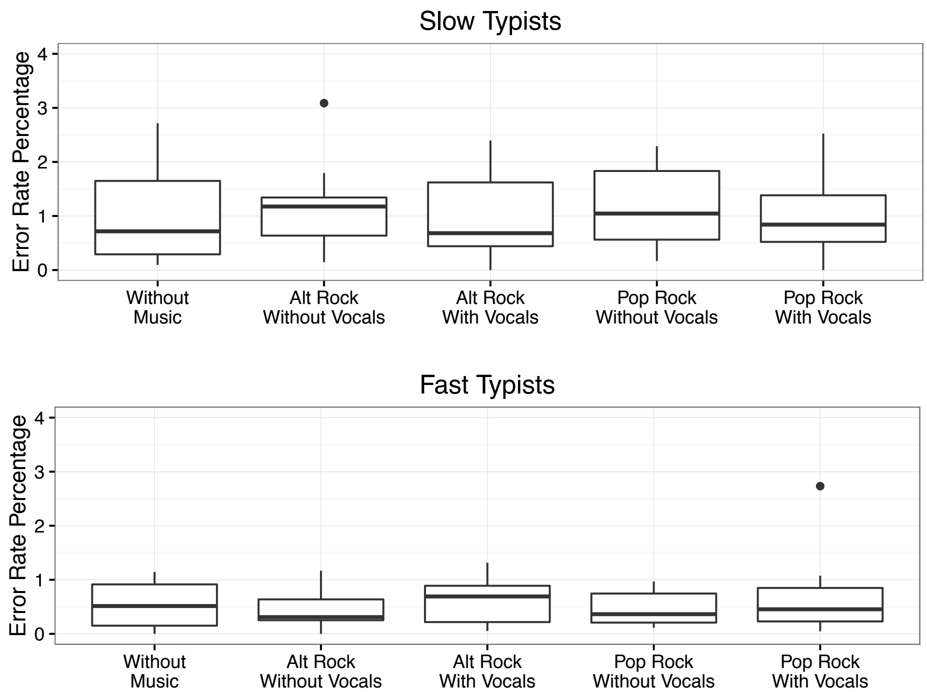
D.1.2 Error Rate



Error rate box plots for all typists

All Typists					
Music	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
		s	**z_s**	**k**	**z_k**
Without music	$W(28) = 0.881$, $p = 0.004$	-0.36	-0.81	-1.02	-1.19
Alt rock without vocals	$W(28) = 0.876$, $p = 0.003$	-0.31	-0.70	-1.10	-1.28
Alt rock with vocals	$W(28) = 0.917$, $p = 0.029$	-0.88	-1.99	0.74	0.86
Pop rock without vocals	$W(28) = 0.883$, $p = 0.005$	-0.32	-0.72	-0.81	-0.94
Pop rock with vocals	$W(28) = 0.884$, $p = 0.005$	-0.79	-1.80	0.63	0.73

Error rate Shapiro-Wilk's tests skewness and kurtosis for all typists

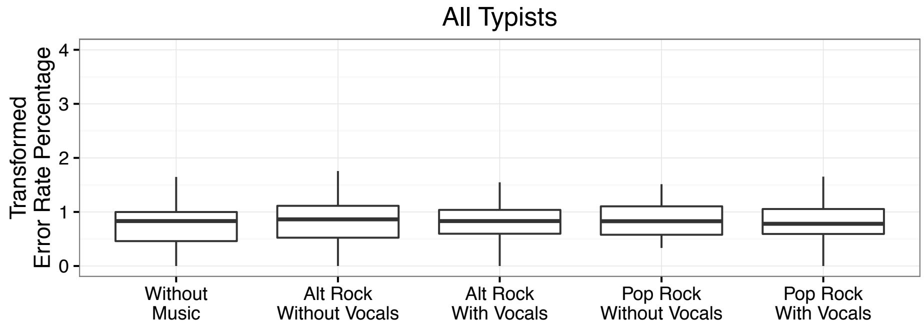


Error rate box plots for slow and fast typists

Slow Typists						
Music	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**		
		s	z_s	**k**	z_k
Without music	$W(16) = 0.897$, *n.s.*	-0.40	-0.70	-1.00	-0.92
Alt rock without vocals	$W(16) = 0.897$, $p = 0.041$	-1.07	-1.89	0.27	0.25
Alt rock with vocals	$W(16) = 0.897$, *n.s.*	-0.01	-0.01	-1.55	-1.42
Pop rock without vocals	$W(16) = 0.924$, *n.s.*	-0.70	-1.24	-0.04	-0.03
Pop rock with vocals	$W(16) = 0.952$, $p = 0.002$	-1.18	-2.08	2.36	2.16
Fast Typists					
Music	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
		s	z_s	**k**	z_k
Without music	$W(12) = 0.879$, *n.s.*	-0.64	-1.00	-1.41	-1.14
Alt rock without vocals	$W(12) = 0.821$, *n.s.*	0.38	0.59	-1.67	-1.36
Alt rock with vocals	$W(12) = 0.935$, *n.s.*	-0.91	-1.43	-0.38	-0.30
Pop rock without vocals	$W(12) = 0.880$, *n.s.*	-0.13	-0.21	-1.36	-1.10
Pop rock with vocals	$W(12) = 0.733$, *n.s.*	-0.32	-0.50	0.57	0.46

Error rate Shapiro-Wilk's tests skewness and kurtosis

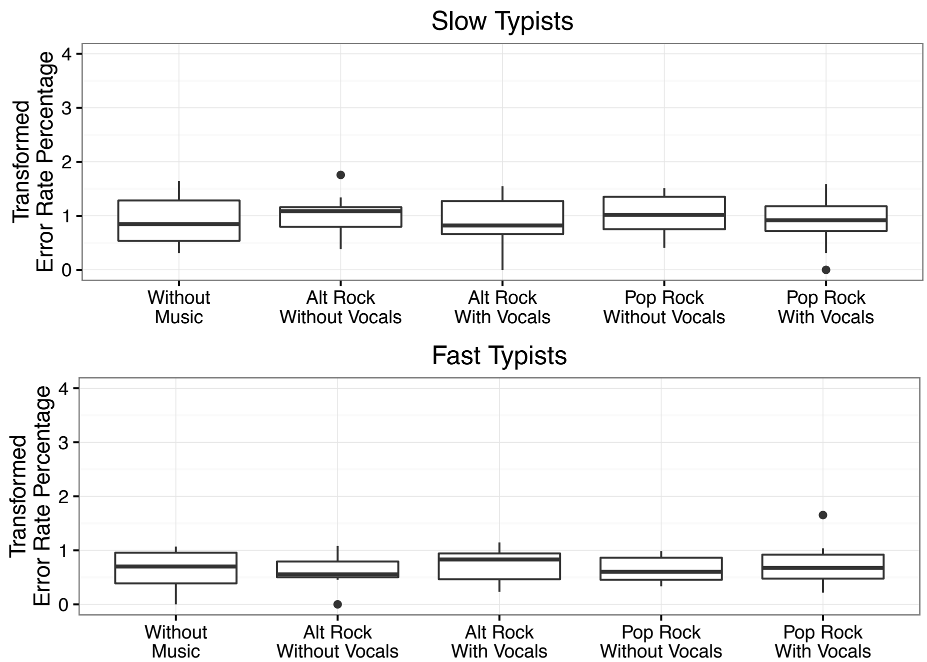
D.1.3 Transformed Error Rate



Transformed error rate box plots for all typists

All Typists					
Music	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
		s	$\mathbf{z_s}$	**k**	$\mathbf{z_k}$
Without music	*W*(28) = 0.938, *n.s.*	0.22	0.51	-0.51	-0.59
Alt rock without vocals	*W*(28) = 0.927, *n.s.*	0.10	0.22	0.00	0.00
Alt rock with vocals	*W*(28) = 0.945, *n.s.*	-0.05	-0.12	-0.33	-0.38
Pop rock without vocals	*W*(28) = 0.958, *n.s.*	0.36	0.83	-0.90	-1.05
Pop rock with vocals	*W*(28) = 0.957, *n.s.*	0.18	0.40	-0.15	-0.17

Transformed error rate Shapiro-Wilk's tests skewness and kurtosis for all typists



Transformed error rate box plots for slow and fast typists

Slow Typists						
Music	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**		
		s	z_s	**k**	z_k	
Without music	$W(16) = 0.945$, *n.s.*	0.26	0.45	-1.15	-1.05	
Alt rock without vocals	$W(16) = 0.842$, $p = 0.010$	-0.18	-0.32	0.30	0.27	
Alt rock with vocals	$W(16) = 0.919$, *n.s.*	-0.30	-0.53	-0.19	-0.17	
Pop rock without vocals	$W(16) = 0.931$, *n.s.*	-0.10	-0.18	-1.22	-1.12	
Pop rock with vocals	$W(16) = 0.927$, *n.s.*	-0.46	-0.82	0.35	0.32	
Fast Typists						
Music	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**		
		s	z_s	**k**	z_k	
Without music	$W(12) = 0.835$, $p = 0.024$	-0.47	-0.73	-0.96	-0.78	
Alt rock without vocals	$W(12) = 0.868$, *n.s.*	-0.26	-0.41	0.80	0.65	
Alt rock with vocals	$W(12) = 0.881$, *n.s.*	-0.38	-0.60	-1.26	-1.02	
Pop rock without vocals	$W(12) = 0.925$, *n.s.*	0.27	0.42	-1.43	-1.16	
Pop rock with vocals	$W(12) = 0.982$, *n.s.*	1.22	1.92	2.26	1.83	

Transformed error rate Shapiro-Wilk's tests skewness and kurtosis for fast typists

Appendix D: Experiment 1 Supplementary Statistical Analyses

Appendix E

Experiment 2 Supplementary Statistical Analyses

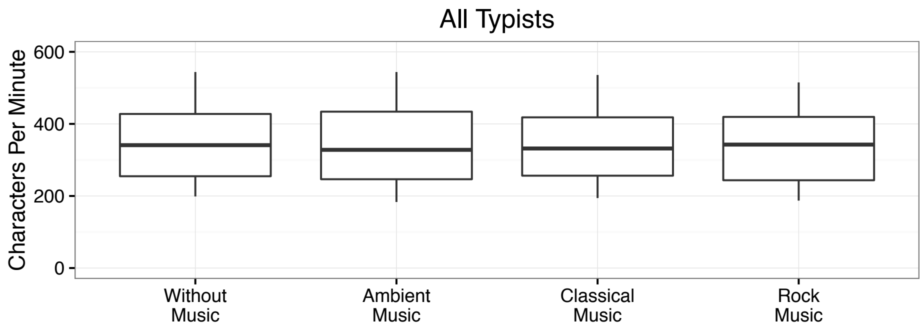
This appendix includes:

- normality tests and box plots for CPM, error rate percentage and the transformed error rate percentage

E.1 Normality Tests and Box Plots

E.1.1 Analysis Paradigm 1

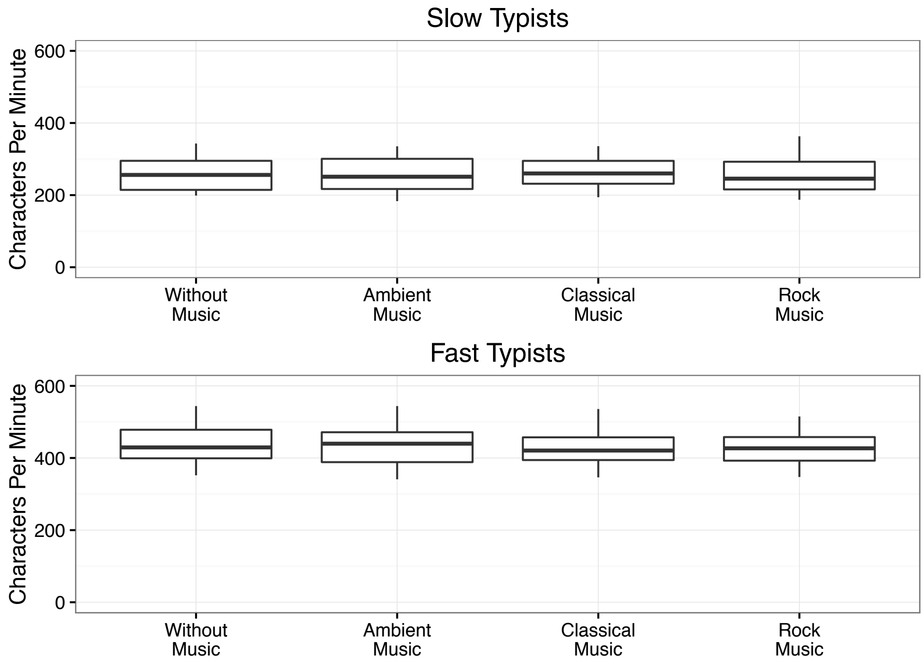
Characters Per Minute



CPM box plot for all typists

All Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Advanced English	Without	$W(42) = 0.943, p = 0.038$	0.21	0.57	-1.13	-1.58
	Ambient	$W(42) = 0.949$, *n.s.*	0.21	0.57	-1.13	-1.57
	Classical	$W(42) = 0.959$, *n.s.*	0.18	0.49	-1.04	-1.45
	Rock	$W(42) = 0.945, p = 0.043$	0.07	0.20	-1.27	-1.78

CPM Shapiro-Wilk's tests, skewness and kurtosis for all typists

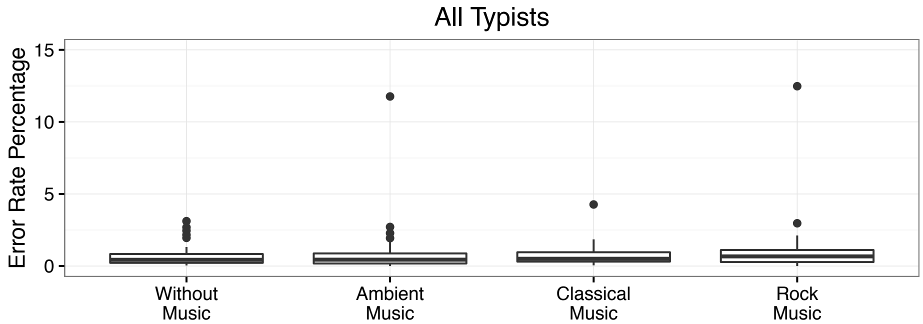


CPM box plots for slow and fast typists

Slow Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Advanced English	Without	*W*(22) = 0.928, *n.s.*	0.37	0.76	-1.00	-1.05
	Ambient	*W*(22) = 0.924, *n.s.*	0.11	0.23	-1.47	-1.54
	Classical	*W*(22) = 0.952, *n.s.*	-0.02	-0.04	-1.11	-1.17
	Rock	*W*(22) = 0.955, *n.s.*	0.47	0.96	-0.52	-0.54
Fast Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Advanced English	Without	*W*(20) = 0.951, *n.s.*	0.49	0.96	-0.57	-0.58
	Ambient	*W*(20) = 0.969, *n.s.*	0.22	0.43	-0.71	-0.72
	Classical	*W*(20) = 0.970, *n.s.*	0.53	1.03	-0.05	-0.05
	Rock	*W*(20) = 0.973, *n.s.*	0.18	0.36	-0.75	-0.76

CPM Shapiro-Wilk's tests, skewness and kurtosis

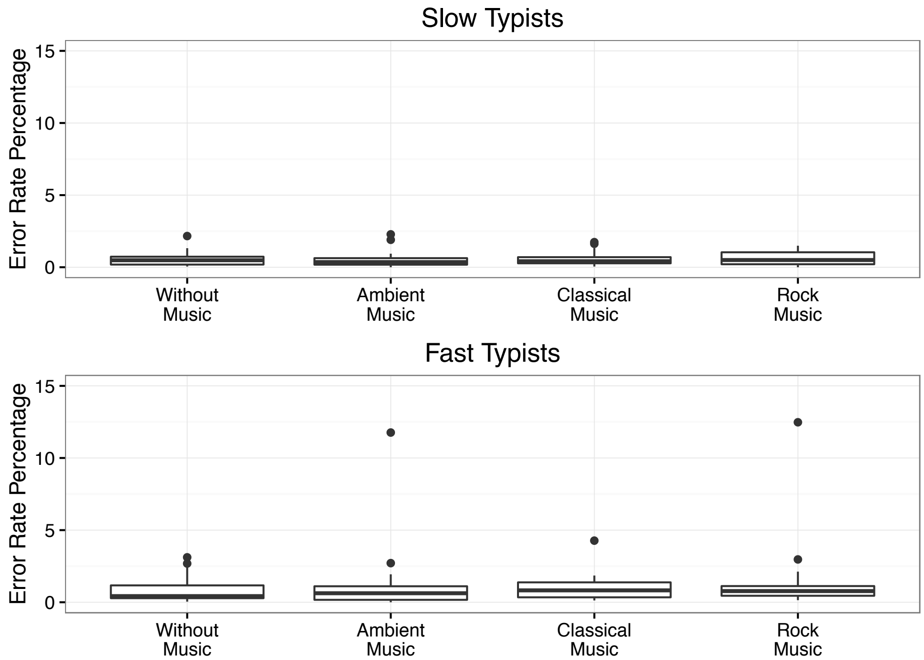
Error Rate



Error rate box plot for all typists

All Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	$\mathbf{z_s}$	**k**	$\mathbf{z_k}$
Advanced English	Without	$W(42) = 0.785, p < 0.0005$	1.75	**4.79***	2.61	**3.65***
	Ambient	$W(42) = 0.408, p < 0.0005$	5.33	**14.61***	31.58	**44.05***
	Classical	$W(42) = 0.743, p < 0.0005$	2.59	**7.10***	9.68	**13.51***
	Rock	$W(42) = 0.388, p < 0.0005$	5.55	**15.19***	33.54	**46.77***

Error rate Shapiro-Wilk's tests, skewness and kurtosis for all typists
(* values that exceed acceptable range for skewness or kurtosis of population)

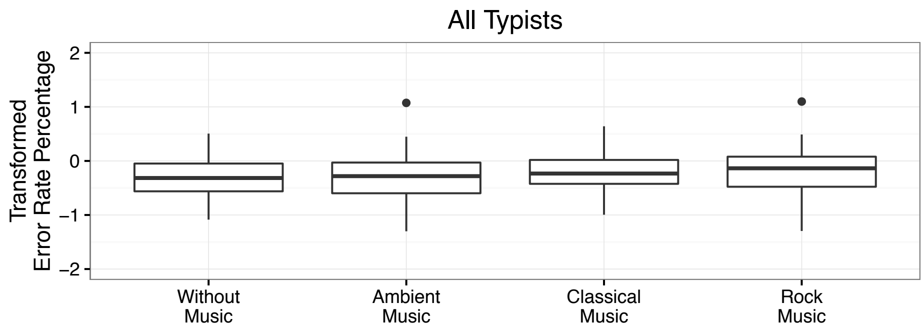


Error rate box plots for slow and fast typists

Slow Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Advanced English	Without	$W(22) = 0.834, p = 0.002$	1.67	**3.39***	3.54	**3.72***
	Ambient	$W(22) = 0.753, p < 0.0005$	2.08	**4.23***	4.33	4.54*
	Classical	$W(22) = 0.843, p = 0.003$	1.41	**2.86***	1.33	1.39
	Rock	$W(22) = 0.897, p = 0.026$	0.62	1.27	-1.01	-1.06
Fast Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Advanced English	Without	$W(20) = 0.798, p = 0.001$	1.35	**2.63***	0.67	0.67
	Ambient	$W(20) = 0.468, p < 0.0005$	3.94	**7.70***	16.56	**16.70***
	Classical	$W(20) = 0.771, p < 0.0005$	2.20	**4.30***	6.49	**6.54***
	Rock	$W(20) = 0.445, p < 0.0005$	4.01	**7.84***	16.99	**17.13***

Error rate Shapiro-Wilk's tests, skewness and kurtosis
(* values that exceed acceptable range for skewness or kurtosis of population)

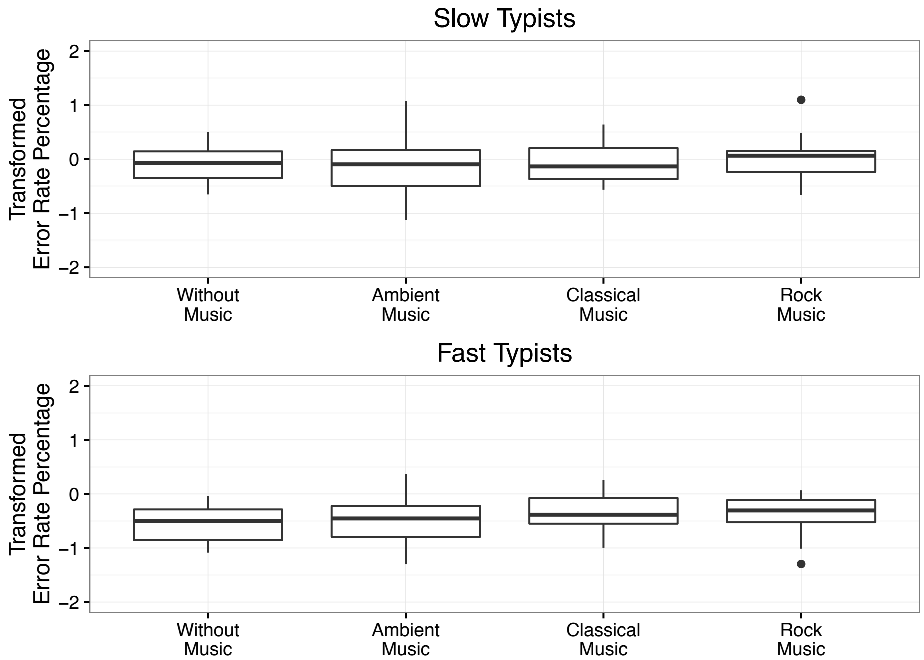
Transformed Error Rate



Transformed error rate box plot for all typists

All Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	z_s	**k**	z_k
Advanced English	Without	$W(42) = 0.972$, *n.s.*	-0.10	-0.26	-0.51	-0.71
	Ambient	$W(42) = 0.980$, *n.s.*	-0.19	-0.52	-0.56	-0.78
	Classical	$W(42) = 0.984$, *n.s.*	0.13	0.35	-0.23	-0.32
	Rock	$W(42) = 0.971$, *n.s.*	-0.55	-1.50	0.53	0.74

Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis with outlier capping



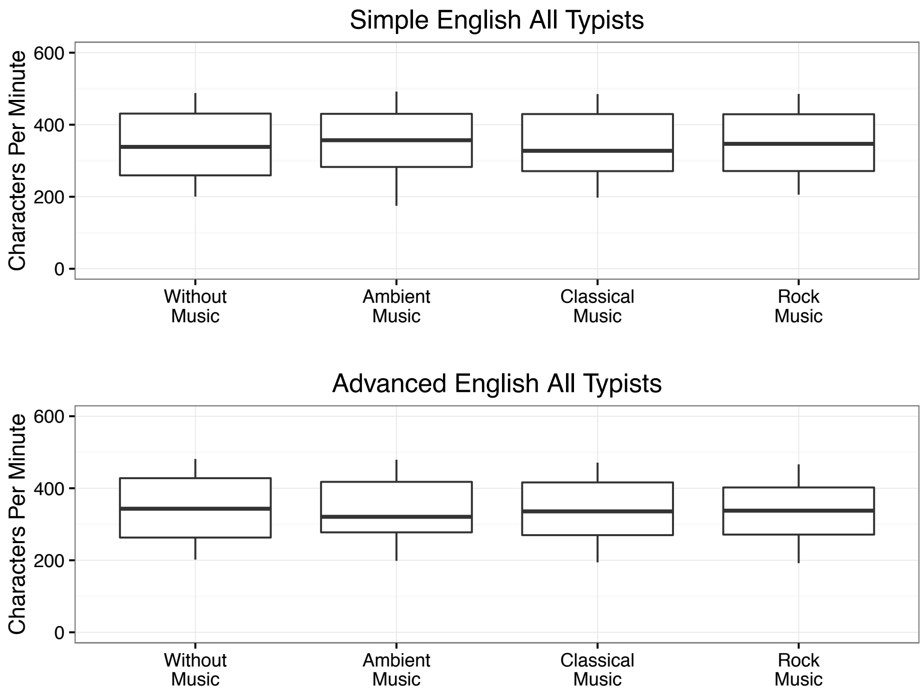
Transformed error rate box plots for slow and fast typists

Slow Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	z_s	**k**	z_k
Advanced English	Without	*W*(22) = 0.948, *n.s.*	-0.10	-0.21	-1.07	-1.12
	Ambient	*W*(22) = 0.974, *n.s.*	0.20	0.41	0.88	0.93
	Classical	*W*(22) = 0.950, *n.s.*	0.40	0.81	-0.59	-0.62
	Rock	*W*(22) = 0.932, *n.s.*	-0.48	-0.97	-0.61	-0.64
Fast Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	z_s	**k**	z_k
Advanced English	Without	*W*(20) = 0.906, *n.s.*	-0.39	-0.76	-1.18	-1.19
	Ambient	*W*(20) = 0.981, *n.s.*	-0.03	-0.06	-0.21	-0.21
	Classical	*W*(20) = 0.980, *n.s.*	-0.06	-0.12	-0.53	-0.54
	Rock	*W*(20) = 0.924, *n.s.*	-0.78	-1.53	0.15	0.15

Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis with outlier capping

E.1.2 Analysis Paradigm 2

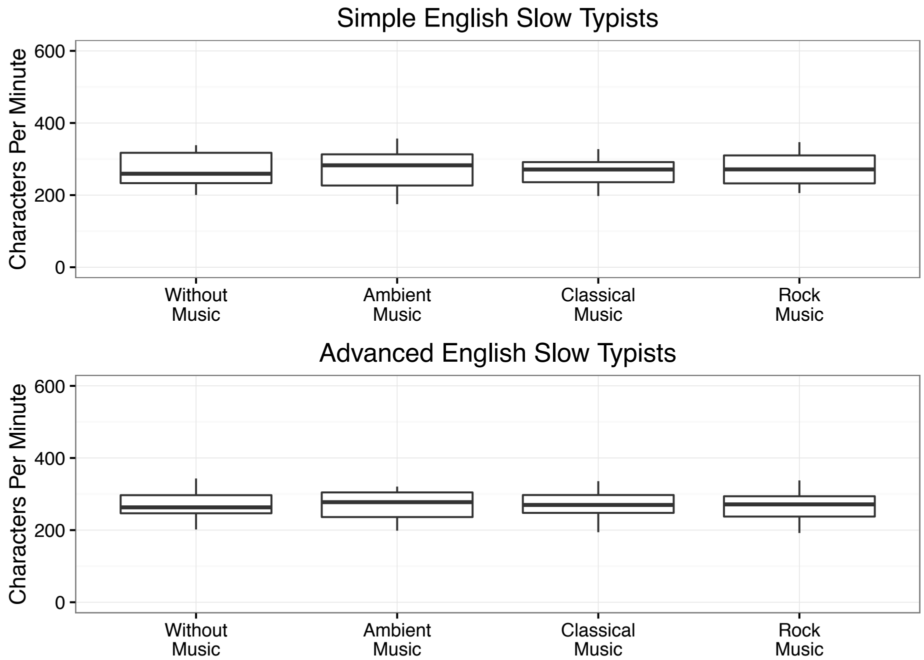
Characters Per Minute



CPM box plots for simple and advanced English tasks, for all typists

All Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Simple English	Without	$W(21) = 0.932$, *n.s.*	-0.07	-0.14	-1.43	-1.47
	Ambient	$W(21) = 0.952$, *n.s.*	-0.20	-0.40	-1.11	-1.14
	Classical	$W(21) = 0.920$, *n.s.*	-0.06	-0.12	-1.53	-1.57
	Rock	$W(21) = 0.923$, *n.s.*	-0.11	-0.22	-1.49	-1.53
Advanced English	Without	$W(21) = 0.926$, *n.s.*	-0.03	-0.07	-1.46	-1.50
	Ambient	$W(21) = 0.939$, *n.s.*	-0.03	-0.06	-1.34	-1.38
	Classical	$W(21) = 0.937$, *n.s.*	-0.08	-0.16	-1.35	-1.39
	Rock	$W(21) = 0.939$, *n.s.*	-0.08	-0.15	-1.34	-1.38

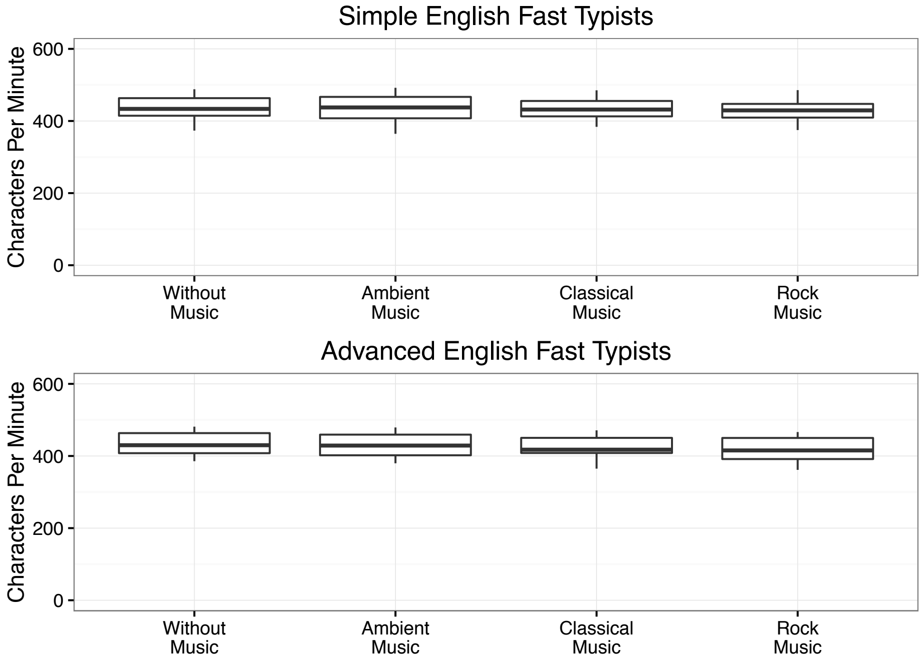
CPM Shapiro-Wilk's tests, skewness and kurtosis for all typists



CPM box plots for simple and advanced English tasks, for slow typists

Slow Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	z_s	**k**	z_k
Simple English	Without	$W(11) = 0.928$, *n.s.*	0.13	0.19	-1.52	-1.19
	Ambient	$W(11) = 0.956$, *n.s.*	-0.34	-0.51	-0.97	-0.76
	Classical	$W(11) = 0.961$, *n.s.*	-0.07	-0.11	-0.90	-0.70
	Rock	$W(11) = 0.946$, *n.s.*	0.26	0.39	-1.26	-0.99
Advanced English	Without	$W(11) = 0.963$, *n.s.*	0.26	0.40	-0.61	-0.47
	Ambient	$W(11) = 0.920$, *n.s.*	-0.45	-0.69	-1.29	-1.01
	Classical	$W(11) = 0.969$, *n.s.*	-0.23	-0.34	-0.57	-0.45
	Rock	$W(11) = 0.967$, *n.s.*	-0.06	-0.08	-0.77	-0.60

CPM Shapiro-Wilk's tests, skewness and kurtosis

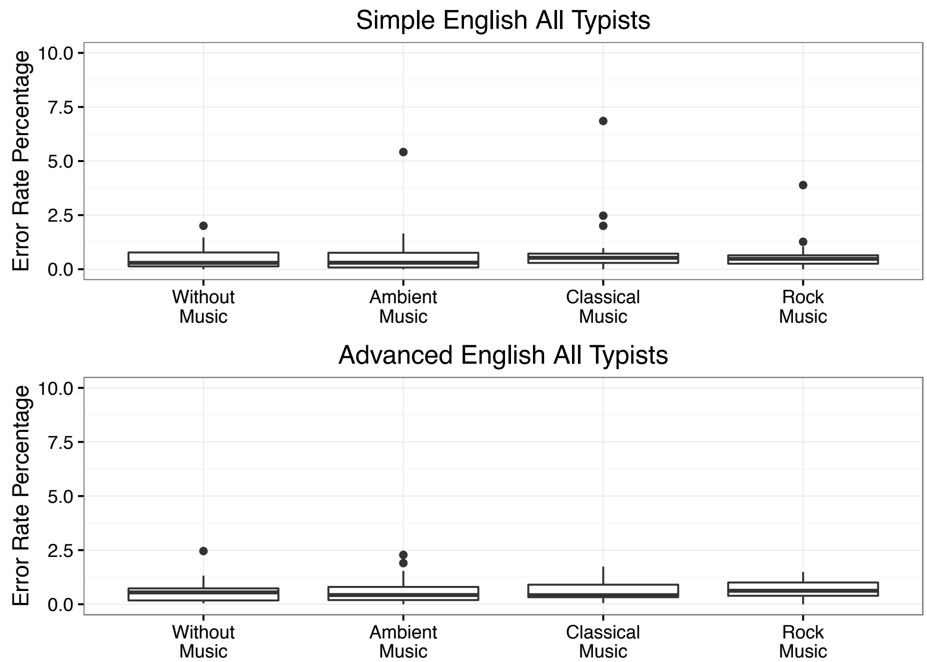


CPM box plots for simple and advanced English tasks, for fast typists

Fast Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Simple English	Without	$W(10) = 0.965$, *n.s.*	-0.20	-0.30	-0.78	-0.59
	Ambient	$W(10) = 0.955$, *n.s.*	-0.37	-0.54	-0.88	-0.66
	Classical	$W(10) = 0.978$, *n.s.*	0.09	0.13	-0.40	-0.30
	Rock	$W(10) = 0.979$, *n.s.*	0.06	0.08	0.31	0.24
Advanced English	Without	$W(10) = 0.909$, *n.s.*	0.11	0.16	-1.25	-0.94
	Ambient	$W(10) = 0.934$, *n.s.*	0.01	0.02	-1.49	-1.12
	Classical	$W(10) = 0.958$, *n.s.*	-0.12	-0.17	-0.66	-0.49
	Rock	$W(10) = 0.941$, *n.s.*	-0.07	-0.10	-1.42	-1.06

CPM Shapiro-Wilk's tests, skewness and kurtosis

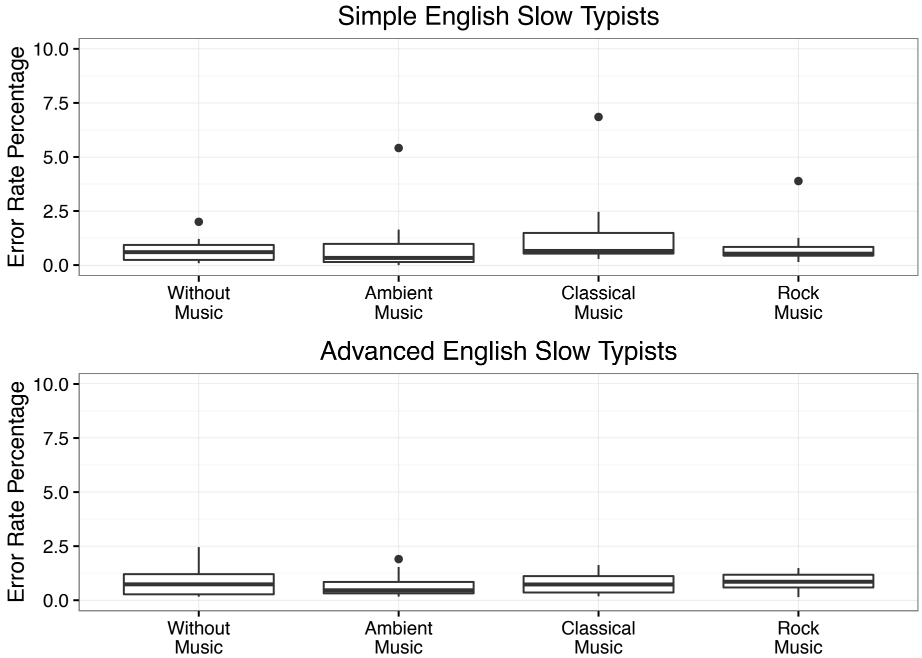
Error Rate



Error rate box plots for simple and advanced English tasks, for all typists

All Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Simple English	Without	$W(21) = 0.855, p = 0.005$	1.35	**2.70***	1.55	1.59
	Ambient	$W(21) = 0.531, p < 0.0005$	3.66	**7.31***	14.87	**15.30***
	Classical	$W(21) = 0.537, p < 0.0005$	3.54	**7.06***	13.79	**14.19***
	Rock	$W(21) = 0.604, p < 0.0005$	3.42	**6.82***	13.50	**13.89***
Advanced English	Without	$W(21) = 0.826, p = 0.002$	1.82	**3.64***	4.36	**4.48***
	Ambient	$W(21) = 0.815, p = 0.001$	1.60	**3.20***	2.12	**2.18***
	Classical	$W(21) = 0.871, p = 0.010$	0.99	1.98	-0.13	-0.13
	Rock	$W(21) = 0.975$, *n.s.*	0.19	0.37	-0.80	-0.82

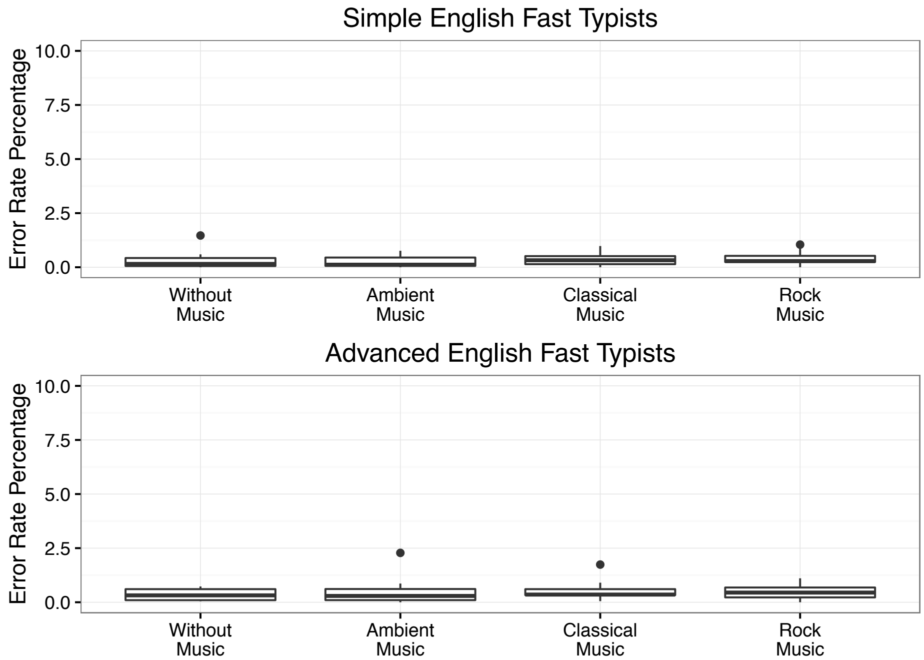
Error rate Shapiro-Wilk's tests, skewness and kurtosis for all typists
(* values that exceed acceptable range for skewness or kurtosis of population)



Error rate box plots for simple and advanced English tasks, for slow typists

Slow Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Simple English	Without	$W(11) = 0.905$, *n.s.*	1.17	1.77	1.63	1.27
	Ambient	$W(11) = 0.636$, $p < 0.0005$	2.70	**4.08***	7.89	**6.17***
	Classical	$W(11) = 0.617$, $p < 0.0005$	2.64	**3.99***	7.40	**5.78***
	Rock	$W(11) = 0.625$, $p < 0.0005$	2.77	**4.19***	8.26	**6.45***
Advanced English	Without	$W(11) = 0.867$, *n.s.*	1.31	1.98	2.03	1.59
	Ambient	$W(11) = 0.847$, $p = 0.038$	1.30	1.96	0.93	0.72
	Classical	$W(11) = 0.896$, *n.s.*	0.56	0.85	-1.12	-0.88
	Rock	$W(11) = 0.983$, *n.s.*	-0.23	-0.35	-0.56	-0.44

Error rate Shapiro-Wilk's tests, skewness and kurtosis for slow typists
(* values that exceed acceptable range for skewness or kurtosis of population)

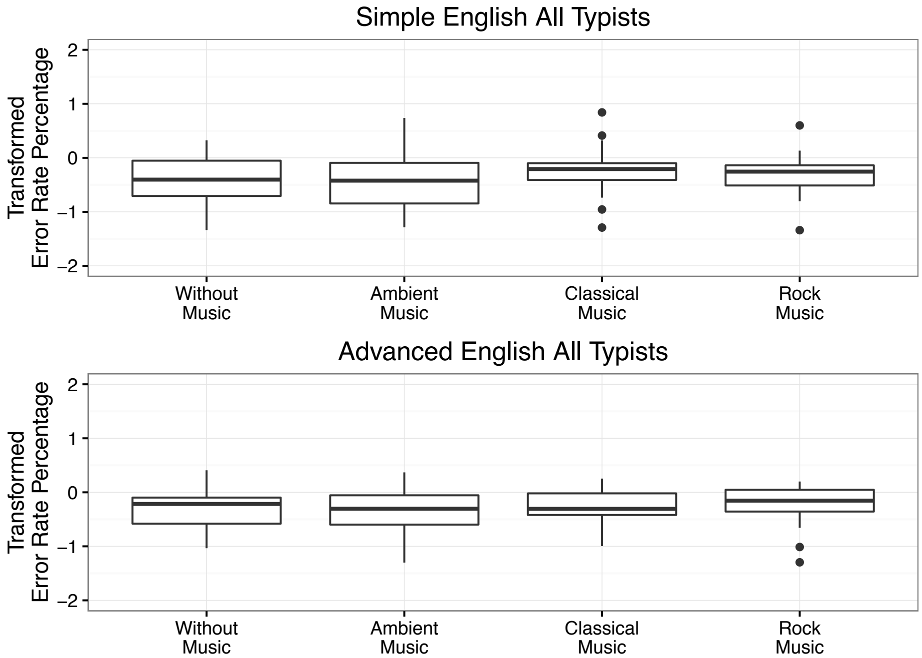


Error rate box plots for simple and advanced English tasks, for fast typists

Fast Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Simple English	Without	$W(10) = 0.731, p = 0.002$	2.16	**3.15***	5.07	**3.80***
	Ambient	$W(10) = 0.848$, *n.s.*	0.91	1.33	-0.68	-0.51
	Classical	$W(10) = 0.934$, *n.s.*	0.86	1.25	0.54	0.40
	Rock	$W(10) = 0.886$, *n.s.*	1.03	1.50	0.22	0.16
Advanced English	Without	$W(10) = 0.882$, *n.s.*	0.09	0.13	-1.82	-1.36
	Ambient	$W(10) = 0.727, p = 0.002$	2.26	**3.29***	5.72	**4.29***
	Classical	$W(10) = 0.805, p = 0.017$	1.90	**2.76***	4.18	3.13*
	Rock	$W(10) = 0.967$, *n.s.*	0.45	0.65	-0.17	-0.13

Error rate Shapiro-Wilk's tests, skewness and kurtosis for fast typists
(* values that exceed acceptable range for skewness or kurtosis of population)

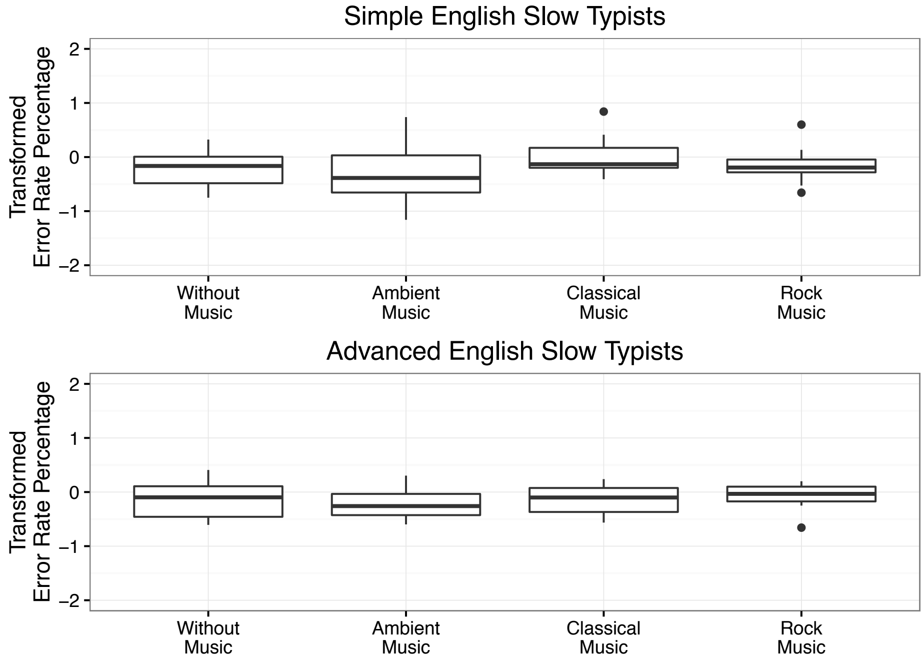
Transformed Error Rate



Transformed error rate box plots for simple and advanced English tasks, for all typists

All Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Simple English	Without	$W(21) = 0.965$, *n.s.*	-0.35	-0.69	-0.65	-0.67
	Ambient	$W(21) = 0.967$, *n.s.*	0.41	0.82	-0.27	-0.28
	Classical	$W(21) = 0.931$, *n.s.*	0.12	0.23	-0.49	-0.50
	Rock	$W(21) = 0.959$, *n.s.*	-0.15	-0.29	-0.78	-0.80
Advanced English	Without	$W(21) = 0.954$, *n.s.*	-0.32	-0.64	-0.40	-0.41
	Ambient	$W(21) = 0.978$, *n.s.*	-0.34	-0.68	0.11	0.11
	Classical	$W(21) = 0.959$, *n.s.*	-0.29	-0.57	0.07	0.07
	Rock	$W(21) = 0.920$, *n.s.*	-0.50	-1.00	-0.89	-0.92

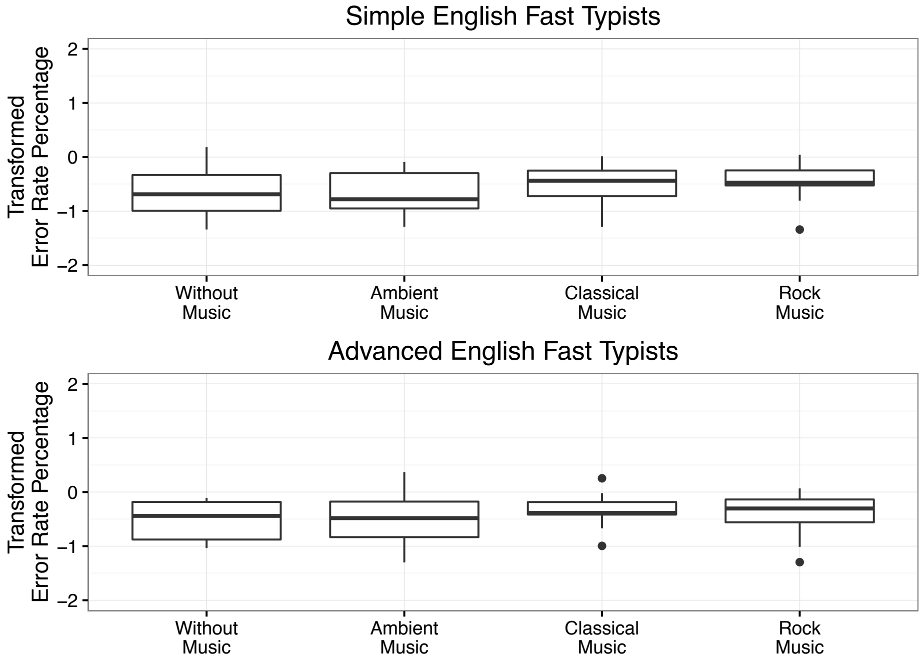
Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis with outlier capping



Transformed error rate box plots for simple and advanced English tasks, for slow typists

Slow Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Simple English	Without	$W(11) = 0.951$, *n.s.*	-0.27	-0.41	-0.95	-0.74
	Ambient	$W(11) = 0.967$, *n.s.*	0.10	0.15	-0.43	-0.34
	Classical	$W(11) = 0.875$, *n.s.*	0.67	1.02	-0.96	-0.75
	Rock	$W(11) = 0.920$, *n.s.*	-0.01	-0.01	-0.66	-0.51
Advanced English	Without	$W(11) = 0.931$, *n.s.*	-0.04	-0.06	-1.08	-0.85
	Ambient	$W(11) = 0.944$, *n.s.*	0.34	0.51	-0.82	-0.64
	Classical	$W(11) = 0.931$, *n.s.*	-0.02	-0.03	-1.50	-1.17
	Rock	$W(11) = 0.928$, *n.s.*	-0.11	-0.16	-1.52	-1.19

Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis with outlier capping for slow typists



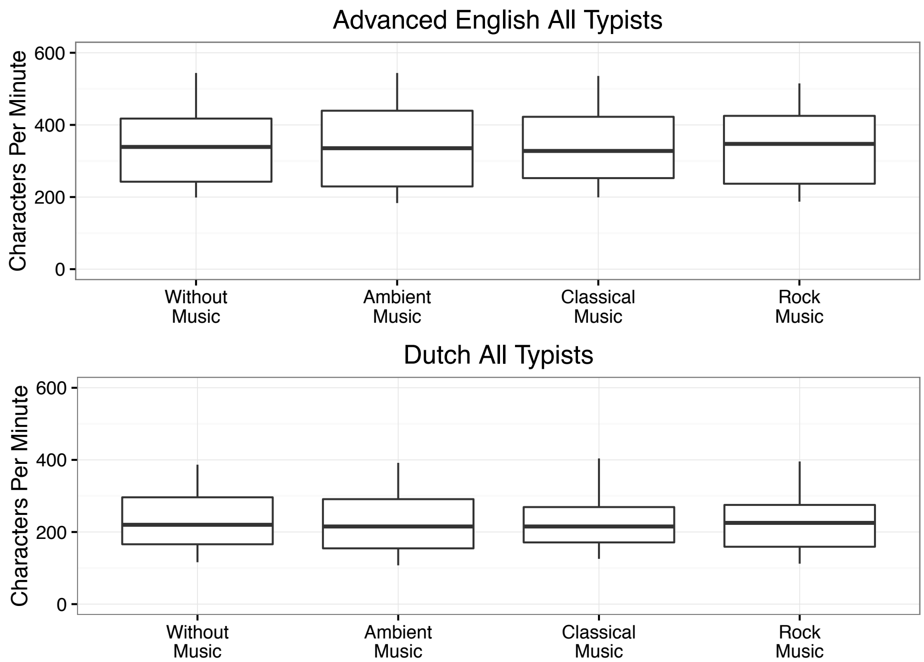
Transformed error rate box plots for simple and advanced English tasks, for fast typists

Fast Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	z_s	**k**	z_k
Simple English	Without	$W(10) = 0.969$, *n.s.*	0.25	0.36	-0.58	-0.43
	Ambient	$W(10) = 0.938$, *n.s.*	0.12	0.17	-1.33	-1.00
	Classical	$W(10) = 0.955$, *n.s.*	-0.68	-0.99	-0.06	-0.05
	Rock	$W(10) = 0.93$, *n.s.*	0.27	0.39	-0.61	-0.46
Advanced English	Without	$W(10) = 0.842$, $p = 0.047$	-0.51	-0.75	-1.58	-1.18
	Ambient	$W(10) = 0.981$, *n.s.*	0.08	0.12	-0.23	-0.17
	Classical	$W(10) = 0.914$, *n.s.*	-0.09	-0.13	-0.87	-0.65
	Rock	$W(10) = 0.901$, *n.s.*	-0.77	-1.12	-0.44	-0.33

Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis with outlier capping

E.1.3 Analysis Paradigm 3

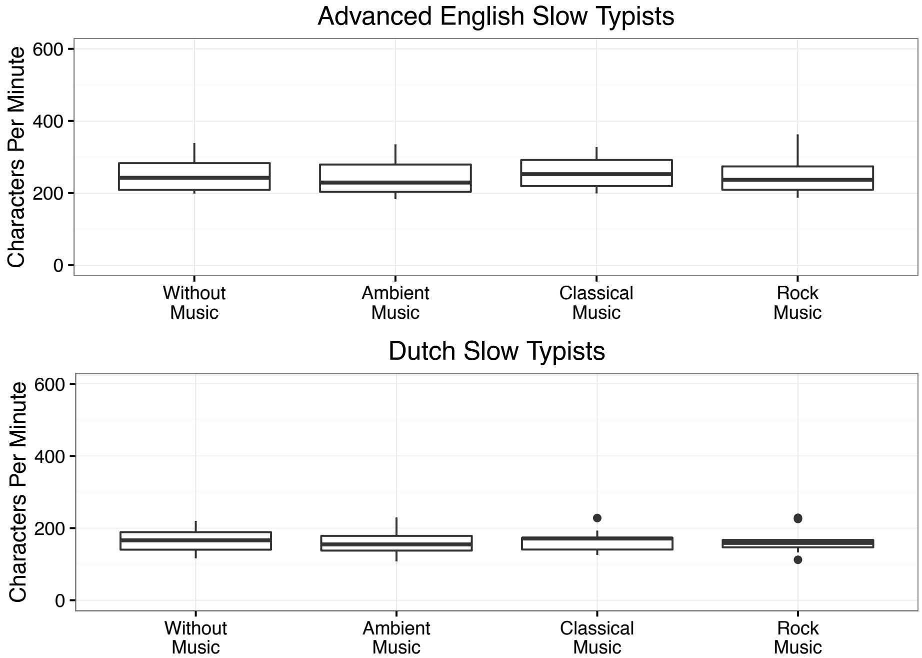
Characters Per Minute



CPM box plots for advanced English and Dutch tasks, for all typists

All Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Advanced English	Without	*W*(21) = 0.920, *n.s.*	0.36	0.72	-1.12	-1.15
	Ambient	*W*(21) = 0.919, *n.s.*	0.36	0.72	-1.20	-1.24
	Classical	*W*(21) = 0.951, *n.s.*	0.36	0.73	-0.89	-0.91
	Rock	*W*(21) = 0.932, *n.s.*	0.17	0.34	-1.35	-1.39
Dutch	Without	*W*(21) = 0.933, *n.s.*	0.53	1.07	-0.67	-0.69
	Ambient	*W*(21) = 0.926, *n.s.*	0.63	1.25	-0.72	-0.74
	Classical	*W*(21) = 0.909, *n.s.*	0.79	1.58	-0.57	-0.58
	Rock	*W*(21) = 0.925, *n.s.*	0.70	1.39	-0.73	-0.75

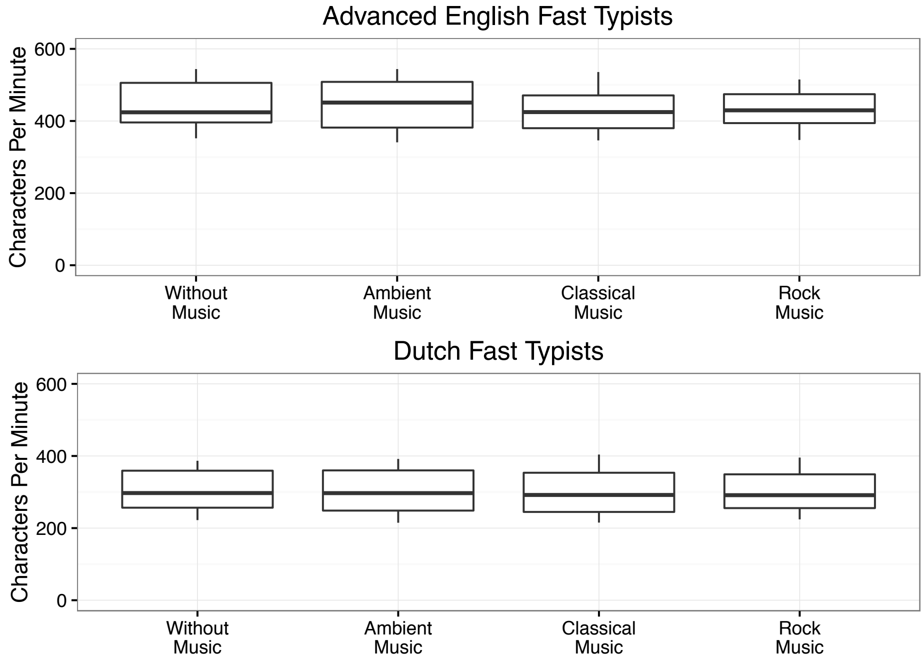
CPM Shapiro-Wilk's tests, skewness and kurtosis for all typists



CPM box plots for advanced English and Dutch tasks, for slow typists

Slow Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	$\mathbf{z_s}$	**k**	$\mathbf{z_k}$
Advanced English	Without	$W(11) = 0.890$, *n.s.*	0.70	1.06	-0.84	-0.66
	Ambient	$W(11) = 0.898$, *n.s.*	0.78	1.18	-0.67	-0.52
	Classical	$W(11) = 0.940$, *n.s.*	0.19	0.29	-1.34	-1.05
	Rock	$W(11) = 0.915$, *n.s.*	1.01	1.52	0.57	0.44
Dutch	Without	$W(11) = 0.943$, *n.s.*	0.08	0.12	-1.30	-1.01
	Ambient	$W(11) = 0.942$, *n.s.*	0.65	0.98	-0.32	-0.25
	Classical	$W(11) = 0.914$, *n.s.*	-0.01	-0.01	-1.37	-1.07
	Rock	$W(11) = 0.831$, $p = 0.024$	-0.80	-1.22	-0.83	-0.65

CPM Shapiro-Wilk's tests, skewness and kurtosis for slow typists

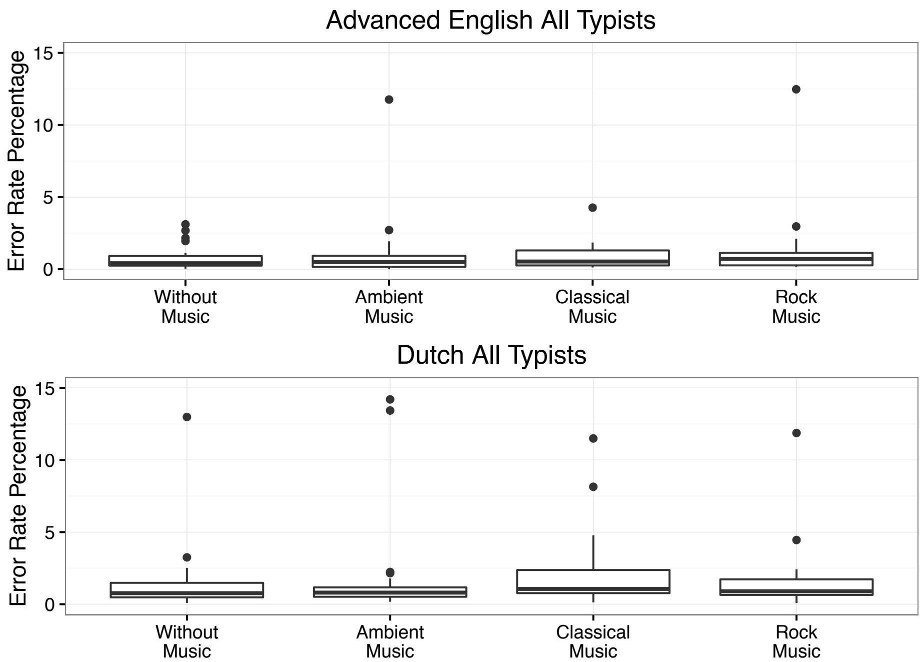


CPM box plots for advanced English and Dutch tasks for fast typists

Fast Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Advanced English	Without	$W(10) = 0.911$, *n.s.*	0.26	0.38	-1.57	-1.18
	Ambient	$W(10) = 0.928$, *n.s.*	-0.02	-0.02	-1.56	-1.17
	Classical	$W(10) = 0.939$, *n.s.*	0.49	0.71	-0.89	-0.67
	Rock	$W(10) = 0.961$, *n.s.*	0.00	0.00	-1.19	-0.89
Dutch	Without	$W(10) = 0.895$, *n.s.*	0.36	0.52	-1.39	-1.04
	Ambient	$W(10) = 0.918$, *n.s.*	0.21	0.31	-1.42	-1.07
	Classical	$W(10) = 0.931$, *n.s.*	0.32	0.46	-1.42	-1.06
	Rock	$W(10) = 0.924$, *n.s.*	0.43	0.63	-1.30	-0.97

CPM Shapiro-Wilk's tests, skewness and kurtosis with outlier capping

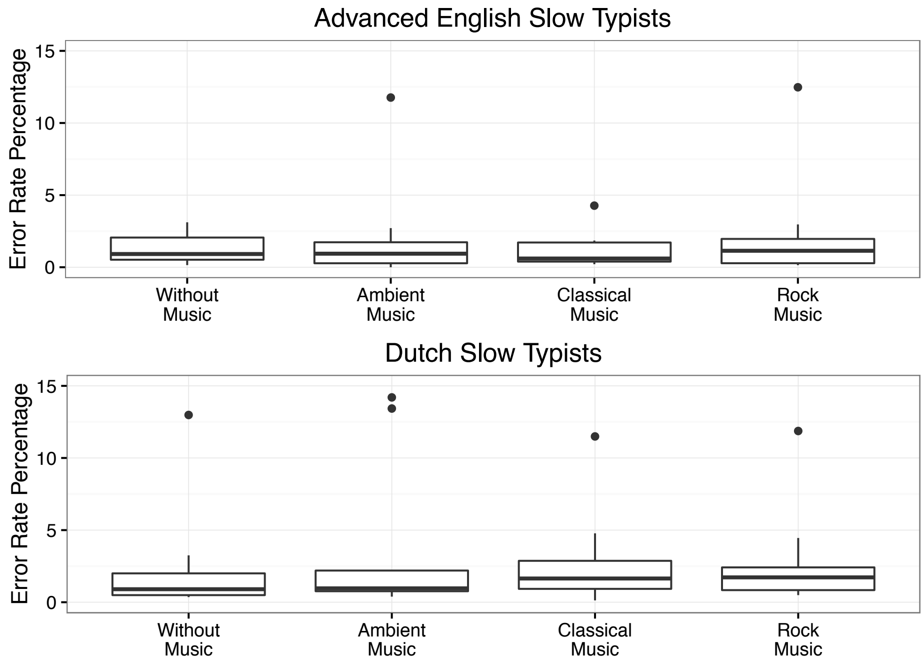
Error Rate



Error rate box plots for advanced English and Dutch tasks for all typists

All Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	z_s	**k**	z_k
Advanced English	Without	$W(21) = 0.782, p < 0.0005$	1.51	**3.02***	1.30	1.34
	Ambient	$W(21) = 0.445, p < 0.0005$	4.06	**8.11***	17.50	**18.00***
	Classical	$W(21) = 0.729, p < 0.0005$	2.45	**4.89***	7.50	**7.72***
	Rock	$W(21) = 0.453, p < 0.0005$	4.04	**8.05***	17.35	**17.84***
Dutch	Without	$W(21) = 0.473, p < 0.0005$	3.98	**7.94***	16.97	**17.46***
	Ambient	$W(21) = 0.468, p < 0.0005$	2.87	**5.73***	7.18	**7.38***
	Classical	$W(21) = 0.676, p < 0.0005$	2.42	**4.82***	5.87	**6.04***
	Rock	$W(21) = 0.535, p < 0.0005$	3.60	**7.18***	14.28	**14.69***

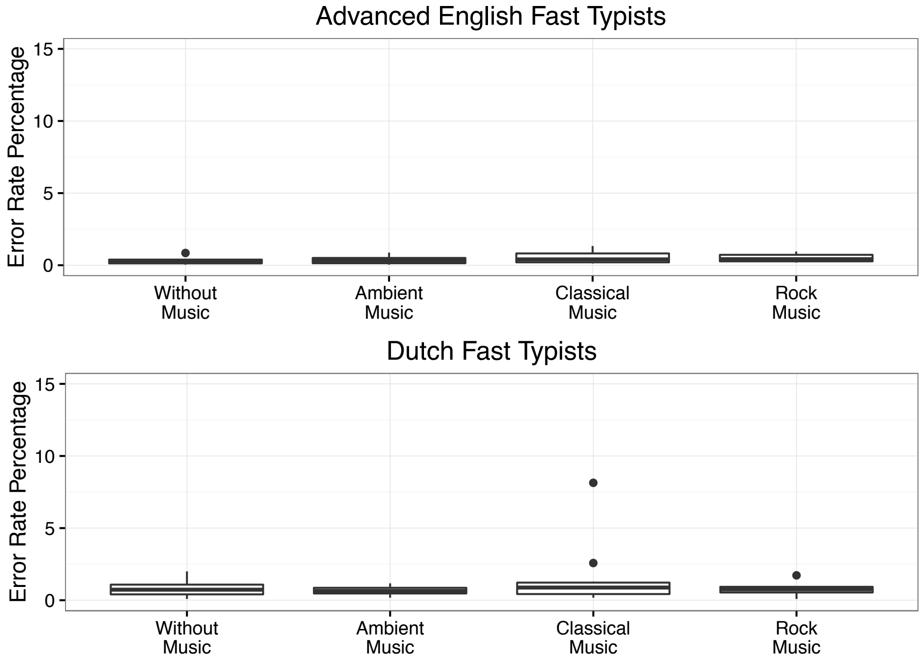
Error rate Shapiro-Wilk's tests, skewness and kurtosis for all typists
(* values that exceed acceptable range for skewness or kurtosis of population)



Error rate box plots for advanced English and Dutch tasks for slow typists

Slow Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Advanced English	Without	$W(11) = 0.901$, *n.s.*	0.68	1.02	-0.94	-0.74
	Ambient	$W(11) = 0.572, p < 0.0005$	2.96	**4.47***	9.20	**7.19***
	Classical	$W(11) = 0.765, p = 0.003$	1.94	**2.93***	4.25	**3.32***
	Rock	$W(11) = 0.577, p < 0.0005$	2.94	**4.44***	9.12	**7.13***
Dutch	Without	$W(11) = 0.554, p < 0.0005$	2.96	**4.48***	9.17	**7.17***
	Ambient	$W(11) = 0.598, p < 0.0005$	1.86	**2.82***	1.92	1.50
	Classical	$W(11) = 0.701, p < 0.0005$	2.43	**3.68***	6.48	**5.07***
	Rock	$W(11) = 0.635, p < 0.0005$	2.67	**4.04***	7.67	**6.00***

Error rate Shapiro-Wilk's tests, skewness and kurtosis for slow typists
(* values that exceed acceptable range for skewness or kurtosis of population)

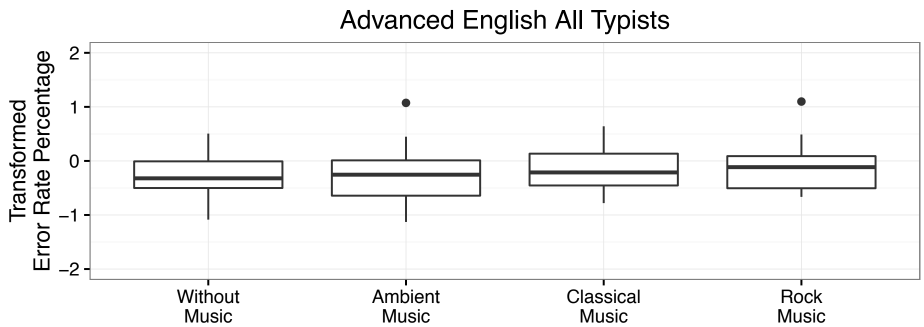


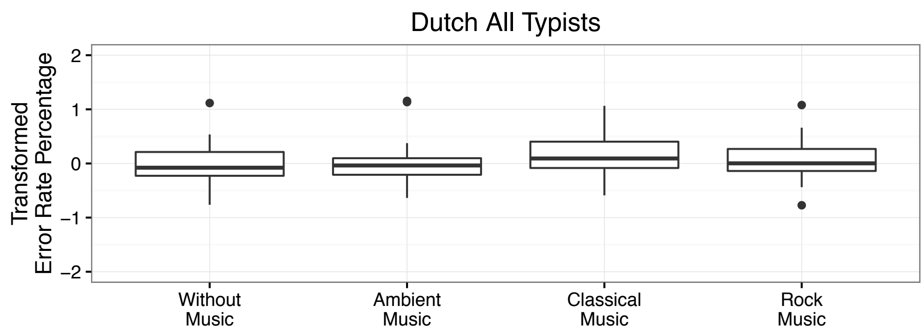
Error rate box plots for advanced English and Dutch tasks for fast typists

Fast Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	z_s	**k**	z_k
Advanced English	Without	$W(10) = 0.881$, *n.s.*	1.36	1.98	2.53	1.90
	Ambient	$W(10) = 0.929$, *n.s.*	0.64	0.93	-0.33	-0.25
	Classical	$W(10) = 0.856$, *n.s.*	0.80	1.16	-0.84	-0.63
	Rock	$W(10) = 0.893$, *n.s.*	0.41	0.60	-1.57	-1.18
Dutch	Without	$W(10) = 0.896$, *n.s.*	0.88	1.29	-0.19	-0.14
	Ambient	$W(10) = 0.947$, *n.s.*	0.09	0.13	-0.86	-0.64
	Classical	$W(10) = 0.609$, $p < 0.0005$	2.71	**3.94***	7.71	**5.78***
	Rock	$W(10) = 0.938$, *n.s.*	0.78	1.13	1.66	1.25

Error rate Shapiro-Wilk's tests, skewness and kurtosis
(* values that exceed acceptable range for skewness or kurtosis of population)

Transformed Error Rate

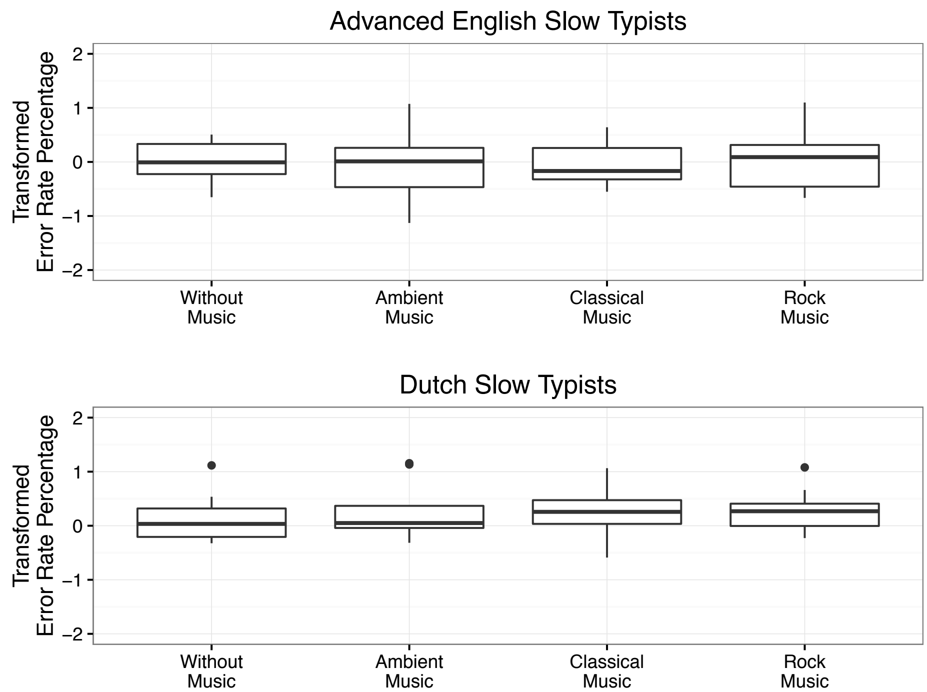




Transformed error rate box plots for advanced English and Dutch tasks for all typists

All Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	z_s	**k**	z_k
Advanced English	Without	$W(21) = 0.967$, *n.s.*	-0.01	-0.01	-0.58	-0.60
	Ambient	$W(21) = 0.963$, *n.s.*	-0.13	-0.25	-0.89	-0.92
	Classical	$W(21) = 0.975$, *n.s.*	0.33	0.66	-0.50	-0.51
	Rock	$W(21) = 0.929$, *n.s.*	0.27	0.54	-1.20	-1.24
Dutch	Without	$W(21) = 0.972$, *n.s.*	-0.08	-0.17	-0.13	-0.13
	Ambient	$W(21) = 0.947$, *n.s.*	-0.29	-0.58	-0.25	-0.26
	Classical	$W(21) = 0.968$, *n.s.*	0.29	0.57	-0.19	-0.19
	Rock	$W(21) = 0.951$, *n.s.*	0.40	0.79	-0.31	-0.31

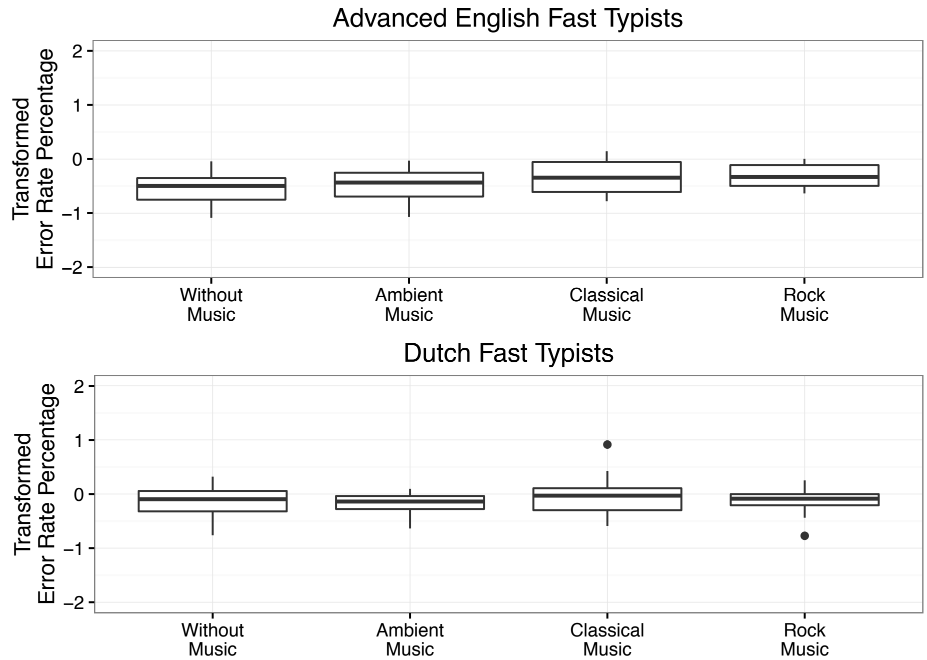
Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis with outlier capping



Transformed error rate box plots for advanced English and Dutch tasks for slow typists

Slow Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Advanced English	Without	$W(11) = 0.951$, *n.s.*	-0.33	-0.50	-0.95	-0.74
	Ambient	$W(11) = 0.967$, *n.s.*	-0.07	-0.10	-0.02	-0.01
	Classical	$W(11) = 0.945$, *n.s.*	0.46	0.69	-0.71	-0.56
	Rock	$W(11) = 0.932$, *n.s.*	0.36	0.54	-0.13	-0.10
Dutch	Without	$W(11) = 0.904$, *n.s.*	0.35	0.52	-1.45	-1.13
	Ambient	$W(11) = 0.875$, *n.s.*	-0.35	-0.53	-1.30	-1.02
	Classical	$W(11) = 0.979$, *n.s.*	-0.02	-0.03	1.03	0.81
	Rock	$W(11) = 0.942$, *n.s.*	0.11	0.17	-1.07	-0.84

Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis with outlier capping



Error rate box plots for advanced English and Dutch tasks for fast typists

Fast Typists						
Text	**Music**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Advanced English	Without	$W(10) = 0.935$, *n.s.*	-0.39	-0.57	-0.42	-0.32
	Ambient	$W(10) = 0.953$, *n.s.*	-0.43	-0.62	-0.86	-0.65
	Classical	$W(10) = 0.927$, *n.s.*	0.10	0.15	-1.57	-1.18
	Rock	$W(10) = 0.913$, *n.s.*	0.02	0.03	-1.78	-1.34
Dutch	Without	$W(10) = 0.962$, *n.s.*	-0.47	-0.68	0.20	0.15
	Ambient	$W(10) = 0.913$, *n.s.*	-0.78	-1.14	-0.22	-0.17
	Classical	$W(10) = 0.940$, *n.s.*	0.01	0.02	-0.96	-0.72
	Rock	$W(10) = 0.944$, *n.s.*	-0.12	-0.18	-0.12	-0.09

Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis with outlier capping

Appendix E: Experiment 2 Supplementary Statistical Analyses

Appendix F

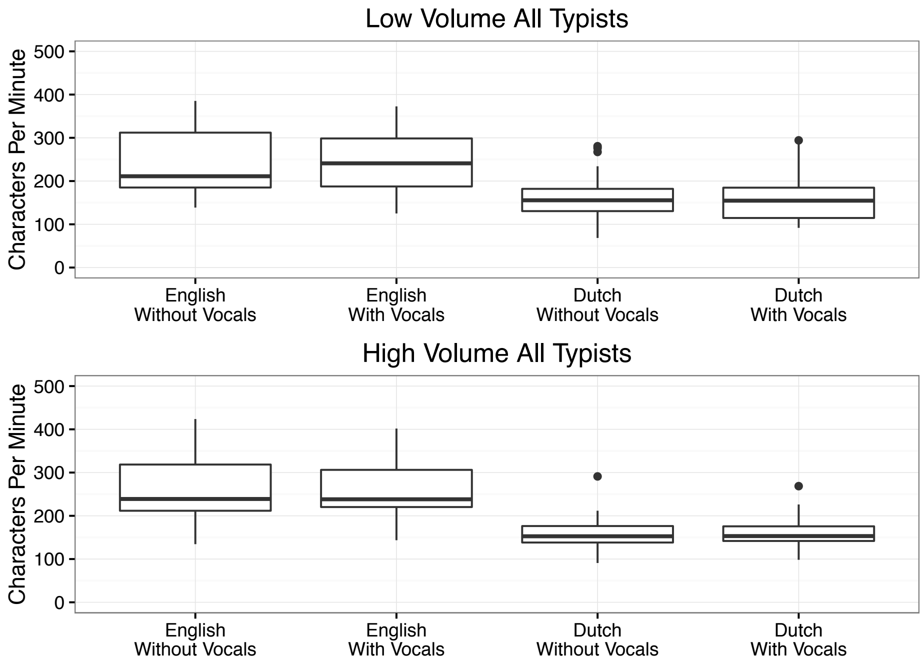
Experiment 3 Supplementary Statistical Analyses

This appendix includes:

- normality tests and box plots for CPM, error rate percentage and the transformed error rate percentage
- further descriptive statistics for task difficulty and music distraction ratings

F.1 Normality Tests and Box Plots

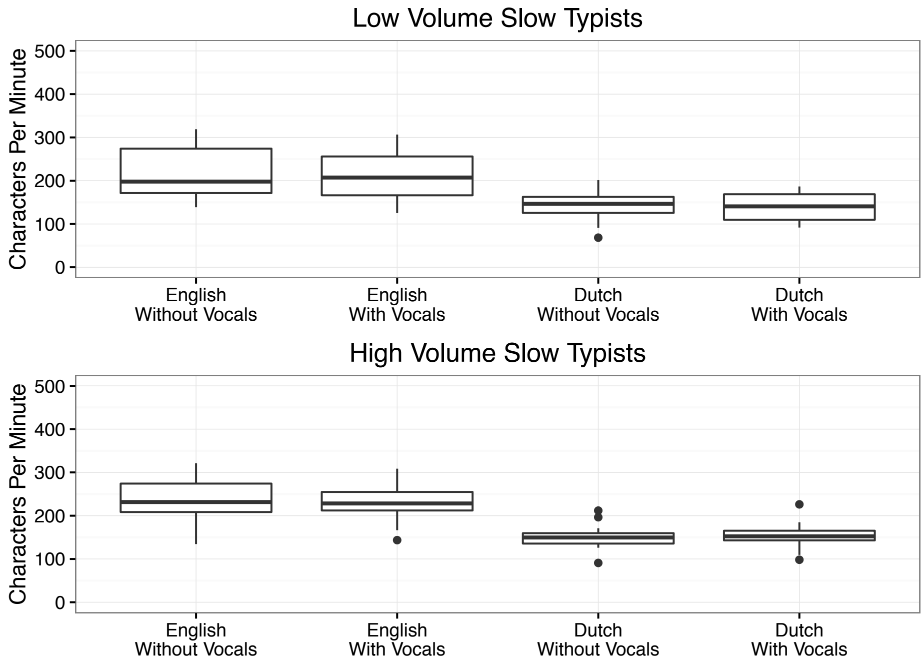
F.1.1 Characters Per Minute



CPM box plots for low and high volume groups, for all typists

Text	Vocals	Shapiro-Wilk's	Skewness		Kurtosis	
Low Volume All Typists						
			s	z_s	k	z_k
English	Without	$W(26) = 0.891$, $p = 0.010$	0.24	0.54	-1.12	-1.26
	With	$W(26) = 0.947$, *n.s.*	0.43	0.95	-1.31	-1.48
Dutch	Without	$W(26) = 0.939$, *n.s.*	0.91	2.00	0.05	0.05
	With	$W(26) = 0.900$, $p = 0.016$	0.72	1.59	0.14	0.16
High Volume All Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	z_s	k	z_k
English	Without	$W(27) = 0.947$, *n.s.*	0.41	0.91	-0.41	-0.47
	With	$W(27) = 0.961$, *n.s.*	0.40	0.89	-0.72	-0.82
Dutch	Without	$W(27) = 0.879$, $p = 0.005$	1.14	**2.54***	2.59	**2.96***
	With	$W(27) = 0.964$, *n.s.*	1.50	**3.35***	4.13	**4.73***

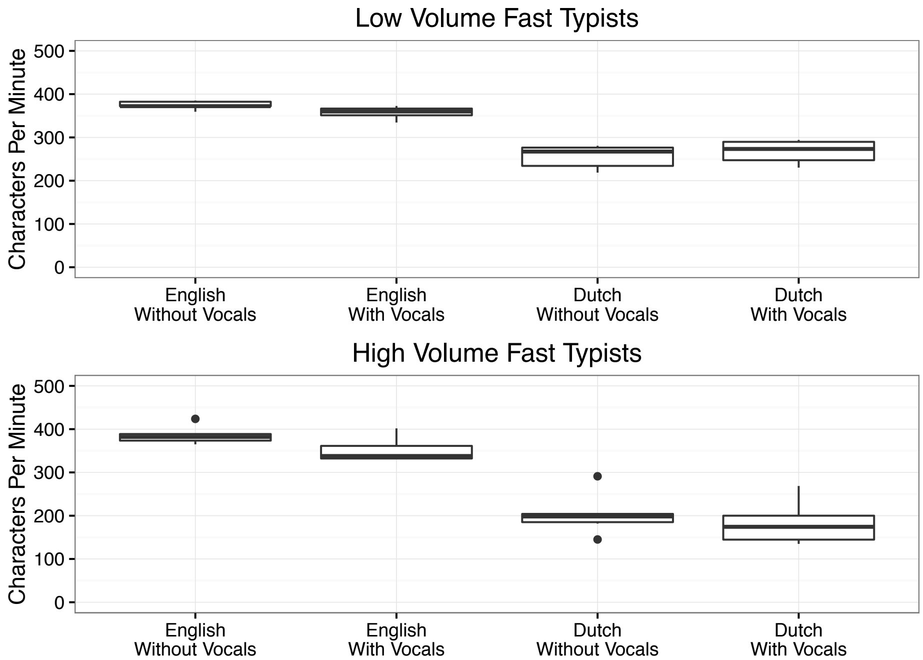
CPM Shapiro-Wilk's tests, skewness and kurtosis for all typists, high volume group
(* values that exceed acceptable range for skewness or kurtosis of population)



CPM box plots for low and high volume groups, for slow typists

Low Volume Slow Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(21) = 0.881, p = 0.015$	0.13	0.26	-1.18	-1.21
	With	$W(21) = 0.945$, *n.s.*	0.57	1.15	-1.20	-1.24
Dutch	Without	$W(21) = 0.986$, *n.s.*	-0.05	-0.10	-1.44	-1.48
	With	$W(21) = 0.926$, *n.s.*	-0.42	-0.83	0.13	0.14
High Volume Slow Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(21) = 0.967$, *n.s.*	-0.09	-0.18	-0.15	-0.15
	With	$W(21) = 0.963$, *n.s.*	-0.12	-0.24	-0.24	-0.25
Dutch	Without	$W(21) = 0.936$, *n.s.*	0.35	0.70	1.43	1.47
	With	$W(21) = 0.961$, *n.s.*	0.45	0.90	2.01	2.07

CPM Shapiro-Wilk's tests, skewness and kurtosis for slow typists, both volume groups

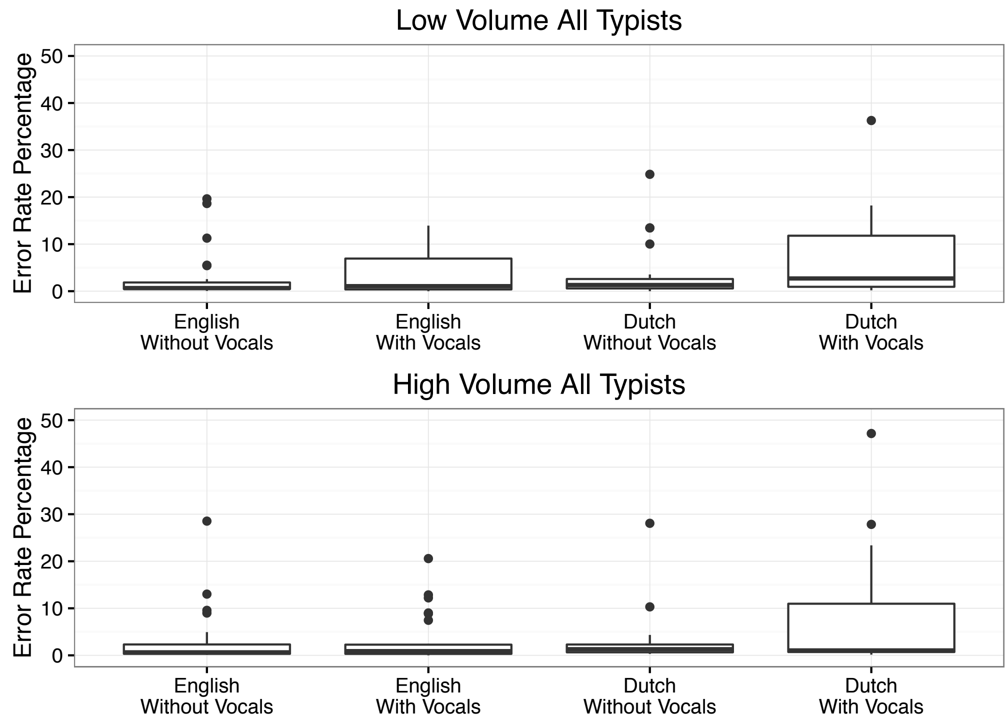


CPM box plots for low and high volume groups, for fast typists

Low Volume Fast Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(5) = 0.927$, *n.s.*	-0.86	-0.94	0.27	0.13
	With	$W(5) = 0.954$, *n.s.*	-0.58	-0.64	-0.15	-0.07
Dutch	Without	$W(5) = 0.874$, *n.s.*	-0.50	-0.54	-1.95	-0.97
	With	$W(5) = 0.917$, *n.s.*	-0.65	-0.72	-2.19	-1.09
High Volume Fast Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(6) = 0.884$, *n.s.*	1.47	1.73	1.32	0.76
	With	$W(6) = 0.790$, $p = 0.047$	1.45	1.72	2.61	1.50
Dutch	Without	$W(6) = 0.876$, *n.s.*	1.06	1.25	0.77	0.44
	With	$W(6) = 0.929$, *n.s.*	1.26	1.49	2.96	1.70

CPM Shapiro-Wilk's tests, skewness and kurtosis for fast typists, both volume groups

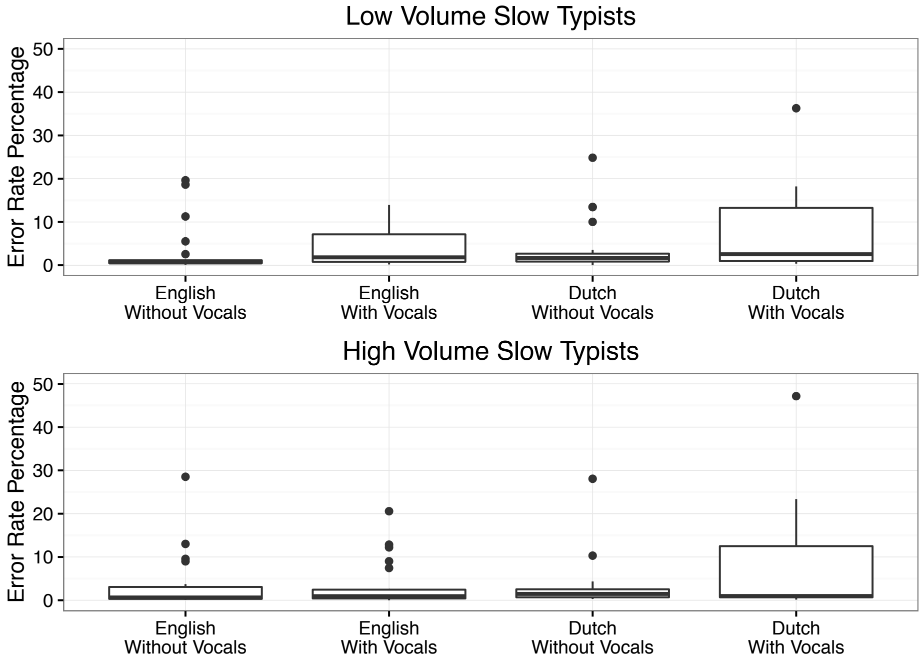
F.1.2 Error Rate



Error rate box plots for low and high volume groups, for all typists

Low Volume All Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(26) = 0.552, p < 0.0005$	2.52	**5.53***	5.59	**6.30***
	With	$W(26) = 0.776, p < 0.0005$	1.10	2.40	-0.37	-0.42
Dutch	Without	$W(26) = 0.588, p < 0.0005$	2.68	**5.87***	7.43	**8.37***
	With	$W(26) = 0.746, p < 0.0005$	1.97	**4.32***	4.62	**5.21***
High Volume All Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(27) = 0.530, p < 0.0005$	3.28	**7.32***	12.13	**13.91***
	With	$W(27) = 0.654, p < 0.0005$	2.08	**4.65***	4.06	**4.66***
Dutch	Without	$W(27) = 0.430, p < 0.0005$	4.25	**9.49***	19.47	**22.32***
	With	$W(27) = 0.642, p < 0.0005$	2.36	**5.27***	6.10	**7.00***

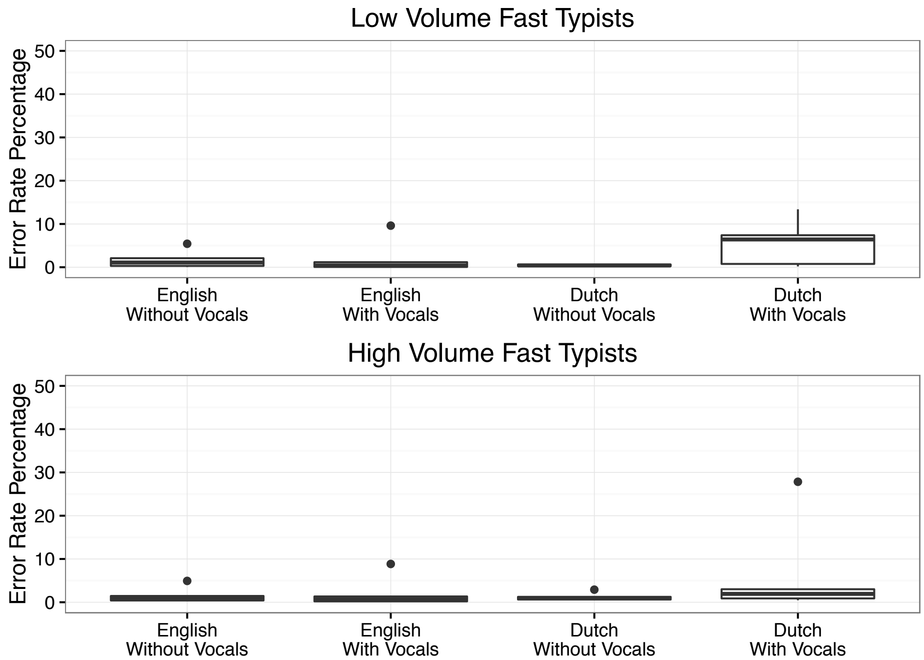
Error rate Shapiro-Wilk's tests, skewness and kurtosis for all typists, both volume groups
(* values that exceed acceptable range for skewness or kurtosis of population)



Error rate box plots for low and high volume groups, for slow typists

Low Volume Slow Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(21) = 0.548, p < 0.0005$	2.29	**4.57***	4.19	**4.31***
	With	$W(21) = 0.783, p < 0.0005$	1.02	2.03	-0.55	-0.57
Dutch	Without	$W(21) = 0.633, p < 0.0005$	2.37	**4.72***	5.64	**5.80***
	With	$W(21) = 0.726, p < 0.0005$	1.92	**3.83***	4.02	**4.13***
High Volume Slow Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(21) = 0.559, p < 0.0005$	2.90	**5.79***	9.30	**9.57***
	With	$W(21) = 0.671, p < 0.0005$	1.96	**3.91***	3.35	**3.44***
Dutch	Without	$W(21) = 0.469, p < 0.0005$	3.76	**7.51***	15.14	**15.57***
	With	$W(21) = 0.639, p < 0.0005$	2.48	**4.96***	7.14	**7.35***

Error rate Shapiro-Wilk's tests, skewness and kurtosis for slow typists, both volume groups
(* values that exceed acceptable range for skewness or kurtosis of population)

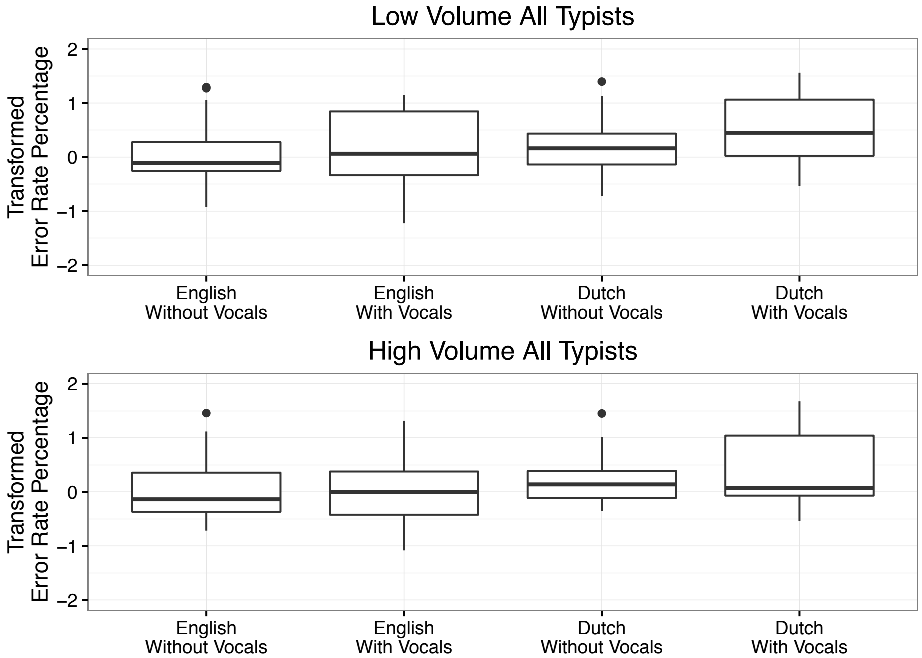


Error rate box plots for low and high volume groups, for fast typists

Low Volume Fast Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(5) = 0.841$, *n.s.*	1.57	1.72	2.44	1.22
	With	$W(5) = 0.646$, $p = 0.002$	2.17	2.37	4.74	2.37
Dutch	Without	$W(5) = 0.969$, *n.s.*	0.09	0.10	-1.73	-0.87
	With	$W(5) = 0.917$, *n.s.*	0.52	0.57	-0.61	-0.31
High Volume Fast Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(6) = 0.744$, $p = 0.017$	2.02	2.39	4.27	2.45
	With	$W(6) = 0.635$, $p = 0.001$	2.31	**2.74***	5.46	**3.14***
Dutch	Without	$W(6) = 0.801$, *n.s.*	1.83	2.16	3.72	2.14
	With	$W(6) = 0.587$, $p < 0.0005$	2.40	**2.84***	5.79	**3.33***

Error rate Shapiro-Wilk's tests, skewness and kurtosis for fast typists, both volume groups
(* values that exceed acceptable range for skewness or kurtosis of population)

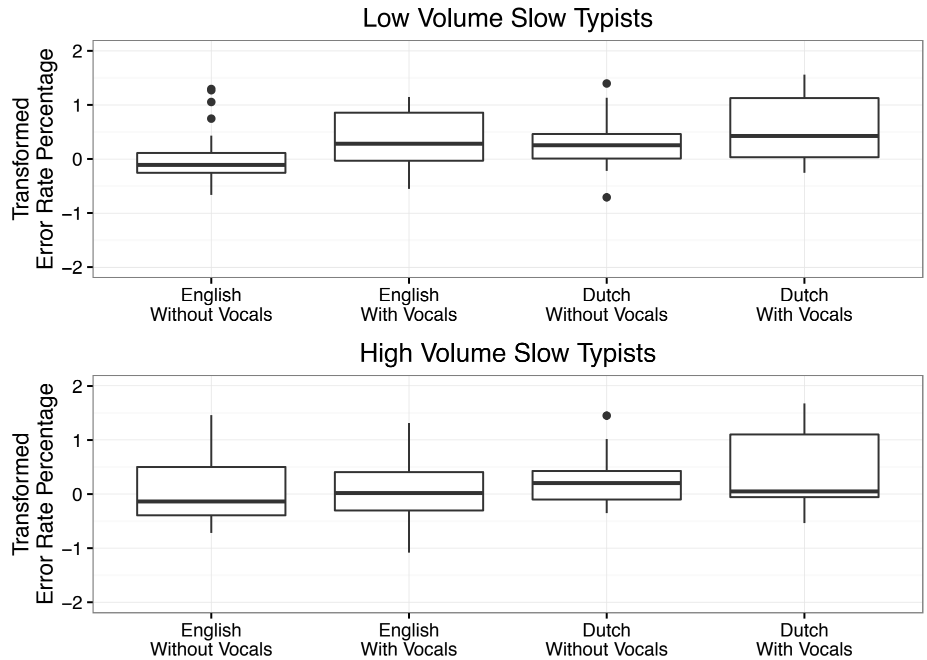
F.1.3 Transformed Error Rate



Transformed error rate box plots for low and high volume groups, for slow typists

Low Volume All Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(26) = 0.913$, $p = 0.031$	0.65	1.42	-0.02	-0.02
	With	$W(26) = 0.948$, *n.s.*	-0.16	-0.36	-0.77	-0.87
Dutch	Without	$W(26) = 0.950$, *n.s.*	0.35	0.78	-0.21	-0.24
	With	$W(26) = 0.947$, *n.s.*	0.13	0.28	-1.19	-1.34
High Volume All Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(27) = 0.889$, $p = 0.007$	0.83	1.84	-0.51	-0.58
	With	$W(27) = 0.955$, *n.s.*	0.22	0.50	-0.71	-0.81
Dutch	Without	$W(27) = 0.934$, *n.s.*	0.84	1.87	0.28	0.32
	With	$W(27) = 0.891$, $p = 0.008$	0.66	1.46	-0.95	-1.08

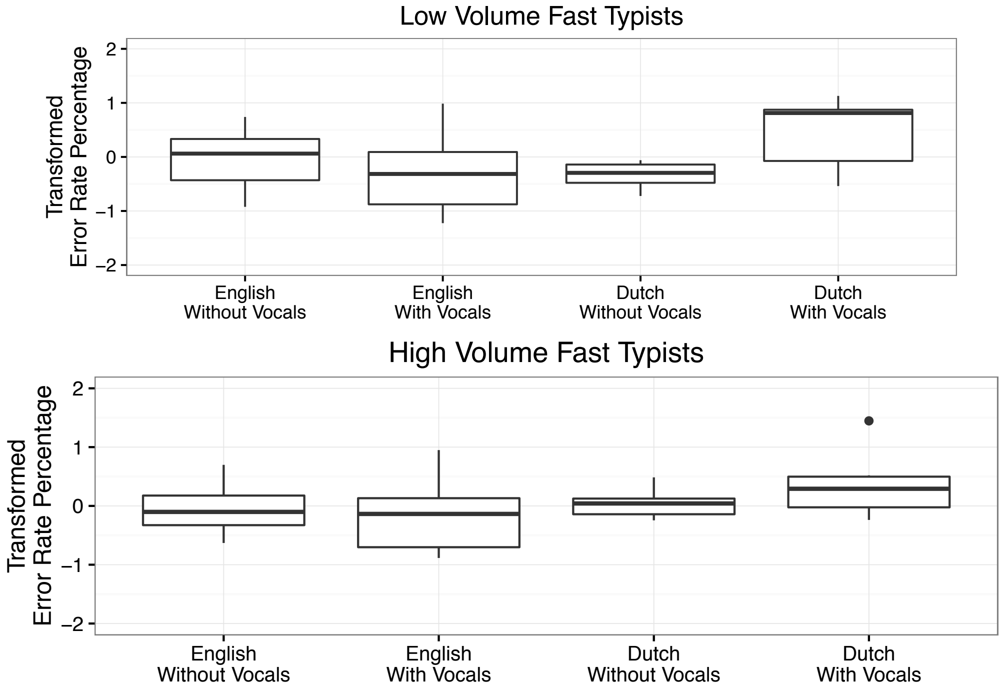
Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis for all typists, both volume groups with outlier capping



Transformed error rate box plots for low and high volume groups, for slow typists

Low Volume Slow Typists						
Text	Vocals	Shapiro-Wilk's	Skewness		Kurtosis	
			s	z_s	k	z_k
English	Without	$W(21) = 0.897, p = 0.031$	0.25	0.51	-0.53	-0.54
	With	$W(21) = 0.916$, *n.s.*	0.11	0.22	-1.40	-1.44
Dutch	Without	$W(21) = 0.894, p = 0.27$	0.71	1.42	-0.47	-0.48
	With	$W(21) = 0.915$, *n.s.*	0.40	0.79	-1.25	-1.29
High Volume Slow Typists						
Text	Vocals	Shapiro-Wilk's	Skewness		Kurtosis	
			s	z_s	k	z_k
English	Without	$W(21) = 0.865, p = 0.008$	0.84	1.67	-0.66	-0.68
	With	$W(21) = 0.960$, *n.s.*	0.17	0.35	-0.63	-0.65
Dutch	Without	$W(21) = 0.935$, *n.s.*	0.72	1.43	-0.03	-0.03
	With	$W(21) = 0.870, p = 0.009$	0.62	1.24	-1.12	-1.15

Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis for slow typists, both volume groups with outlier capping



Transformed error rate box plots for low and high volume groups, for fast typists

Low Volume Fast Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(5) = 0.984$, *n.s.*	-0.31	-0.34	-0.87	-0.44
	With	$W(5) = 0.970$, *n.s.*	0.59	0.65	-0.25	-0.13
Dutch	Without	$W(5) = 0.953$, *n.s.*	-0.64	-0.70	-0.75	-0.38
	With	$W(5) = 0.884$, *n.s.*	-0.71	-0.78	-1.81	-0.90
High Volume Fast Typists						
Text	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
English	Without	$W(5) = 0.967$, *n.s.*	0.53	0.63	-0.14	-0.08
	With	$W(5) = 0.923$, *n.s.*	0.51	0.60	-0.10	-0.06
Dutch	Without	$W(5) = 0.939$, *n.s.*	0.80	0.95	0.75	0.43
	With	$W(5) = 0.875$, *n.s.*	-0.49	-0.58	-1.89	-1.09

Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis for fast typists, both volume groups with outlier capping

F.2 Further Descriptive Statistics

F.2.1 Task Difficulty Ratings

Low Volume Group									
Text	**Vocals**	Slow Typists (n = 21)				Fast Typists (N = 5)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
English	Without	2.00	3.00	4.00	2.90	1.00	2.00	2.50	1.80
	With	3.00	3.00	4.00	3.24	2.00	3.00	3.00	2.60
Dutch	Without	6.00	6.00	6.00	5.62	3.00	3.00	4.50	3.60
	With	5.00	5.00	6.00	5.52	3.50	5.00	6.00	4.80
High Volume Group									
Text	**Vocals**	Slow Typists (n = 21)				Fast Typists (N = 6)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
English	Without	2.00	3.00	4.00	3.10	1.75	2.00	3.00	2.17
	With	2.00	3.00	4.50	3.38	2.75	3.00	4.25	3.33
Dutch	Without	3.00	6.00	6.00	4.95	3.00	5.00	7.00	5.00
	With	5.00	6.00	6.50	5.62	3.00	4.00	7.00	4.00

Descriptive statistics of task difficulty ratings with typing speed classification

F.2.2 Music Distraction Ratings

Low Volume Group									
Text	**Vocals**	Slow Typists (n = 21)				Fast Typists (N = 5)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
English	Without	1.00	2.00	3.00	2.00	1.00	1.00	3.50	2.00
	With	1.00	2.00	3.00	2.19	1.00	4.00	5.00	3.20
Dutch	Without	1.00	2.00	3.50	2.43	1.00	1.00	5.00	2.60
	With	1.00	2.00	3.00	2.19	2.00	2.00	4.00	2.80
High Volume Group									
Text	**Vocals**	Slow Typists (n = 21)				Fast Typists (N = 6)			
		Q1	*Mdn*	*Q3*	*M*	*Q1*	*Mdn*	*Q3*	*M*
English	Without	1.50	2.00	4.00	2.81	1.75	2.50	3.25	2.50
	With	1.50	3.00	5.00	3.33	1.75	3.00	5.00	3.17
Dutch	Without	1.00	2.00	4.50	2.86	1.00	2.50	5.50	3.17
	With	1.50	3.00	5.00	3.14	2.75	3.00	4.25	3.33

Descriptive statistics of music distraction ratings with speed classification

Appendix G

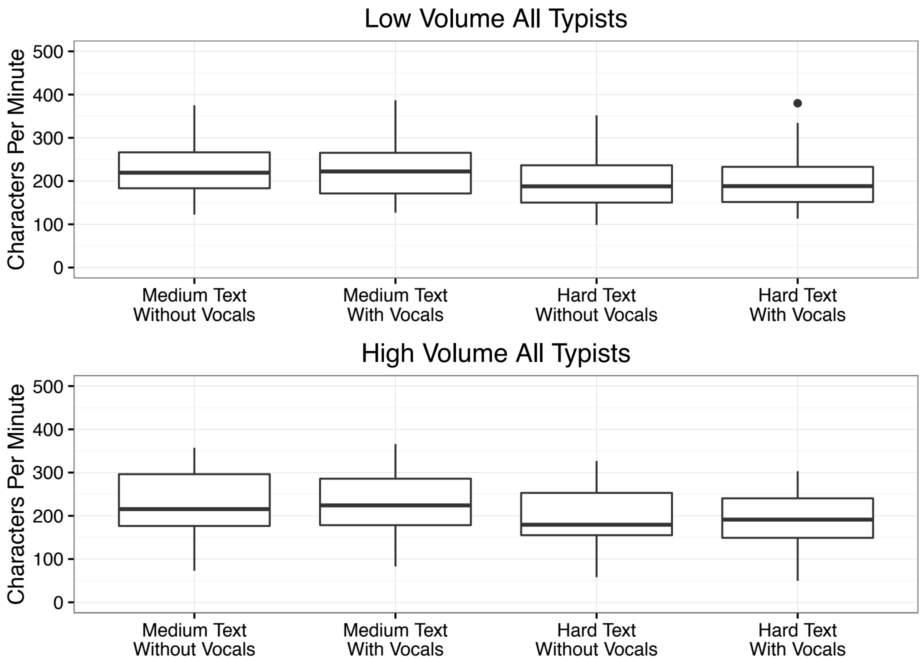
Experiment 4 Supplementary Statistical Analyses

This appendix includes:

- normality tests and box plots for CPM, error rate percentage and the transformed error rate percentage,
- further inferential analysis of transformed error rate and task difficulty ratings data with the speed classification applied.

G.1 Normality Tests and Box Plots

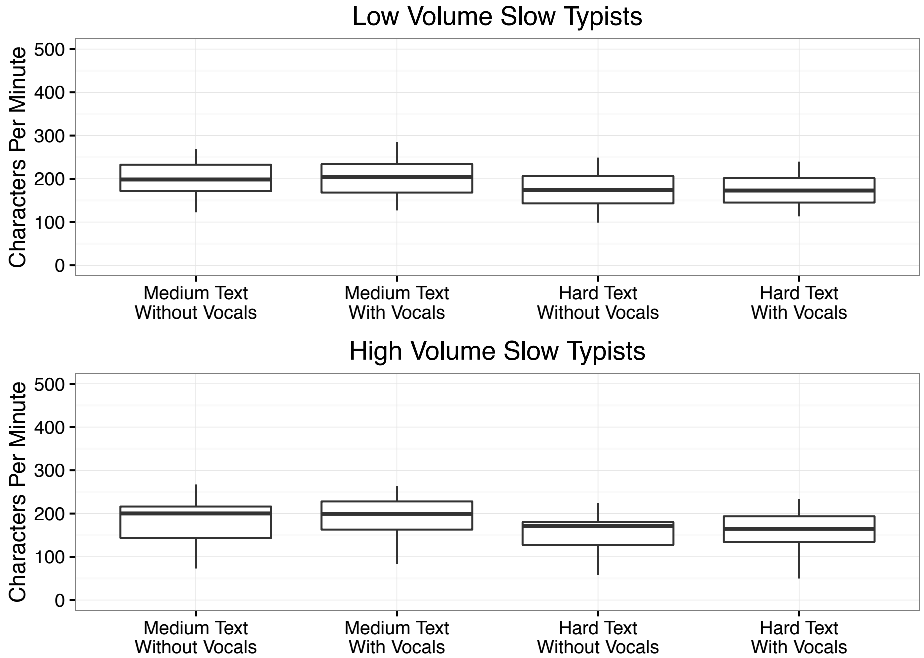
G.1.1 Characters Per Minute



CPM box plots for low and high volume groups, for all typists

Low Volume All Typists						
Difficulty	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	$\mathbf{z_s}$	**k**	$\mathbf{z_k}$
Medium	Without	$W(53) = 0.966$, *n.s.*	0.51	1.57	-0.44	-0.69
	With	$W(53) = 0.947$, $p = 0.021$	0.63	1.94	-0.28	-0.43
Hard	Without	$W(53) = 0.962$, *n.s.*	0.57	1.74	-0.26	-0.40
	With	$W(53) = 0.940$, $p = 0.010$	0.87	2.67	0.45	0.70
High Volume All Typists						
Difficulty	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	$\mathbf{z_s}$	**k**	$\mathbf{z_k}$
Medium	Without	$W(60) = 0.970$, *n.s.*	-0.03	-0.10	-0.90	-1.48
	With	$W(60) = 0.979$, *n.s.*	-0.02	-0.07	-0.82	-1.35
Hard	Without	$W(60) = 0.972$, *n.s.*	0.06	0.19	-0.82	-1.36
	With	$W(60) = 0.979$, *n.s.*	-0.08	-0.25	-0.69	-1.14

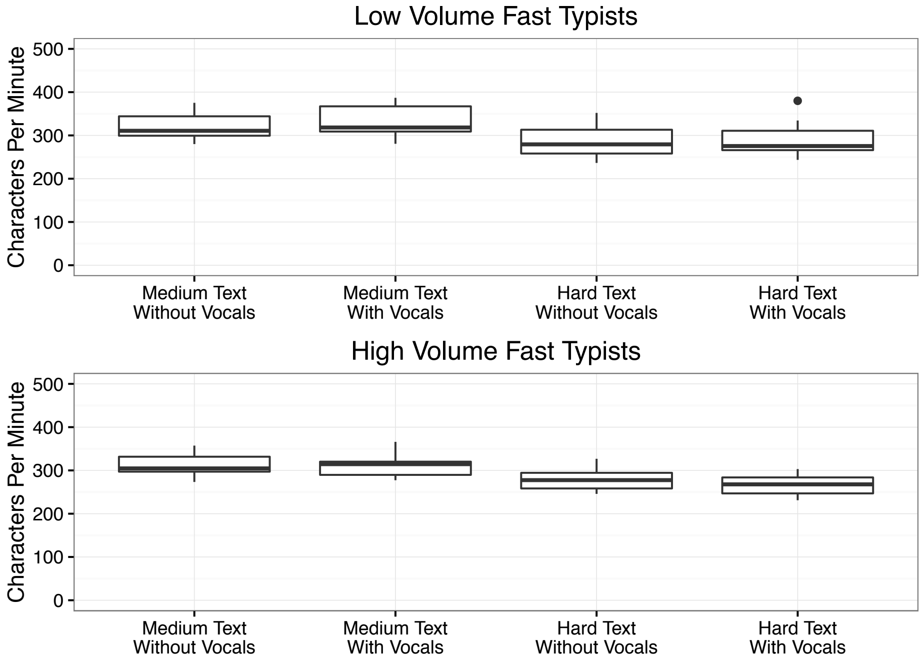
CPM Shapiro-Wilk's tests, skewness and kurtosis for all typists, both volume groups



CPM box plots for low and high volume groups, for slow typists

Low Volume Slow Typists						
Difficulty	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Medium	Without	$W(41) = 0.971$, *n.s.*	-0.03	-0.09	-0.93	-1.29
	With	$W(41) = 0.966$, *n.s.*	-0.01	-0.02	-1.04	-1.44
Hard	Without	$W(41) = 0.974$, *n.s.*	0.10	0.26	-0.87	-1.20
	With	$W(41) = 0.963$, *n.s.*	0.14	0.37	-1.06	-1.47
High Volume Slow Typists						
Difficulty	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Medium	Without	$W(41) = 0.961$, *n.s.*	-0.36	-0.98	-0.61	-0.84
	With	$W(41) = 0.959$, *n.s.*	-0.41	-1.11	-0.70	-0.97
Hard	Without	$W(41) = 0.950$, *n.s.*	-0.54	-1.47	-0.43	-0.59
	With	$W(41) = 0.971$, *n.s.*	-0.59	-1.59	0.01	0.01

CPM Shapiro-Wilk's tests, skewness and kurtosis for slow typists, both volume groups

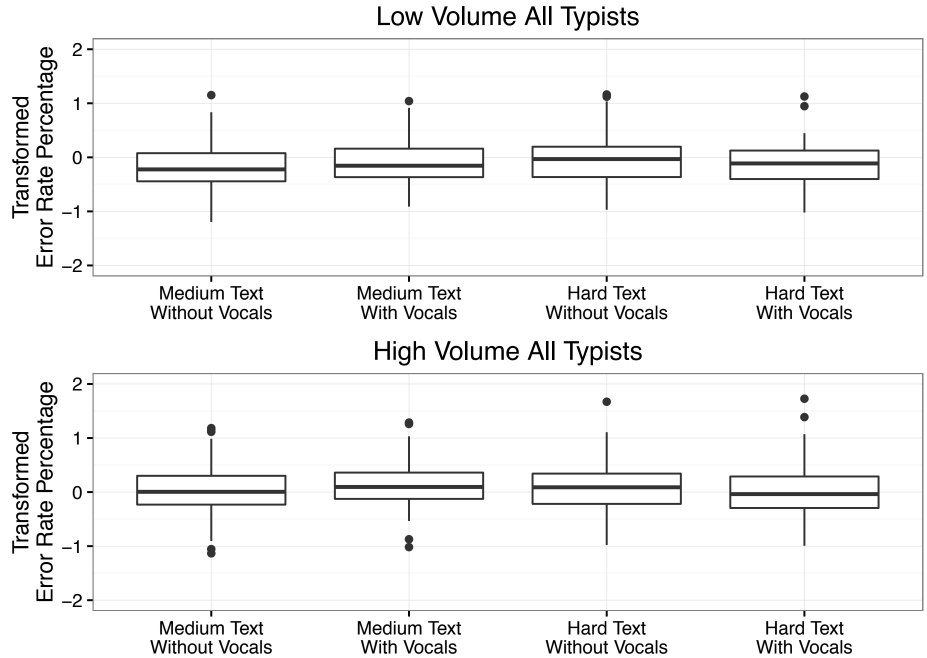


CPM box plots for low and high volume groups, for fast typists

Low Volume Fast Typists						
Difficulty	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Medium	Without	$W(12) = 0.938$, *n.s.*	0.38	0.60	-0.96	-0.78
	With	$W(12) = 0.901$, *n.s.*	0.39	0.61	-1.30	-1.06
Hard	Without	$W(12) = 0.934$, *n.s.*	0.51	0.80	-0.96	-0.78
	With	$W(12) = 0.902$, *n.s.*	1.20	1.88	1.24	1.01
High Volume Fast Typists						
Difficulty	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Medium	Without	$W(19) = 0.913$, *n.s.*	0.54	1.03	-0.79	-0.78
	With	$W(19) = 0.918$, *n.s.*	0.75	1.43	0.14	0.14
Hard	Without	$W(19) = 0.942$, *n.s.*	0.45	0.86	-0.55	-0.54
	With	$W(19) = 0.941$, *n.s.*	-0.02	-0.03	-1.23	-1.21

CPM Shapiro-Wilk's tests, skewness and kurtosis for fast typists, both volume groups

G.1.2 Transformed Error Rate



Transformed error rate box plots for low and high volume groups, for fast typists

Low Volume All Typists						
Difficulty	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Medium	Without	$W(53) = 0.980$, *n.s.*	0.01	0.04	-0.06	-0.09
	With	$W(53) = 0.938$, $p = 0.008$	0.86	**2.64***	0.86	1.34
Hard	Without	$W(53) = 0.934$, $p = 0.006$	0.58	1.79	0.33	0.51
	With	$W(53) = 0.974$, *n.s.*	-0.36	-1.10	-0.22	-0.34
High Volume All Typists						
Difficulty	**Vocals**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Medium	Without	$W(60) = 0.962$, *n.s.*	0.43	1.38	0.01	0.01
	With	$W(60) = 0.957$, $p = 0.033$	0.37	1.21	-0.42	-0.69
Hard	Without	$W(60) = 0.973$, *n.s.*	0.33	1.07	0.09	0.14
	With	$W(60) = 0.955$, $p = 0.026$	0.54	1.75	-0.07	-0.11

Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis for all typists with outlier capping
(* values that exceed acceptable range for skewness or kurtosis of population)

G.2 Further Inferential Analysis with Speed Classification

G.2.1 Low Volume Group

Transformed Error Rate

Test	F	df	p	η_p^2
Vocals	0.191	1, 51	n.s.	-
Vocals x Speed*	0.153	1, 51	n.s.	-
Text	0.600	1, 51	n.s.	-
Text x Speed*	0.113	1, 51	n.s.	-
Text x Vocals	4.288	1, 51	0.043	0.078
Text x Vocals x Speed*	0.004	1, 51	n.s.	-
Speed*	0.672	1, 51	n.s.	-

Inferential analysis of transformed error rate, low volume group with speed classification
(* between groups factor)

Task Difficulty

Test	F	df	p
Vocals	0.906	1, 51	n.s.
Vocals x Speed*	2.080	1, 51	n.s.
Text	65.635	1, 51	< 0.0005
Text x Speed*	0.088	1, 51	n.s.
Text x Vocals	0.111	1, 51	n.s.
Text x Vocals x Speed*	0.199	1, 51	n.s.
Speed*	2.148	1, 51	n.s.

Inferential analysis of task difficulty ratings, low volume group with speed classification
(* between groups factor)

G.2.2 High Volume Group

Transformed Error Rate

Test	F	df	p	η_p^2
Vocals	0.346	1, 58	n.s.	-
Vocals x Speed*	0.479	1, 58	n.s.	-
Text	0.059	1, 58	n.s.	-
Text x Speed*	0.009	1, 58	n.s.	-
Text x Vocals	4.040	1, 58	0.049	0.065
Text x Vocals x Speed*	1.865	1, 58	n.s.	-
Speed*	7.869	1, 58	0.007	0.119

Inferential analysis of transformed error rate, high volume group with speed classification
(* between groups factor)

Task Difficulty

Test	F	df	p
Vocals	1.380	1, 58	n.s.
Vocals x Speed*	0.986	1, 58	n.s.
Text	110.971	1, 58	< 0.0005
Text x Speed*	0.110	1, 58	n.s.
Text x Vocals	0.024	1, 58	n.s.
Text x Vocals x Speed*	0.897	1, 58	n.s.
Speed*	0.558	1, 58	n.s.

Inferential analysis of task difficulty ratings, high volume group with speed classification
(* between groups factor)

Appendix G: Experiment 4 Supplementary Statistical Analyses

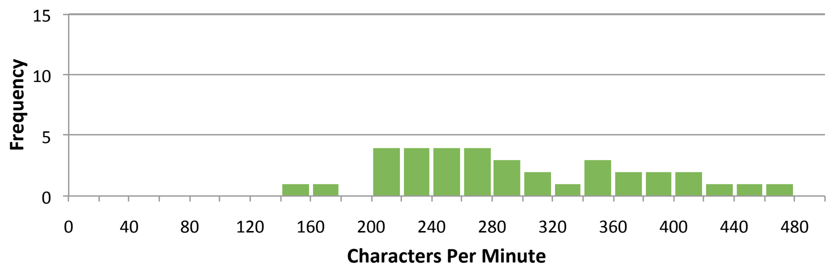
Appendix H

Experiment 5 Supplementary Statistical Analyses

This appendix includes:

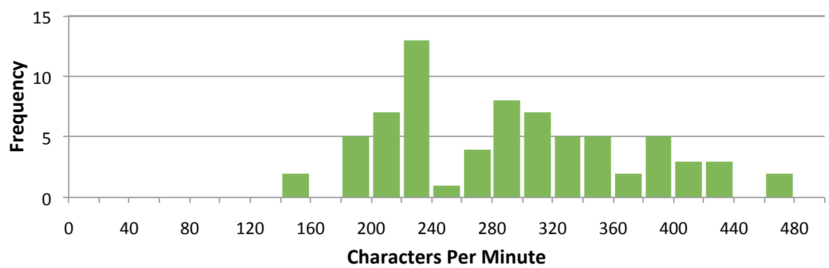
- histograms of CPM for slow, mid and fast tempo conditions,
- normality tests and box plots for CPM and the transformed error rate percentage,

H.1 Histograms of CPM



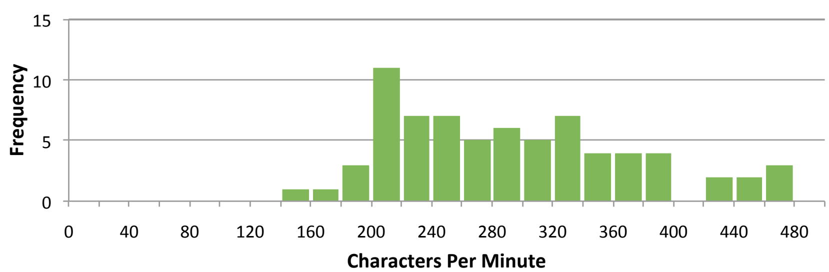
n = 36, *M* = 298.80, *SD* = 79.80

Histogram of CPM for without music condition



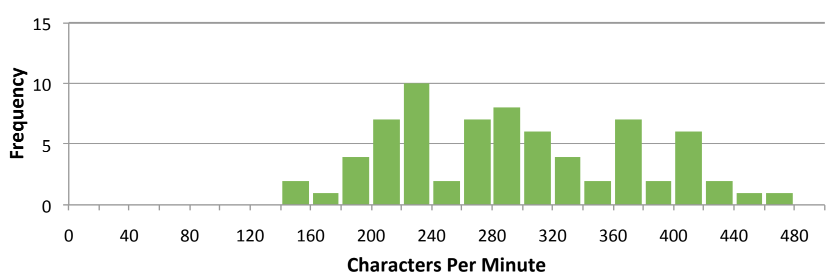
n = 72, *M* = 295.38, *SD* = 78.43

Histogram of CPM for slow tempo conditions



n = 72, *M* = 293.67, *SD* = 79.81

Histogram of CPM for mid tempo conditions

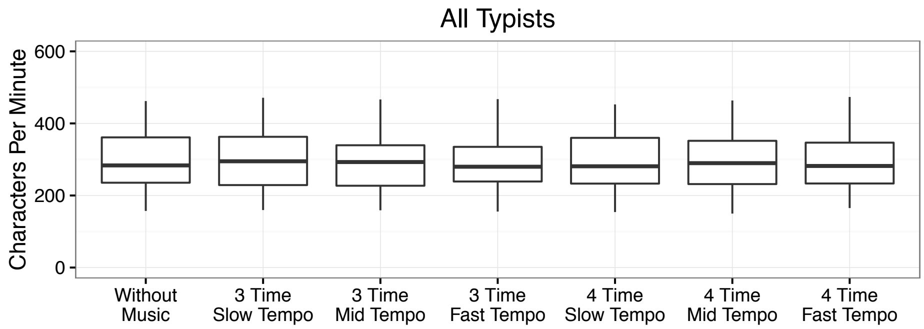


n = 71, *M* = 293.31, *SD* = 78.77

Histogram of CPM for fast tempo conditions

H.2 Normality Tests and Box Plots

H.2.1 Characters Per Minute

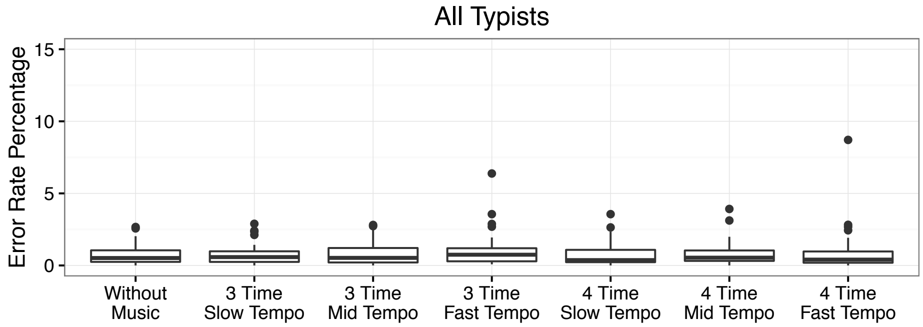


CPM box plots for all typists

All Typists						
Time Signature	**Tempo**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Without music		$W(36) = 0.964$, *n.s.*	0.35	0.90	-0.79	-1.03
3/4	Slow	$W(36) = 0.974$, *n.s.*	0.27	0.69	-0.70	-0.91
	Mid	$W(36) = 0.959$, *n.s.*	0.54	1.37	-0.41	-0.53
	Fast	$W(36) = 0.960$, *n.s.*	0.38	0.97	-0.76	-0.99
4/4	Slow	$W(36) = 0.964$, *n.s.*	0.31	0.79	-0.89	-1.16
	Mid	$W(36) = 0.950$, *n.s.*	0.56	1.42	-0.40	-0.52
	Fast	$W(36) = 0.966$, *n.s.*	0.30	0.76	-0.77	-1.00

CPM Shapiro-Wilk's tests, skewness and kurtosis for all typists

H.2.2 Error Rate

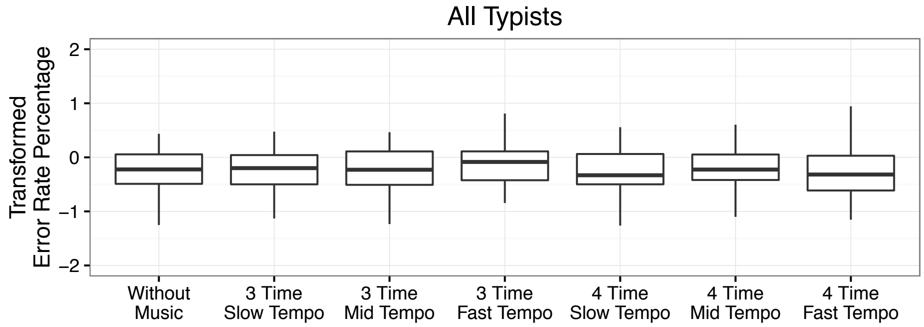


Error rate box plots for all typists

All Typists						
Time Signature	**Tempo**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Without music		*W(36) = 0.865, p < 0.0005*	1.29	**3.27***	1.22	1.59
3/4	Slow	*W(36) = 0.840, p < 0.0005*	1.49	**3.79***	1.80	2.35
	Mid	*W(36) = 0.849, p < 0.0005*	1.34	**3.40***	1.36	1.77
	Fast	*W(36) = 0.713, p < 0.0005*	2.74	**6.97***	9.56	**12.45***
4/4	Slow	*W(36) = 0.771, p < 0.0005*	1.84	**4.69***	3.27	**4.25***
	Mid	*W(36) = 0.764, p < 0.0005*	2.26	**5.75***	6.02	**7.84***
	Fast	*W(36) = 0.537, p < 0.0005*	3.99	**10.15***	19.02	**24.77***

Error rate Shapiro-Wilk's tests, skewness and kurtosis for all typists
(* values that exceed acceptable range for skewness or kurtosis of population)

H.2.3 Transformed Error Rate



Transformed error rate box plots for all typists

All Typists						
Time Signature	**Tempo**	**Shapiro-Wilk's**	**Skewness**		**Kurtosis**	
			s	**z_s**	**k**	**z_k**
Without music		$W(36) = 0.977$, *n.s.*	-0.34	-0.87	-0.23	-0.30
3/4	Slow	$W(36) = 0.976$, *n.s.*	-0.33	-0.84	-0.44	-0.57
	Mid	$W(36) = 0.961$, *n.s.*	-0.40	-1.01	-0.64	-0.84
	Fast	$W(36) = 0.980$, *n.s.*	0.02	0.06	-0.36	-0.47
4/4	Slow	$W(36) = 0.978$, *n.s.*	-0.08	-0.21	-0.15	-0.19
	Mid	$W(36) = 0.984$, *n.s.*	-0.25	-0.63	0.16	0.20
	Fast	$W(36) = 0.974$, *n.s.*	0.50	1.28	0.13	0.16

Transformed error rate Shapiro-Wilk's tests, skewness and kurtosis for all typists

H.3 Pairwise Comparisons

H.3.1 Transformed Error Rate

Condition 1	Condition 2	Mean Difference	p
Without music	3/4 Slow	0.008	n.s.
Without music	3/4 Mid	0.043	n.s.
Without music	3/4 Fast	-0.125	0.037**
Without music	4/4 Slow	0.053	n.s.
Without music	4/4 Mid	-0.026	n.s.
Without music	4/4 Fast	0.044	n.s.
3/4 Slow	3/4 Mid	0.036	n.s.
3/4 Slow	3/4 Fast	-0.133	0.025**
3/4 Slow	4/4 Slow	0.045	n.s.
3/4 Slow	4/4 Mid	-0.034	n.s.
3/4 Slow	4/4 Fast	0.036	n.s.
3/4 Mid	3/4 Fast	-0.168	0.030**
3/4 Mid	4/4 Slow	0.009	n.s.
3/4 Mid	4/4 Mid	-0.07	n.s.
3/4 Mid	4/4 Fast	0.001	n.s.
3/4 Fast	4/4 Slow	0.178	0.005**
3/4 Fast	4/4 Mid	0.099	n.s.
3/4 Fast	4/4 Fast	0.169	0.025**
4/4 Slow	4/4 Mid	-0.079	n.s.
4/4 Slow	4/4 Fast	-0.009	n.s.
4/4 Mid	4/4 Fast	0.071	n.s.

Post hoc tests for omnibus effect of music
(significance at $p < 0.0024$ in 1 by 7 design, significance at $p < 0.0033$ in 1 by 6 design, ** significant at the non-adjusted level)

H.3.2 Music Distraction

Condition 1	**Condition 2**	**z**	**p**
3/4 Slow	3/4 Mid	-0.609	n.s.
3/4 Slow	3/4 Fast	-1.035	n.s.
3/4 Slow	4/4 Slow	-1.037	n.s.
3/4 Slow	4/4 Mid	-0.118	n.s.
3/4 Slow	4/4 Fast	-1.958	0.049**
3/4 Mid	3/4 Fast	-1.631	n.s.
3/4 Mid	4/4 Slow	-0.422	n.s.
3/4 Mid	4/4 Mid	-0.721	n.s.
3/4 Mid	4/4 Fast	-2.306	0.021**
3/4 Fast	4/4 Slow	-1.629	n.s.
3/4 Fast	4/4 Mid	-1.056	n.s.
3/4 Fast	4/4 Fast	-1.054	n.s.
4/4 Slow	4/4 Mid	-1.188	n.s.
4/4 Slow	4/4 Fast	-2.450	0.014**
4/4 Mid	4/4 Fast	-1.966	0.049**

Post hoc tests for omnibus effect of music
(significance at $p < 0.0033$, ** significant at the non-adjusted level, *** not significant at Bonferroni or non-adjusted levels, but $p < 0.1$)

Appendix H: Experiment 5 Supplementary Statistical Analyses

References

Alcott, L. M. (2007). Op Eigen Wieken. (G. W. Elberts, Trans.). Retrieved from http://www.gutenberg.org/files/21946/21946-h/21946-h.htm

Alty, J. L., & Rigas, D. I. (1998). Communicating graphical information to blind users using music: the role of context. *The SIGCHI Conference*, 574–581. http://doi.org/10.1145/274644.274721

Apple. (2010, October 1). GarageBand 6 [Computer Software].

Areni, C. S., & Kim, D. (1993). The Influence of Background Music on Shopping Behavior: Classical Versus Top-Forty Music in a Wine Store. *Advances in Consumer Research*, *20*(1), 336–340.

Arrabito, G. R., Mondor, T. A., & Kent, K. J. (2004). Judging the urgency of non-verbal auditory alarms: a case study. *Ergonomics*, *47*(8), 821–840. http://doi.org/10.1080/0014013042000193282

Baddeley, A., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press.

Baddeley, A., & Salamé, P. (1986). The Unattended Speech Effect: Perception or Memory?., *12*(4), 525–529.

Blattner, M., Sumikawa, D., & Greenberg, R. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction*, *4*(1), 11–44. http://doi.org/10.1207/s15327051hci0401_1

Bosman, E. (1993). Age-Related Differences in the Motoric Aspects of Transcription Typing Skill. *Psychology & Aging*, 8(1), 87–102.

Brewster, S. (2002). Overcoming the Lack of Screen Space on Mobile Computers. *Personal Ubiquitous Computing.*, *6*(3), 188–205. http://doi.org/10.1007/s007790200019

Brewster, S. A. (1997). Using non-speech sound to overcome information overload. *Displays*, *17*(3-4), 179–189. http://doi.org/10.1016/S0141-9382(96)01034-7

Brodsky, W. (2001). The effects of music tempo on simulated driving performance and vehicular control. *Transportation Research Part F: Traffic Psychology and Behaviour*, *4*(4), 219–241. http://doi.org/10.1016/S1369-8478(01)00025-0

Brown, M. H., & Hershberger, J. (1992). Color and sound in algorithm animation. *Computer*, *25*(12), 52–63.

Bulmer, M. G. (1979). Principles of Statistics. Dover Publications Inc.

Burgess, T. W. (2004). The Adventures of Grandfather Frog. Retrieved from http://www.gutenberg.org/files/14375/14375-h/14375-h.htm

Burnett, D. (2016, August 20). Does music really help you concentrate? *The Guardian*. Retrieved from https://www.theguardian.com/education/2016/aug/20/does-music-really-help-you-concentrate

Burroughs, E. R. (2008). The Outlaw of Torn. Retrieved from http://www.gutenberg.org/files/369/369-h/369-h.htm

Burt, J. L., Bartolome, D. S., Burdette, D. W., & Comstock, J. R., JR. (1995). A psychophysiological evaluation of the perceived urgency of auditory warning signals. *Ergonomics*, *38*(11), 2327–2340. http://doi.org/10.1080/00140139508925271

Caldwell, C., & Hibbert, S. A. (2002). The influence of music tempo and musical preference on restaurant patrons' behavior. *Psychology and Marketing*, *19*(11), 895–917. http://doi.org/10.1002/mar.10043

Cassidy, G. G., & MacDonald, R. A. R. (2010). The effects of music on time perception and performance of a driving game. *Scandinavian Journal of Psychology*, *51*(6), 455–464. http://doi.org/10.1111/j.1467-9450.2010.00830.x

Clarke, E., Dibben, N., & Pitts, S. (2010). Music and Mind in Everyday Life. Oxford University Press.

Clayton, M., Sager, R., & Udo, W. (2005). In time with the music: the concept of entrainment and its significance for ethnomusicology. *European Meetings in Ethnomusicology, ESEM Counterpoint*, *11*, 1–82.

Davidson, L. (2016, June 2). This is the kind of music you should listen to at work. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/business/2016/06/02/this-is-the-kind-of-music-you-should-listen-to-at-work/

Ding, C. G., & Lin, C.-H. (2012). How does background music tempo work for online shopping? *Electronic Commerce Research and Applications*, *11*(3), 299–307. http://doi.org/10.1016/j.elerap.2011.10.002

Dixon, L., Trigg, R., & Griffiths, M. (2007). An Empirical Investigation of Music and Gambling Behaviour. *International Gambling Studies*, *7*(3), 315–326. http://doi.org/10.1080/14459790701601471

Edwards, A. D. N. (1989). Soundtrack: an auditory interface for blind users. *Human-Computer Interaction*, *4*(1), 45–66. http://doi.org/10.1207/s15327051hci0401_2

Edworthy, J., & Waring, H. (2006). The effects of music tempo and loudness level on treadmill exercise. *Ergonomics*, *49*(15), 1597–1610. http://doi.org/10.1080/00140130600899104

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving Auditory Warning Design: Relationship between Warning Sound Parameters and Perceived Urgency. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *33*(2), 205–231.

Everitt, B. S., & Howell, D. C. (2005). Winsorized Robust Measures. Wiley-Blackwell.

Everyday. (2016). Everyday. In *Oxford English Dictionary*. Oxford University Press. Retrieved from http://www.oed.com/view/Entry/65342

Flesch, R. (n.d.). How to Write Plain English. Retrieved April 18, 2010, from https://web.archive.org/web/20160712094308/http://www.mang.canterbury.ac.nz/writing_guide/writing/flesch.shtml

Francioni, J. M., Albright, L., & Jackson, J. A. (1991). Debugging parallel programs using sound. *ACM SIGPLAN Notices*, *26*(12), 68–75. http://doi.org/10.1145/127695.122765

Gaver, W. (1986). Auditory Icons: Using Sound in Computer Interfaces. *Human-Computer Interaction*, *2*(2), 167–177. http://doi.org/10.1207/s15327051hci0202_3

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*, *42*(3), 237–288. http://doi.org/10.2307/1169991

Gestalt. (2016). Gestalt. In *Oxford English Dictionary*. Oxford University Press. Retrieved from http://www.oed.com/view/Entry/ 77951

Gopher, D., Brickner, M., & Navon, D. (2005). Different Difficulty Manipulations Interact Differently With Task Emphasis: Evidence for Multiple Resources. *Journal of Experimental Psychology Human Perception and Performance*, *8*(1), 146–157.

Haake, A. B. (2010). *Music listening in UK offices: Balancing internal needs and external considerations*. University of Sheffield.

Hayes, J. R., & Chenoweth, N. A. (2016). Is Working Memory Involved in the Transcribing and Editing of Texts? *Written Communication*, *23*(2), 135–149. http://doi.org/10.1177/0741088306286283

Hear. (2016). Hear. In *Oxford English Dictionary*. Oxford University Press. Retrieved from http://www.oed.com/view/Entry/85048

Henderson, M., Crews, A., & Barlow, J. (1945). A study of the effect of music distraction on reading efficiency. *Journal of Applied Psychology*, *29*(4), 313–317.

Hébert, S., Béland, R., Dionne-Fournelle, O., Crête, M., & Lupien, S. J. (2005). Physiological stress response to video-game playing: the contribution of built-in music. *Life Sciences*, *76*(20), 2371–2380. http://doi.org/10.1016/j.lfs.2004.11.011

Holland, S., Morse, D. R., & Gedenryd, H. (2002). AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface. *Personal Ubiquitous Computing.*, *6*(4), 253–259. http://doi.org/10.1007/s007790200025

Howell, D. C. (2008). Fundamental Statistics for the Behavioural Scientists (7 ed.). Wadsworth Cengage Learning.

Hudson, S. E., & Smith, I. (1996). Electronic mail previews using non-speech audio. *Conference Companion on Human Factors in Computing Systems*, 237–238. http://doi.org/10.1145/257089.257299

Hughett, P. (2006, January 1). Pword14 [Data File]. Retrieved from http://www.med.upenn.edu/bbl/pseudowords.html

Inhoff, A., & Wang, J. (1992). Encoding of Text, Manual Movement Planning, and Eye-Hand Coordination During Copytyping. *Journal of Experimental Psychology: Human Perception & Performance*, 18(2), 437–448.

Jensen, M. B. (1931). The influence of jazz and dirge music upon speed and accuracy of typing. *Journal of Educational Psychology*, *22*(6), 458–462. http://doi.org/10.1037/h0074752

Jiang, X., & Sengupta, A. K. (2011). Effect of music and induced mental load in word processing task (pp. 3261–3266). Presented at the 2011 IEEE International Conference on Systems, Man and Cybernetics - SMC, IEEE. http://doi.org/10.1109/ICSMC.2011.6084172

Kaltcheva, V. D., & Weitz, B. A. (2006). When Should a Retailer Create an Exciting Store Environment? *Journal of Marketing*, *70*(1), 107–118.

Kanai, H., Nakada, T., Hanbat, Y., & Kunifuji, S. (2008). A support system for context awareness in a group home using sound cues (pp. 264–267). Presented at the 2008 Second International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), IEEE. http://doi.org/10.1109/PCTHEALTH.2008.4571085

Kane, S. K., Wobbrock, J. O., Harniss, M., & Johnson, K. L. (2008). TrueKeys: identifying and correcting typing errors for people with motor impairments. *The 13th International Conference*, 349–352. http://doi.org/10.1145/1378773.1378827

Karageorghis, C. I., Cheek, P., Simpson, S. D., & Bigliassi, M. (2017). Interactive effects of music tempi and intensities on grip strength and subjective affect. *Scandinavian Journal of Medicine & Science in Sports*, *28*(3), 1166–1175. http://doi.org/10.1111/sms.12979

Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2011). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music*, *39*(4), 424–448.

Keith, N., & Ericsson, K. A. (2007). A Deliberate Practice Account of Typing Proficiency in Everyday Typists. *Journal of Experimental Psychology: Applied*, 13(3), 135–145. http://doi.org/10.1037/1076-898X.13.3.135

Khalfa, S., Isabelle, P., Jean-Pierre, B., & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters*, *328*(2), 145–149.

Khalfa, S., Roy, M., Rainville, P., Dalla Bella, S., & Peretz, I. (2008). Role of tempo entrainment in psychophysiological differentiation of happy and sad music? *International Journal of Psychophysiology,* 68(1), 17–26. http://doi.org/10.1016/j.ijpsycho.2007.12.001

Kotler, P. (1973). Atmospherics as a Marketing Tool. *Journal of Retailing*, *49*(4), 48.

Landau, S., Wiener, W., Naghshineh, K., & Giusti, E. (2005). Creating Accessible Science Museums With User-Activated Environmental Audio Beacons (Ping!). *Assistive Technology*, *17*(2), 133–143. http://doi.org/10.1080/10400435.2005.10132103

Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. In *Flow and the Foundations of Positive Psychology* (pp. 21–34). Springer.

Lesiuk, T. (2000). The Effect of Music Listening on a Computer Programming Task. *Journal of Computer Information Systems*, *40*(3), 50–57.

Lesiuk, T. (2005). The effect of music listening on work performance. *Psychology of Music*, *33*(2), 173–191. http://doi.org/10.1177/0305735605050650

Levy, C. M., & Marek, P. (1999). Testing components of Kellogg's multicomponent model of working memory in writing: The role of the phonological loop. In M. Torrance & G. C. Jeffrey (Eds.), *The cognitive demands of writing Processing capacity and working memory in text production*.

Lipscomb, S. D., & Zehnder, S. M. (2004). Immersion in the Virtual Environment: The Effect of a Musical Score on the Video Gaming Experience. *Journal of Physiological Anthropology and Applied Human Science, 23*(6), 337–343.

Liquid Mind. (2001). *Continuity, Pt. 1*. Liquid Mind V: Serenity.

Listen. (2016). Listen. In *Oxford English Dictionary*. Oxford University Press. Retrieved from http://www.oed.com/view/Entry/109008

Mansur, D. L., Blattner, M. M., & Joy, K. I. (1985). Sound graphs: A numerical data analysis method for the blind. *Journal of Medical Systems*, *9*(3), 163–174. http://doi.org/10.1007/BF00996201

Martin, R. C., Wogalter, M. S., & Forlano, J. G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*, *27*(4), 382–398. http://doi.org/10.1016/0749-596X(88)90063-0

McGee-Lennon, M., Wolters, M., McLachlan, R., Brewster, S., & Hall, C. (2011). Name that tune: musicons as reminders in the home (pp. 2803–2806). Presented at the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada: ACM. http://doi.org/10.1145/1978942.1979357

Milliman, R. E. (1982). Using Background Music to Affect the Behavior of Supermarket Shoppers. *Journal of Marketing*, *46*(3), 86–91. http://doi.org/10.2307/1251706

Milliman, R. E. (1986). The Influence of Background Music on the Behavior of Restaurant Patrons. *Journal of Consumer Research*, *13*(2), 286–289. http://doi.org/10.2307/2489234

Moosmusic. (n.d.). *Big Dummy Shake*. Retrieved from http://www.cambridge-mt.com/ms-mtk.htm#Pop

Morimoto, C. H., & Amir, A. (2010). Context switching for fast key selection in text entry applications. *The 2010 Symposium*, 271–274. http://doi.org/10.1145/1743666.1743730

Morley, S., Petrie, H., O'Neill, A.-M., & McNally, P. (1999). Auditory navigation in hyperspace: Design and evaluation of a non-visual hypermedia system for blind users. *Behaviour & Information Technology*, *18*(1), 18–26. http://doi.org/10.1080/014492999119219

Mozart, W. A. (2005). *Allegro Assai. Piano Concerto No. [Performed by A. Ligeti, C. Hungaricus & J. Jandó]*. The Very Best of Mozart.

Mozart, W. A. (1996). *Menuetto and Allegro. Symphony No 25 in G Minor, K 183: III and IV [Performed by Berliner Philharmoniker & K. Böhm]*. Mozart: 46 Symphonies, Disc 6.

North, A. C. (2004). Uses of Music in Everyday Life, *22*(1), 41–77. http://doi.org/10.1525/mp.2004.22.1.41

North, A. C., & Hargreaves, D. J. (1998). The Effect of Music on Atmosphere and Purchase Intentions in a Cafeteria. *Journal of Applied Social Psychology*, *28*(24), 2254–2273. http://doi.org/10.1111/j.1559-1816.1998.tb01370.x

North, A. C., Hargreaves, D. J., & McKendrick, J. (1999). The Influence of In-Store Music on Wine Selections. *Journal of Applied Psychology*, *84*(2), 271–276.

Noseworthy, T. J., & Finlay, K. (2009). A Comparison of Ambient Casino Sound and Music: Effects on Dissociation and on Perceptions of Elapsed Time While Playing Slot Machines. *Journal of Gambling Studies*, *25*(3), 331–342. http://doi.org/10.1007/s10899-009-9136-x

Ober, S. (n.d.). Typing Speed Requirements for Administrative Positions. Retrieved June 15, 2016, from http://www.gdpkeyboarding.com/Word_Files/Required_Typing_Speeds.pdf

Open Source. (2012, March 13). Audacity [Computer Software]. Retrieved from http://www.audacityteam.org/

Pareles, J. (2002, June 9). David Bowie, 21st-Century Entrepreneur. *The New York Times*. Retrieved from http://www.nytimes.com/2002/06/09/arts/david-bowie-21st-century-entrepreneur.html

Pendse, A., Pate, M., & Walker, B. N. (2008). The accessible aquarium: identifying and evaluating salient creature features for sonification. *The 10th International ACM SIGACCESS Conference*, 297–298. http://doi.org/10.1145/1414471.1414546

Pool, M. M., Koolstra, C. M., & Van Der Voort, T. H. A. (2006). The Impact of Background Radio and Television on High School Students' Homework Performance. *Journal of Communication*, 53(1), 74–87. http://doi.org/10.1111/j.1460-2466.2003.tb03006.x

Reilly, T., & Smith, D. (1986). Effect of work intensity on performance in a psychomotor task during exercise. *Ergonomics*, *29*(4), 601–606. http://doi.org/10.1080/00140138608968294

Rossier, J., Haeberli, C., & Schenk, F. (2000). Auditory cues support place navigation in rats when associated with a visual cue. *Behavioural Brain Research*, *117*(1-2), 209–214. http://doi.org/10.1016/S0166-4328(00)00293-X

Russell, J., Weiss, A., & Mendelsohn, G. (1989). Affect Grid: A Single-Item Scale of Pleasure and Arousal. *Journal of Personality and Social Psychology*, *57*(3), 493–502.

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 150–164. http://doi.org/10.1016/S0022-5371(82)90521-7

Salthouse, T. A. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological Bulletin*.

Sanders, T., & Cairns, P. (2010). Time perception, immersion and music in videogames. In *Proceedings of the BCS Interaction Specialist Group Conference* (pp. 160–167). Dundee, United Kingdom: British Computer Society.

Sanderson, P. M., Watson, M. O., Russell, W. J., Jenkins, S., Liu, D., Green, N., et al. (2008). Advanced Auditory Displays and Head-Mounted Displays: Advantages and Disadvantages for Monitoring by the Distracted Anesthesiologist. *Anesthesia & Analgesia KW -*, *106*(6).

Saner, E. (2016, June 1). Is M&S right to scrap in-store music? *The Guardian*. Retrieved from https://www.theguardian.com/business/shortcuts/2016/jun/01/is-marks-and-spencer-right-scrap-in-store-music

Senior, M. (n.d.). Mixing Secrets: Multitrack Download Library. Retrieved September 7, 2016, from http://www-cambridge-mt.com/ms-mtk.htm

Shaffer, L. H. (1975). Multiple Attention in Continuous Verbal Tasks. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and Performance V* (pp. 157–167).

Sloboda, J. A. (2010). Music in Everyday Life. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 492–514). Oxford University Press. http://doi.org/10.1093/acprof:oso/9780199230143.003.0018

Sloboda, J. A., & O'Neill, S. A. (2001). Emotions in everyday listening to music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion theory and research* (pp. 415–429). Oxford University Press.

Spenwyn, J., Barrett, D. J. K., & Griffiths, M. D. (2009). The Role of Light and Music in Gambling Behaviour: An Empirical Pilot Study. *International Journal of Mental Health and Addiction*, *8*(1), 107–118. http://doi.org/10.1007/s11469-009-9226-0

Stahl, C. (2007). The roaring navigator: a group guide for the zoo with shared auditory landmark display. *The 9th International Conference*, 383–386. http://doi.org/10.1145/1377999.1378042

Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S., & O'Modhrain, S. (2005). GpsTunes: controlling navigation via audio feedback. *The 7th International Conference*, 275–278. http://doi.org/10.1145/1085777.1085831

Tabachnick, B. G., & Fidell, L. S. (2007). Experimental designs using ANOVA. Thomson/Brooks/Cole.

Teigen, K. H. (2005). Yerkes-Dodson: A Law for All Seasons. *Theory and Psychology*, *4*(4), 525–547.

The Doppler Shift. (n.d.). *Atrophy*. Retrieved from http://www.cambridge-mt.com/ms-mtk.htm#AltRock

Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2012). Fast and loud background music disrupts reading comprehension. *Psychology of Music*, *40*(6), 700–708. http://doi.org/10.1177/0305735611400173

Van Dyck, E., Moens, B., Buhmann, J., Demey, M., Coorevits, E., Bella, S. D., & Leman, M. (2015). Spontaneous Entrainment of Running Cadence to Music Tempo. *Sports Medicine - Open*, *1*(1), 1–15. http://doi.org/10.1186/s40798-015-0025-9

Vickers, P., & Alty, J. L. (1996). Caitlin: A musical program auralisation tool to assist novice programmers with debugging. Presented at the ICAD Third International Conference on Auditory Display.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. http://doi.org/10.1080/14639220210123806

Wiener, E. L., Curry, R. E., & Faustina, M. L. (1984). Vigilance and Task Load: In Search of the Inverted U. *Human Factors*, *26*(2), 215–222. http://doi.org/10.1177/001872088402600208

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. http://doi.org/10.1002/cne.920180503