Memory for Melody

Investigating the link between experience, perception, and memory formation

Steffen A. Herff

A dissertation submitted for the degree of Doctor of Philosophy

MARCS Institute for Brain, Behaviour, and Development

Western Sydney University, Australia

March 2017

Wer für etwas brennt, der fackelt nicht lange.

- Aaron Feuermal

Dedication

Dedicated to my grandparents who never stopped believing in me.

Der Jung macht das schon.

- Opa Herff 1931-2016 & Oma Herff 1934-2017

Acknowledgements

I am forever indebted to Prof. Roger Dean and Dr. Kirk Olsen. I don't think there could have been two better supervisors for me. Your doors were always open, you provided endless support, and you still gave me all the freedom to pursue my own research. Your thoughts have been invaluable and you will remain the first people I turn to when in need of good advice. I have thrived under your guidance and I cannot thank you enough for what you have done for me.

I would like to thank my friend and colleague Dr. Andrew Milne, mostly for being Andy. Thank you for countless discussions, I enjoyed every single one of them. With *XronoMorph* we built an inspiring tool for the algorithmical composition of rhythms that will always be my prime example of how basic research can contribute to society and that, surprisingly, people actually care about what we do.

My gratitude goes to all members of *Illusions of Grandeur* and the *Seafarer's Rest project*: Andy, Kirk, Felix, Jenni, Ian, and Jules. Over the three years of my PhD, the band provided musical and creative stimulation and kept reminding me why I wanted to research music perception to begin with. At this point, I would also like to thank the MARCS Tech-team for their support as well as for letting us use the performance studio for our after-hours 'work'.

I am very grateful to the MARCS Institute and all of its support staff. Especially Kate, Karen, Craig, Darlene, Sonya, Gail, Jan, and Farah. Special thanks also to the entire *Music, Cognition, and Action* group for providing a friendly, constructive, and generally pleasant environment. I would also like to thank Jon Prince for sharing his data. Out of this initial act of generosity a new, exciting, and on-going collaboration was formed.

My deepest respect and love goes to my parents and grandparents who were always there for me. Especially in recent years you have been inspiring examples of strong characters in the face of tough circumstances. The same goes for my brother Christian. As long as I can remember, you have been looking out for me. You are, and always will be, Babu to me and your growing family brings great joy to my life.

My final thanks go to the people that know me the best and, despite that, have still decided to be friends with me. None of this would have been possible without you. Thank you for everything and more Felix, Simon, Denise, Kim, Jules, Phil, and Lauren.

Declaration

I hereby declare that this submission is my own work and, to the best of my knowledge, it contains no material previously published or written by another person, nor material which has been accepted for the award of any other degree or diploma at the Western Sydney University, or at any other educational institution, except where due acknowledgement is made in the dissertation.

I also declare that the intellectual content of this dissertation is the product of my own work, except to the extent that assistance from others in the project's design and conception is acknowledged.

____

Steffen A. Herff

Table of Contents

Note that chapters 2-5 have been either submitted or accepted for publication in peer-reviewed journals. Figures, tables and experiments in these chapters have been relabelled relative to their position in the dissertation.

Chapter 2: Herff, S.A., Olsen, K.N., & Dean, R.T. (2017). Resilient memories for melodies: The number of intervening melodies does not influence novel melody recognition. *Quarterly Journal of Experimental Psychology*. Accepted March 2, 2017.

Chapter 3: Herff, S.A., Olsen, K.N., Dean, R.T., & Prince, J. (2017). Memory for melodies in unfamiliar tuning systems: Investigating effects of recency and number of intervening items. *Quarterly Journal of Experimental Psychology*.

Chapter 4: Herff, S.A., Olsen, K.N., Prince, J., & Dean, R.T. (2017). Interference in memory for pitch-only and rhythm-only sequences. *Musicae Scientiae,* Advance online publication. doi:10.1177/1029864917695654

Chapter 5: Herff, S.A., Dean, R.T., & Olsen, K.N. (submitted). Inter-rater agreement in memory for melody as a measure of listeners' similarity in music perception. *Psychomusicology: Music, Mind & Brain*.

List of Tables

List of Figures

Abstract

Music exhibits remarkable properties in the context of memory. For example, memory for melodies is long lasting, persistent, and spared from some forms of dementia and severe brain injuries. Over the course of 10 experiments, we here attempt to shed light on the question of what makes music 'special' by specifically investigating interference in memory for melody. Most domains show a decrease in recognition performance between the first and second presentation of an object as the number of intervening items increases. We tested this cumulative disruptive interference effect in the context of memory for melodies, with results showing that memory for melodies is not much affected by the number of intervening items. Specifically, the probability of correctly recognising a melody was statistically identical and above chance between 1 and up to nearly 200 intervening melodies. To explain these findings, we provide a new Regenerative-Multiple-Representations (RMR) conjecture. The conjecture describes a crucial link between prior experience, perception, and subsequent formation of memories. Using the theoretical framework of this conjecture, we further explored memory for melodies in a series of experiments. In the process, we revealed how to disrupt and shape memory for melodies' resilience to cumulative interference using melodies in an unfamiliar tuning system and with pitch-only and rhythm-only sequences. In a final analysis, we used memory as a window into perception in more general terms, and analysed data from all experiments combined to measure the degree of similarity in listeners' perception of music. The findings of this dissertation contribute to our understanding of fundamental memory phenomena, while providing practical implementations and elucidating further the mechanisms that explain why music may indeed be 'special'.

Chapter 1

Introduction

1.1 Memory

Memory has rightfully earned its place amongst the most studied subjects of cognitive psychology. This is not surprising considering that some form of memory is involved in nearly all levels of conscious and unconscious human behaviour (Anderson & Bower, 2014; Baddeley, 1997). Memory is not only crucial for any form of skill acquisition, but also essential for emotional responses, perception, and personality. Indeed, dramatic cases of memory loss show that what others perceive as one's personality is deeply anchored in memory (Baddeley, 2014).

Memory is a complex, multi-faceted subject. A plethora of memory models have been developed and, amongst other things, they commonly differentiate between *long-term memory*, the predominant subject of this dissertation, and *short-term memory* (Atkinson & Shiffrin, 1968). Short-term memory refers to durations of approximately 18 to 30 seconds (Peterson & Peterson, 1959; Posner, 1966). Long-term memory roughly describes anything beyond 30 seconds (Shiffrin & Atkinson, 1969) (but also see Cowan (2008) for a more functional definition). Furthermore, memory retrieval, the process of accessing a stored memory representation, can broadly be divided into *recall and recognition* (Cabeza et al., 1997). Recall describes retrieval of previously encountered stimuli; for example, the answer to the question, 'what did you have for breakfast?'. Recognition describes the skill of differentiating between new and previously encountered stimuli. Hearing a song on the radio and immediately knowing that you have heard this song before is one example of recognition. It is recognition that is the particular focus of this dissertation.

1.1.1 Recognition Paradigms

In tests of recognition, participants are usually asked to judge whether or not an item presented in a test phase was present in a preceding learning phase (e.g., Shepard, 1967). Responses are usually given by simply indicating *'Yes/No'* or *'Old/New'* (Finnigan, Humphreys, Dennis, & Geffen, 2002).

A *blocked paradigm* where a learning phase precedes a test phase is perhaps the most common experimental design used to test recognition performance (e.g., McAuley, Stevens, & Humphreys, 2004; Morrison, Demorest, Aylward, Cramer, & Maravilla, 2003; Müllensiefen & Halpern, 2014; Schaal, Javadi, Halpern, Pollok, & Banissy, 2015; Yonelinas, Kroll, Dobbins, & Soltani, 1999). This design has advantages, such as tight experimental control in research that aims to isolate encoding and retrieval processes. For example, some cases of neuroimaging or brain stimulation may require grouping of trials that involve similar neurological processes.

However, blocked designs also come with disadvantages. In a blocked design, participants are able to distinguish a test trial from a learning trial. This means that participants know that they only have to remember items presented in the learning phase, rather than items presented in the test phase. This information could lead to different processing of items in learning and test phases. Participants concentrate on learning during the learning phase and on recognition during the test phase. When investigating recognition in such a design, effectively only test trials can be used for statistical analysis. A blocked paradigm is also somewhat different to the everyday continuous use of recognition, where we often do not know exactly when it is important to focus on recognition or on learning.

A *continuous recognition paradigm* presents a solution to the problem of participants being capable to distinguish a test trial from a learning trial (Shepard & Teghtsoonian, 1961). Stimuli are presented continuously and sometimes an item repeats. After each trial, participants are required to respond whether or not the item was previously presented. As a result, participants perform the same task on every trial. All experiments reported in this dissertation use a continuous recognition paradigm to ensure that participants cannot distinguish a test trial from a learning trial (Dowling, 1991). A continuous recognition paradigm is also very useful to investigate disruptive effects of interference and decay on recognition memory that can ultimately lead to forgetting (Berman, Friedman, & Cramer, 1991; Campeanu, Craik, Backer, & Alain, 2014; Ferris, Crook, Clark, Mccarthy, & Rae, 1980; Friedman, 1990b; Hockley, 1992; Sadeh, Ozubko, Winocur, & Moscovitch, 2014).

1.1.2 Forgetting: Interference and Decay

Sometimes memory fails: a colleague's name eludes us or we miss the birthday of a family member. Forgetting broadly describes the process that leads to some memories becoming inaccessible or distorted; it is a common, but not necessarily negative process (Eysenck & Keane, 2015, pp. 242-256). There are some factors that facilitate forgetting. Two of the most important factors are *decay* and *interference*. Decay describes fading of memory representations due to the passing of time. The more time that has passed, the harder it is to access a memory. Interference describes a continuous decrease in memory performance, not because of the passing of time, but from the additional information that is learned after or even before the encoding of a stimulus into memory. For example, it is relatively easy to recognise a name on a guest list when the person has just introduced themselves. However, after being introduced to 60 people over the course of an evening, it can be quite hard to accurately recognise which name has previously been encountered. The present project predominantly investigates memory in the framework of recognition and interference.

Interference. Research investigating interference and recognition usually produces results like those shown in Figure 1.1. The x-axis represents the number of intervening items and the y-axis represents some form of recognition performance. Typically, as the number of intervening items increases, the cumulative disruptive effect resulting in a decrease in recognition performance also increases. These findings are very common for most stimuli (Bui, Maddox, Zou, & Hale, 2014; Donaldson & Murdock, 1968; Hockley, 1992; Konkle, Brady, Alvarez, & Oliva, 2010; Nickerson, 1965; Olson, 1969; Poon & Fozard, 1980; Rakover & Cahlon, 2001; Sachs, 1967; Sadeh et al., 2014). However, this dissertation will demonstrate that this phenomenon is not universal.

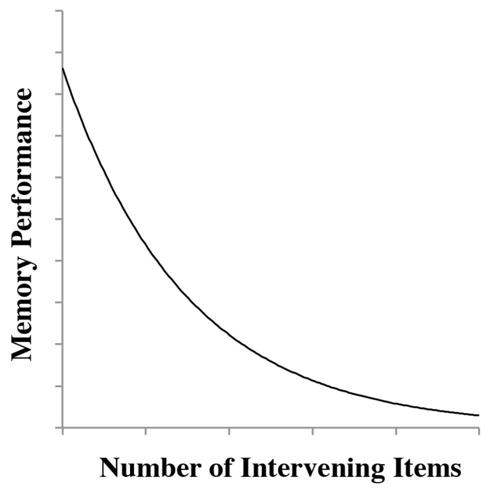


Figure 1.1: Schematic representation of traditional inference curves. The exact functions may vary; however generally, memory performance decreases as the number of intervening items increases.

1.1.3 The Stimulus Specificity of Memory

As mentioned above, memory is a complex and multi-faceted subject. One reason for this is the stimulus specificity of memory. For example, research on face perception has identified the fusiform face area within the brain as specialized region involved in storing and comparing faces (Kanwisher & Yovel, 2006). This means that memory phenomena observed within a specific stimulus domain are not necessarily true for other stimuli. For instance, a series of experiments suggest that memory for prose declines over time (Sachs, 1967), but the mere passing of time may have only minimal disruptive effects on memory for poetry (Tillmann & Dowling, 2007).

In the study by Tillmann and Dowling (2007), participants listened to short prose stories or poems. After a retention interval of up to ~29 sec, one of three possible test items were presented. The test item could be an identical repetition of a part of the story or poem. Alternatively, the test item could be only semantically identical, but come with surface changes such as a paraphrased form of the original sentence. A third option was an item that was semantically different. Participants were asked to judge whether each subsequent test item had been presented before. Results indicate that in general, discrimination performance between identical and only semantically identical versions of an item declines over time. Interestingly, discrimination performance between identical and paraphrased versions only declined in prose, but not in poetry. To interpret these results, Tillmann and Dowling (2007) offer an interesting explanation: multiple features of a stimulus may be bound together in memory as a coherent whole, and prose and poetry might use different weighting during the binding process because poetry emphasises different perceptual features compared to prose (e.g., rhythm) (also see Kroll, Knight, Metcalfe, Wolf, and Tulving, 1996, for support that the hippocampus might play a crucial role in the binding of underlying perceptual components). This explanation is explored in greater detail in this dissertation and ultimately, is combined with other findings in the literature to form the basis of a novel Regenerative Multiple Representations (RMR) conjecture that is presented in Chapter 2

Memory's stimulus specificity poses a challenge and an opportunity alike. On the one hand, findings using a specific kind of stimulus cannot be generalised to other domains. For example, pictures are differently processed by the brain compared to words (Erdelyi & Becker, 1974). On the other hand, similarities and dissimilarities in the way memory treats different kinds of stimuli can tell us a lot about memory in general. For example, research on picture memory has revealed interesting phenomena such as change blindness (Simons & Levin, 1997), a phenomenon in which observers fail to notice changes in a visual scene (for example due to flickering), which greatly informs our knowledge of human perception.

A study reported in Isola, Xiao, Torralba, and Olivia (2011) shows that memorability of a visual picture is a stable property across human observers. This means that pictures that are more memorable for one person are also most likely more memorable for another person. However, the features used in that particular study were specific to the visual domain and are not applicable to other domains such as audition. The auditory domain, however, is worth investigating. In particular, music is a stimulus that shows intriguing findings in the context of memory. In this dissertation, Chapter 5 reports a series of experiments that addresses the stability of memory in the context of music.

1.1.4 Memory for Music

Music is a cultural universal and a culture without music has yet to be found (Cross, 1999, 2001, 2014). In terms of memory, everyone is acquainted with the feeling of a melody being "stuck" in one's ear. *Involuntary music imagery* (e.g., *earworms)* is a common example of how long lasting and persistent memory for music can be (Bailes, 2007, 2015; Halpern & Bartlett, 2011; Williams, 2015). Background music can improve learning of new content, and combining simple melodies with lyrics (e.g. nursery rhymes) is an efficient way to learn complex concepts (Chazin & Neuschatz, 1990; Colley, 2016; Smith, 1985). Remarkably, memory for music also seems to be spared by some forms of dementia and severe brain injuries (Baird & Samson, 2014; Cuddy & Duffin, 2005; Jacobsen et al., 2015; Schulkind, 2009). This is an interesting phenomenon, especially considering that dementia is one of the biggest challenges faced by our ageing population. The relationship between memory and music is further discussed in Chapter 2

1.2 Aim and Summary

The present project aims to deepen our understanding of human memory, specifically the process of recognition, by investigating a particular domain that has previously been shown to yield intriguing memory phenomena: music. Vice versa, we aim to shed light on music, a cultural universal, using the psychological framework of one of humans' most cherished cognitive systems: memory. The main body of empirical work presented in this dissertation is reported in a series of manuscripts either submitted or published in peer-reviewed journals. As a result of this format, there will be a certain degree of repetition. Chapters that include a peer-reviewed manuscript are indicated throughout. The Human Research Ethics Committee of the Western Sydney University approved the present project (H10847, the general participant information sheet can be found in Appendix F).

The empirical work in this dissertation begins by demonstrating in Chapter 2 that memory for melody behaves very differently compared with most domains when long-term interference of recognition is considered. As part of the investigation reported in Chapter 2 a novel memory conjecture is developed to explain the intriguing memory phenomena in music and provide a theoretical framework for further research. The conjecture is introduced in 2.9 and is termed Regenerative Multiple Representations (RMR) conjecture. The RMR conjecture makes predictions within, but also outside, the domain of music. Some of the predictions within the domain of music are tested in Chapter 3 where preliminary support for the RMR conjecture is reported in the context of memory for melodies in an unfamiliar tuning system.

In Chapter 4 we further explore memory for melody by dissecting melodies into their pitch and rhythmic sequences. This chapter provides insight into memory for two main components of melody. Indeed, results here show that memory for pitch sequences and rhythmic sequences both contribute to memory for melody; however, they may behave differently to one another in terms of interference. If combined, they show some interactions that cannot be explained by the two sequences individually. This is also the first experimental investigation in the memory and music literature of long-term interference effects in rhythmic information.

On a more technical note, we implement a novel data-driven approach to account for participants' response tendency shifts during an experiment, described in Section 2.4.1. Sometimes, participants change their response behaviour over the course of an experiment (Berch, 1976; Donaldson & Murdock, 1968; Snodgrass & Corwin, 1988). These shifts in response tendencies are a general problem for continuous memory tasks, thus our statistical data-driven approach of accounting for them may also be of use for memory research in other domains. We also provide a protocol for an in-depth stimulus-selection-procedure used to draw stimuli from a large stimulus corpus. The procedure is based on clustering stimuli within a multi-dimensional feature space as well as perceptual piloting. This stimulus selection procedure is described in Section 2.6.1 and Appendix B (p. 250). This stimulus selection procedure may also be of use within and outside the domain of music.

After in-depth investigations of recognition and interference in memory for melody, our final analysis in Chapter 5 assumes a broader perspective. Drawing from the wealth of data provided by 10 experiments, we investigate the similarity of listeners' music perception by using their recognition response patterns as a window into perception. Theoretical as well as practical implications of the results are discussed in Chapter 6

We conclude that indeed, memory for melody does have intriguing properties that can be used to inform the domains of music and memory, both individually and together. We report that memory for melody shows virtually no cumulative disruptive effects on recognition from the number of intervening melodies. We replicate this result and extend it to up to ~200 intervening melodies. Interestingly, this phenomenon is no longer observed when melodies are presented in unfamiliar tuning systems, rather than familiar tuning systems such as those in the Western tonal tradition. When melodies' individual rhythmic and pitch components are tested, we also observe cumulative interference. All of this points to the basic tenets of the RMR conjecture, that prior knowledge informs perception, which in turn influences subsequent formation of memory representations. The finding of this dissertation contributes to our general understanding of memory by demonstrating the stimulus specificity of memory as well as offering a conjecture to account for it. The findings also shed light on human music perception and will inform computational models that aim to predict melodies that are memorable and those that are not. We end with some future research directions that follow from the RMR conjecture, as well as practical implications. The next sections provide an overview of the general approach we used to collect and analyse data.

1.2.1 General Experimental Approach

Participants. An experimental approach was used to investigate interference effects on recognition of musical stimuli. This means we recruied large numbers of participants and conducted quantitative statistical analyses of the results. All experiments of this dissertation followed a similar experimental protocol. The vast majority of the participants were undergraduate students from Western Sydney University. We predominantly recruited participants with little musical expertise. This is because musical expertise can have substantial influences on music perception, as well as memory (M. A. Cohen, Evans, Horowitz, & Wolfe, 2011; Halpern & Bower, 1982; Schellenberg & Weiss, 2013). This is generally observed in a main-effect of musicians outperforming non-musicians (Bigand & Poulin-

Charronat, 2006). For example, in Herff and Czernochowski (in press), expert musicians vastly outperformed non-musicians in a melody recognition task. When attention was divided during a condition that required simultaneous monitoring of digits, musicians' melody recognition performance suffered to a similar extent compared to non-musician's melody recognition performance. However, during the divided attention condition the musician's outperformed non-musicians in the digit-monitoring task, even though musicians generally do not perform better in such tasks in the absence of an additional music perception task (Brandler & Rammsayer, 2003; Fernandes & Moscovitch, 2000; Rodrigues, Loureiro, & Caramelli, 2014).

Stimuli. We used *monophonic* or *single line* melodies as stimuli throughout. Melodies are an essential aspect of almost all music, including non-Western music (Corrigall & Trainor, 2014). We are exposed to many melodies on a day-to-day basis (Krause, North, & Hewitt, 2014). Often many new melodies are heard before we an old one is encountered again. This makes melody recognition a fitting domain to investigate recognition and interference in the context of music. Of course, there are many other important aspects to music; for example, musical tuning, tonality, or polyphony (i.e., multiple notes simultaneously), and this dissertation does indeed explore the relationship of tuning system and melody recognition in Chapter 3 Tonality, polyphony, and many other important aspects of music, however, are beyond the scope of this dissertation and will be explored in future research.

Memory for *novel* melodies (never heard before an experiment) and *familiar* melodies (heard prior to an experiment) elicit some remarkable differences from listeners (McAuley et al., 2004). For example, discrimination between studied and non-studied melodies is significantly better in familiar melodies (Peretz & Gaudreau, 1998). The present project is a thorough investigation of memory for *novel* melodies. However, replicating the experiments of this dissertation with melodies that are *familiar* to participants has the potential to further elucidate the difference between memory for familiar and memory for unfamiliar melodies.

Experiment 1 (in 2.4) and Experiment 2 (in 2.5) explore novel melodies that resemble modern advertisement jingles or themes commonly encountered in movies. Experiment 3 (in 2.6) and Experiment 4 (in 2.7) use a sophisticated computational procedure to select a representative yet diverse set of melodies from a corpus of European folk songs. This procedure is detailed in Appendix B (p. 250). Experiment 5 (in 3.4) and Experiment 6 (in 3.5) use melodies in a tuning system unfamiliar to participants. Experiment 7 (in 3.6) and Experiment 8 (in 4.3) use a different set of melodies in a different tuning system, which was also unfamiliar to listeners. Experiment 9 (in 4.4) explores memory for pitch-only sequences; that is, using a set of melodies where each melody has the same isochronous rhythm. Experiment 10 (in 4.5) explores memory for rhythm-only sequences; that is, a set of stimuli where each note is sounded in the same pitch. These different melody corpora are presented to participants using the experimental procedure below.

Experimental procedure. In the 10 experiments of this dissertation, we exclusively used a continuous recognition paradigm to investigate recognition and interference in the context of memory for melody. Monophonic melodies were presented one after another, and participants were required to make some form of memory judgement on each trial. The majority of experiments required participants to make recognition judgements. In Experiment 4 (in 2.7) and Experiment 6 (in 3.5), however, we asked participants to indicate their perceived familiarity on a computer-based slider, rather than making a recognition judgement such as Old/New. This allowed us to observe recognition and interference in an indirect memory paradigm where participants were not informed about the memory nature of the experiment, and thus compared the results with the observations in direct recognition paradigms.

To investigate cumulative disruptive effects from the number of intervening items in memory for melody, we manipulated the number of intervening items in each experiment. In Experiment 1 (in 2.4) we started with small numbers of intervening items (i.e., up to 6). In the subsequent experiments, we investigated up to nearly 200 intervening items. Using the same experimental task in almost all experiments enabled us to make overarching comparisons between the data of each experiment. The results of this analysis are presented in Chapter 5 Below, we describe the general statistical approach.

1.2.2 General Statistical Approach

We predominantly used mixed effects models to investigate the cumulative disruptive effects of the number of intervening items on memory for melody. The advantage of mixed effects models over more traditional methods such as Analysis of Variance (ANOVA) is that mixed effects models take into account crossed random effects of subjects and melodies that possess different levels of memorability (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Judd, Westfall, & Kenny, 2012; Kass & Raftery, 1995; Kruschke, 2010, 2013; Nathoo & Masson, 2016). This is important for the present investigation in the context of music because participants show considerably different memory performance, and melodies differ dramatically in their memorability (Müllensiefen & Halpern, 2014). The mixed effects models were implemented in R (R-Core-Team, 2013) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2013).

To evaluate the models we either reported model comparisons (Kruschke, 2011; Wilks, 1938) or coefficient *p*-values of models with maximal random effect structure, as justified by each experiment's design (Barr, Levy, Scheepers, & Tily, 2013). We used Kenward-Roger approximations (Kenward & Roger, 1997) to adjust the degrees of freedom for linear models (Halekoh & Højsgaard, 2014a, 2014b). For generalized linear models, *p*-values as calculated by the lme4 package were reported. All models incorporated random intercepts for *Participant* and *Melody*.

Accounting for Participants' Response Biases. Recognition paradigms are prone to effects of response biases (Snodgrass & Corwin, 1988). For example, if a participant is correct for 80% of the repetition trials (hits) by accurately making an 'old' response, this is not sufficient evidence that the participant was capable of accurately recognizing melodies. This is because the participant may have strong biases towards pressing the 'old' button. This can be revealed by a very high percentage of 'old' responses toward new melodies (false alarms). Vice versa, if a participant only correctly responds 'old' to 40% of the melody repetitions, this is not sufficient evidence that the participant was incapable of recognizing melodies. This is because if this participant accurately rejects all new melodies (0% 'old' responses, which would be false alarms), then this participant demonstrates successful melody recognition. Therefore, both hit rates *and* false alarm rates should be taken into account in the analysis to deal with this phenomenon of response biases. Here, mixed effects models assessed whether overall performance was above chance. This was achieved by testing if melody repetition predicted significantly more 'old' responses than there were false alarms (on first stimulus presentation) while taking random participant response biases and melody variation into account. All figures that depict recognition performances used bias-corrected hit rates (unless specified otherwise). As a result of this transformation, performance at or near zero indicated incapability of recognizing the melodies (Snodgrass & Corwin, 1988).

Dynamic Response Tendency shifts. We also incorporated a novel procedure to control for shifts of participants' response biases over the course of an experiment. We labelled such shifts *dynamic response tendencies shifts* and we used the same terminology within the following chapters. As discussed above, static response biases were captured by the random intercept for participant. However, random intercepts cannot capture response tendency *shifts*. Dynamic response tendency shifts occur when participants' response tendencies shift over the course of the experiment. For example, a participant might be very liberal in classifying a melody as 'old' at the beginning of the experiment, however, as the experiment, progresses they become increasingly more conservative. Indeed, response tendency shifts are often observed in recognition experiments (Berch, 1976; Donaldson & Murdock, 1968) yet rarely accounted for in statistical models.

Therefore, we modelled response tendencies of each participant in each experiment using the participant's responses on trials that present melodies that have not been presented before. We assessed how the participant's behaviour on the new melody trials changed over the course of the experiment. The dynamic response probabilities were then incorporated as a participant-wise *Dynamic Response Tendency* fixed factor in the mixed effects models that predicted actual recognition performance on trials that presented melodies that had been presented before. Effectively, the procedure provided a dynamic baseline over the course of the experiment for each participant. It normalised the probability of producing a correct 'old' response with the base probability of a given participant producing an 'old' response in general as a result of their response tendency. Details of this procedure can be found in Section 2.4.1. We now turn to Chapter 2 which reports the first series of four experiments on recognition memory for melody.

Chapter 2

Resilient Memory for Melodies

Chapter 2 has been peer reviewed and accepted for publication:

Herff, S.A., Olsen, K.N., & Dean, R.T. (2017). Resilient memories for melodies: The number of intervening melodies does not influence novel melody recognition. *Quarterly Journal of Experimental Psychology*.

Note: Kirk N. Olsen and Roger T. Dean are the author's supervisors.

2.1 Abstract

In many memory domains, a decrease in recognition performance between the first and second presentation of an object is observed as the number of intervening items increases. However, this effect is not universal. Within the auditory domain, this form of interference has been demonstrated in word and single-note recognition, but has yet to be substantiated using relatively complex musical material such as a melody. Indeed, it is becoming clear that music shows intriguing properties when it comes to memory. This study investigated how the number of intervening items influences memory for melodies. Melodies were continuously presented, one per trial, and after each trial participants indicated whether they had heard the melody in the experiment before by responding ‘old’ or ‘new’ (Experiments 1, 2, and 3). In Experiment 4, participants rated perceived familiarity for every melody without being told that melodies reoccur. In four experiments using two corpora of music, two different memory tasks, transposed and untransposed melodies, and up to 195 intervening melodies, no sign of a disruptive effect from the number of intervening melodies beyond the first was observed. We propose a new ‘Regenerative Multiple Representations’ conjecture to explain why intervening items increase interference in recognition memory for most domains but not music. This conjecture makes several testable predictions and has the potential to strengthen our understanding of domain specificity in human memory, while moving one step closer to explaining the ‘paradox’ that is memory for melody.

2.2 Resilient Memory for Melodies

The number of intervening melodies does not influence novel melody recognition

Fast and accurate recognition of stimuli such as faces, names, smells, phone numbers, street signs, book titles, animals, food, and music relies on an important interplay between perception and memory. The demand of this difficult day-to-day task is exacerbated when we are faced with numerous items of similar nature. In human memory, cumulative interference from an increase in the number of intervening items impairs recognition performance in a variety of domains (Sadeh et al., 2014). For example, in written word recognition, a systematic decrease in recognition performance is observed as the number of intervening items increases (Poon & Fozard, 1980). Digit tasks (Donaldson & Murdock, 1968), letter trigrams (Olson, 1969), word lists and pairs (Bui et al., 2014; Hockley, 1992), faces (Rakover & Cahlon, 2001), and a wide variety of everyday visual objects (Konkle et al., 2010; Nickerson, 1965) have been thoroughly investigated with results that report similar behavioural phenomena.

However, evidence suggests that the detrimental effect of intervening items on recognition performance is not universal. In the visual domain, there is a large decrease in word recognition accuracy as the number of intervening words between the first and second presentation of a target word increases from 2 to 32 (Friedman, 1990b). Interestingly, this effect is not observed with drawings, where recognition performance remains constant between 2, 8, and 32 intervening items (Berman et al., 1991; Friedman, 1990a, 1990b). In the auditory domain, a disruptive effect of intervening items has been demonstrated in recognition tasks using spoken words and single-notes (Campeanu et al., 2014; D. Deutsch, 1970, 1975). In single-note recognition, interference arises from intervening pitches but not intervening spoken numbers (D. Deutsch, 1970). This suggests strong stimulus-specific interference effects in single-note recognition. However, cumulative disruptive effects have yet to be substantiated in other real-world auditory contexts, such as multi-note melody recognition in the musical domain. Indeed, as detailed below, it has been suggested that memory for melodies may be resilient to interference from intervening items (Dowling, 1991; Dowling, Kwak, & Andrews, 1995). Thus a systematic investigation of the possible interference effects in music and melodies is important to further elucidate memory's domain specificity (Fougnie, Zughni, Godwin, & Marois, 2015) and the question of whether memory for music is 'special' (Jackendoff & Lerdahl, 2006; Schulkind, 2009; Stevens, 2015). This paper conducts such an investigation.

2.2.1 Memory for Melody

Melodies are one of the most ubiquitous aspects of music and are crucial for musical enculturation (Corrigall & Trainor, 2014). We are exposed to many melodies on a day-to-day basis (Krause et al., 2014) and often many new melodies are heard before we encounter an old one again. Listeners who have experienced involuntary music imagery (e.g., earworms) know that simple melodic sequences of notes (i.e., a simple melody) have great potential for memorability (Bailes, 2007, 2015; Halpern & Bartlett, 2011; Williams, 2015). However, memory for melody has been described as a 'paradox' (Halpern & Bartlett, 2010; Schulkind, 2009; Stevens, 2015). It can be long lasting (Bailes, 2007; Baird & Samson, 2014; Cuddy & Duffin, 2005; Halpern & Bartlett, 2011; Jacobsen et al., 2015; Vanstone, Cuddy, Duffin, & Alexander, 2009), yet "memory for novel melodies is surprisingly poor" (Lange & Czernochowski, 2013, p. 137) and "(…) even the simplest kind of recognition test for melodies shows how poor musical memory can be, in comparison to other kinds of memory" (Halpern & Bartlett, 2010, p. 234). These findings raise important questions about the nature of fundamental memory phenomena in memory for melody. Here, we systematically investigate two fundamental memory phenomena in memory for melody: recency-in-memory and cumulative disruptive effects from the number of intervening melodies (both discussed in detail below).

2.2.2 Interference and Temporal Delay in Novel Melody Recognition

A potential source of confusion in the literature on memory for music is that different forms of memory – and the conditions under which memory is measured – are often not well articulated. This means that apparent contradictory evidence may occur when in fact different results arise because researchers evaluate different forms of memory under different experimental conditions. In the context of memory for melody in particular, apparent contradictory evidence such as the claims reviewed above, that memory for melody is sometimes 'surprisingly poor' yet sometimes 'surprisingly good', provides one such example. To avoid this potential source of confusion in the present study, it is important to note that we specifically investigate continuous recognition of novel monophonic melodies that resemble melodies that are encountered in normal day-to-day experiences.

Recently, Schellenberg and Habashi (2015) showed that recognition performance of novel ~30 s melodies is not disrupted by the mere passing of time over periods up to a week. However, their study focused on the passing of time and not on whether the number of intervening items influenced melody recognition. This is important because memory decay over time is only one of two main mechanisms of forgetting (Eysenck & Keane, 2015). Interference is the other mechanism, which describes a continuous decrease in memory performance, not because of the passing of time, but from the additional information that is learned after or even before the encoding of a stimulus into memory. Establishing the extent of interference as well as effects of decay in memory for melody provides an important step towards a full account of memory for melody in particular, and memory for complex auditory information in general. This will, in turn, assist in a comprehensive understanding of memory phenomena observed in the context of music; for example, the observation that memory for music is often 'spared' in people who suffer from debilitating deficits in memory, such as those associated with dementia and severe brain injury (Baird & Samson, 2014; Cuddy & Duffin, 2005; Jacobsen et al., 2015; Schulkind, 2009).

In the context of interference and decay, a series of melody recognition experiments conducted by Dowling and colleagues (Dowling, 1991; Dowling et al., 1995) used very short (~ 3 s) melodies that recurred after periods of silence or periods filled with other melodies. Results indicated that melody recognition predominantly relied on pitch-interval information when delays are filled with other intervening melodies. Pitch-interval information refers to the relative pitch distance between notes, rather than their absolute pitches (i.e., pitch intervals like "a major third" rather than absolute pitches like 'C4' and 'E4'). After observing relatively stable memory performance over the first two minutes, Dowling et al. (1995) proposed that there might be underlying processes involved in pitch-interval based recognition of novel melodies that may be resilient to the presentation of intervening items (in this case, melodies). Such resilience to the effects of interference in memory for melodies is investigated here by systematically manipulating the number of intervening melodies in four melody recognition experiments.

2.2.3 Pitch Information and Melody Transposition

As mentioned above, several kinds of pitch information are important for accurate melody recognition: surface information such as absolute pitch and abstract information of pitch relations expressed, for example, as pitch intervals or pitch contour (Bartlett & Dowling, 1980; Dowling & Fujitani, 1971; Krumhansl, 2000; Levitin, 1994; Schellenberg, Stalinski, & Marks, 2014). Transposition of a melody into another key only retains relative pitch information. While participants are capable of recognizing untransposed melodies, recognition of transposed melodies tends to be worse (Dowling & Fujitani, 1971; Plantinga & Trainor, 2005; Schellenberg et al., 2014). Some literature suggests that, for delays beyond one minute, memory for melody predominantly uses relative pitch information (Bachem, 1954) (see also, Krumhansl, 2000, for a review). However, this does not necessarily imply that memory is resilient to transposition. Indeed, more recent research shows that transposing melodies also disrupts memory for melody after longer delays of up to a day (Schellenberg & Habashi, 2015).

Taken together, these findings suggest that interference in melody recognition might depend on whether or not both absolute or relative frequency information is retained. This hypothesis is investigated in the present study. Specifically, Experiments 1 and 2 investigate melody recognition with the second occurrence of each melody untransposed, whereas in Experiments 3 and 4 the second occurrence of each melody is transposed.

2.2.4 Task Awareness in Melody Recognition

Participants' awareness of the nature of a musical memory task can potentially influence performance (Halpern & Bartlett, 2010). For example, age-related decline in memory appears to be smaller with indirect (Fleischman, Wilson, Gabrieli, Bienias, & Bennett, 2004) rather than direct memory tasks (Gaudreau & Peretz, 1999; Halpern & O'Connor, 2000). Therefore, the four experiments reported here used two different continuous memory tasks with varying degrees of task-specific awareness (Shepard & Teghtsoonian, 1961).

Continuous memory tasks are commonly used to investigate recognition performance and the disruptive effects of intervening items (Berman et al., 1991; Campeanu et al., 2014; Ferris et al., 1980; Friedman, 1990b; Hockley, 1992; Sadeh et al., 2014). In such a task, stimuli are presented successively and participants are asked to respond after each item; most commonly, whether they perceive the present stimulus to be 'old' or 'new' (Shepard & Teghtsoonian, 1961). In this case, 'old' refers to items that have been presented before, and vice versa. Importantly, the memory-related aspects of a continuous memory task can be expressed with varying degrees of memory-task awareness. For example, the 'old' or 'new' request is a direct memory task, while rating the familiarity of each melody is an indirect memory task. Therefore, in the present study, Experiments 1, 2, and 3 implemented direct memory task instructions, whereas Experiment 4 implemented indirect memory task instructions.

2.2.5 Recency-In-Memory

In a continuous memory task, the case of zero intervening melodies is the only condition that reflects immediate repetition. Due to recency-in-memory effects, we expected that immediate repetition of a melody should lead to higher memory performance when compared to any case that has at least one intervening item (Berz, 1995; Dowling, 1973; Greene & Samuel, 1986; Jahnke, 1963; Roberts, 1986). Indeed, recognition performance should be lower for all other numbers of intervening items, as the first intervening melody will disrupt any recency-in-memory effect. No observed effect of the number of intervening melodies beyond the first would also support the findings of Schellenberg and Habashi (2015), who reported no significant disruptive effect of temporal delay on melody recognition. Here, Experiment 1 specifically investigates a small number of intervening items (up to 6) to compare the zero intervening melody condition to each of the other intervening melody conditions. Experiments 2, 3, and 4 use a far larger number (up to 197) of intervening melodies to investigate longer-term unfolding cumulative disruptive effects on recognition. In the following section, we provide a summary of the aim, design, and hypothesis of the study as a whole.

2.3 Aim, Design, and Hypotheses

Four experiments were designed to further elucidate mechanisms that underpin memory for novel melodies by investigating the effect of distinctively different intervening melodic items (details in the stimulus section). The importance of absolute and relative frequency information was also investigated by implementing melody transposition, and the potential for differential outcomes between memory tasks was investigated by measuring both recognition and perceived familiarity.

Following the literature reviewed above, a significant performance decrease as the number of intervening items increases is evidence that memory for melody is similar to other memory domains. A lack of such an effect supports Dowling, et al.'s (1995) suggestion that memory for melody might be special because other intervening melodies do not disrupt it. Furthermore, such results would also provide evidence of the domain specificity of human memory. Finally, immediate repetition of a melody was hypothesized to lead to better performance than when a melody's repetition occurs after any number of different intervening melodies.

2.4 Experiment 1: Melody Recognition with Zero to Six Intervening Melodies

As a starting point, Experiment 1 was designed to investigate how the first six intervening melodies affect melody recognition performance. Specifically, the number of intervening items varied between zero and six in a continuous recognition paradigm while recognition performance was measured. The choice of six intervening item conditions was deemed a sufficient starting point to demonstrate cumulative disruptive effects, as has been the case in other domains such as written and spoken word recognition (Campeanu et al., 2014; Friedman, 1990b).

2.4.1 Method

Participants. Twenty-eight undergraduate students were recruited from the Western Sydney University (M_{age} = 22.1 years, SD_{age} = 8.5, seven male). Recruitment criteria were less than two years of musical training and no hearing impairments. Participants volunteered for this experiment and were reimbursed with chocolate and the opportunity to learn about the research after the experiment.

Stimuli and Equipment. Testing took place in a sound attenuated booth at the MARCS Institute for Brain, Behaviour, and Development, Sydney, Australia. The experiment was programmed in E-Prime (Psychology Software Tools, 2012). Stimuli were presented through Sennheiser 25 HD headphones at a volume comfortable to the user. Experiment 1 focused on recognition of novel melodies, and 60 novel monophonic melodies were composed by the first author of this article. All melodies were 12 seconds in duration and unmistakably tonal. The melodies were composed in 12-tone equal temperament, the most common tuning system used in Western tonal music (Milne, Sethares, & Plamondon, 2007). The key for each melody was randomly chosen before composition; half of the melodies were composed in major and the other half in minor. All notes were sounded with the same grand piano timbre at the same velocity using Pianoteq 4 STAGE (Version 4.2). The meter was balanced across the melodies between 4/4 and 3/4, the two dominant meters in Western tonal music (London, 2012). The tempi were pseudo randomized between 80-165 beats per minute (bpm) with a mean bpm set to the most common 120 bpm (Franek, van Noorden, & Rezny, 2014; Moelants, 2002). Not all tempi between 80 and 165 bpm were possible to realize, as the meter was fixed at 4/4 and 3/4 and the duration fixed at 12 seconds. The rhythmic structure was kept simple with not more than two levels of metrical division (Winold, 1975). Musical scores relating to representative examples of melodies presented in Experiment 1 can be found in Appendix A (p. 249). The online supplement S1-Stimuli.zip contains the stimuli of all Experiments as well as a musical feature analysis of all melodies (more detail is provided in the Stimuli and Material section of Experiment 3).

Even though all melodies were novel, there was a slight possibility that some melodies might have resembled familiar tunes. Therefore, a pilot study was conducted and every melody that was perceived to sound similar to another known melody by at least one participant was removed from the corpus. Twelve researchers from the MARCS institute who were not involved in this project volunteered to participate in the pilot study. All participants were oblivious to the origin of the stimuli. A total of five melodies were removed from the initial pool of 60. Representative examples of the remaining stimuli can be seen in Appendix A. An uninvolved expert listener with an extensive and sophisticated background in music (*Ollen Musical Sophistication Index* of 845, see Ollen, 2006, where on a scale of 0-1000, >500 is deemed to be musically sophisticated) described the melodies as follows:

> *"(…) I guessed they were theme tunes from TV programs, film music, or adverts. They sounded like the sort of melodies one would typically come across in everyday life."*

Procedure. Up to three participants could participate in the experiment simultaneously, but each participant could neither hear the stimuli presented to others, nor see the other participants as they completed the task. Participants provided informed consent and sat comfortably in front of a computer. All instructions were presented on the computer screen. Participants were instructed that "*in this experiment, you will hear many different melodies, one after another. However, sometimes a melody will be repeated. Each melody may repeat more than once OR may not repeat at all*." They were asked to "*listen to each melody and respond to whether you have heard this melody before in this experiment*." Responses were made using two different keys of the keyboard (the 'A' key and the '-' key), one associated with '*New*' and the other with '*Old*'. The response keys were counterbalanced between participants. While the melodies were played, the screen showed 'Listen!' in black letters on a white background. As soon as the melody was finished, the 'Listen!' text disappeared. The next trial was initiated as soon as a participant gave a response. Participants had the opportunity to practice the task in six practice trials and were allowed to adjust the volume to their personal preference during the practice trials. However, the volume was then fixed to their chosen level for the main experiment trials. After completion of the experiment, participants were asked to fill out a short demographic questionnaire. A schematic example of a sequence of trials is shown in *Figure 2.1*.

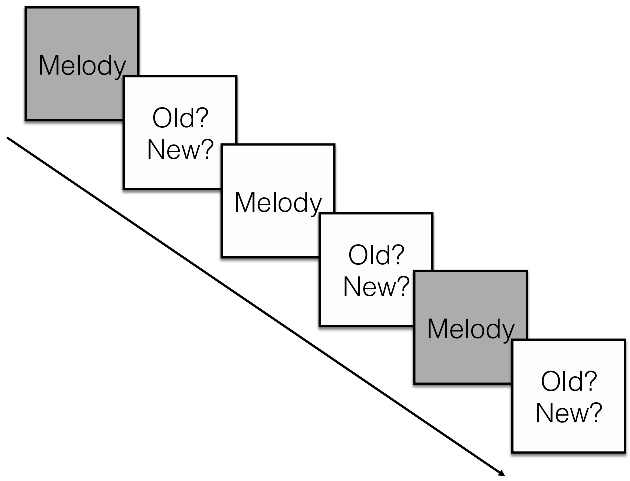


Figure 2.1: Schematic example of a sequence of trials in Experiment 1, with 'Melody' representing a stimulus and 'Old? New?' representing a participant's response. The two grey fields represent the same melody. Therefore, this shows an example of one intervening melody between the first presentation of the 'grey' melody and its repetition. Here, we systematically manipulate the number of intervening melodies until a target melody repeats (see text for details).

The number of intervening melodies was manipulated between zero and six. To avoid list order effects, the order of the melodies was randomized for every participant. In a continuous recognition task this means that both embedded-trials and overlap-trials are possible. When trials are embedded within each other, then a larger intervening-items condition contains first and second melody presentations of a smaller intervening-item condition. When trials overlap, much like the links in a chain, the intervening items between the two presentations of a target melody contains the first, but not the second presentation of another target melody. For example, if A1 and B1 are the first presentations of melodies A and B, and A2 and B2 are the second presentations of melodies A and B, then both scenarios are possible: A1B1A2B2 (overlap), A1B1B2A2 (embedded). Since the number of intervening melodies was manipulated from zero to six with the same number of trials in each gap size, the possible permutations of list order were highly constrained. As a result, 49 of the 55 melodies were presented twice over the course of the experiment with controlled numbers of intervening items. The remaining six melodies were used as 'dummy' melodies. Dummy melodies randomly filled the remaining item list slots and were not included in the analyses. Overall, every participant listened to 130 trials that included seven melodies in each of the different numbers of intervening melody conditions (0, 1, 2, 3, 4, 5, 6), and 32 dummy trials that were not included in the analysis but enabled a fully randomized list order for each participant.

Accounting for Participant Response Tendencies. Recognition paradigms are prone to effects of response biases (Snodgrass & Corwin, 1988). For example, there is evidence that response tendencies can change over the course of an experiment (Berch, 1976; Donaldson & Murdock, 1968). In order to take this into account, we built participant-wise *Dynamic Response Tendency* models that predict the baseline tendency for each participant to press 'old', and how this tendency changes throughout the experiment. These generalized linear mixed effects models were trained on 'old' responses on first presentations based on trial number. The fitted models were then used to predict the probability of pressing 'old' on repetition trials based solely on trial number. As a result, the Dynamic Response Tendency models were used in statistical analyses in each Results section to account for the individual response tendencies of each participant, and the way such tendencies may change over the course of the experiment.[1]

Statistical Approach. The statistical approach is similar in all experiments and will be detailed here. We used generalized linear mixed effects models to investigate the effect of the number of intervening melodies on binary melody recognition data (Experiments 1, 2, and 3) and linear mixed effects models for continuous perceived familiarity data (Experiment 4) (Baayen et al., 2008). This approach takes into account crossed random effects of subjects and melodies that possess different levels of memorability (Baayen, 2008; Baayen et al., 2008; Judd et al., 2012; Kass & Raftery, 1995; Kruschke, 2010, 2013; Nathoo & Masson, 2016). The models were implemented in the R software platform (R-Core-Team, 2013) using the lme4 package (Bates et al., 2013).

A model comparison approach was used to compare evidence in favour of the null hypothesis relative to evidence in favour of a competing hypothesis (Kruschke, 2011). Bayesian information criteria (BIC) (Schwarz, 1978) are reported. BIC values, which penalize additional parameters strongly, were used as the basis of model selection. Differences in BIC values between models are reported as ΔBIC for significant model improvements. A ΔBIC of two or greater is assumed "positive" in favour of the model with lower BIC. A ΔBIC difference of six or greater is

---

[1] Detailed assessment showed that there were no response tendency shifts in Experiments 1, 2, and 4 (all dynamic response tendency coefficient p-values > .20). Experiment 3 showed dynamic response tendency shifts over the 196 trials ($p < .001$). The model controlled for the shifts. The nature of the response tendency shifts in Experiment 3 is detailed in Appendix C.

considered “strong” evidence (Kass & Raftery, 1995). ΔBIC can be used to estimate the Bayes factor, a measurement of how much evidence there is supporting one hypothesis or model compared to another (Kass & Raftery, 1995; Lewandowsky & Farrell, 2010, p. 186; Nathoo & Masson, 2016; Wagenmakers, 2007). A ΔBIC of six represents a Bayes factor of twenty, which can be interpreted as twenty times more evidence for the model with the lower BIC. A Bayes factor cannot only provide evidence against the Null-Hypothesis, but also evidence for it. Furthermore, direct model comparisons using goodness-of-fit were conducted with likelihood-ratio tests, with *p*-values reported throughout (Wilks, 1938). The models were provided with a random intercept for *Melody* in order to account for possible effects of individual melodies. Accounting for participant response tendencies was achieved in each model by implementing random intercepts for *Subject* as well as a fixed factor for the aforementioned *Dynamic Response Tendency* models.

At the beginning of each results section, mixed effects models were used to assess whether overall performance is above chance. This was achieved by testing if melody repetition predicts significantly more ‘old’ responses (Experiments 1-3) or higher perceived familiarity (Experiment 4) while taking random participant response biases and melody variation into account. Coefficient *p*-values are reported in the beginning of each results section. All figures that depict recognition performances in each Results section below report response bias-corrected hit rates by subtracting the participant-wise false alarm rates from participant-wise hit rates. As a result of this transformation, performance around zero indicates the inability to recognize the melodies (Snodgrass & Corwin, 1988).

At the end of each Results section for each experiment, a final assessment of the effects of the number of intervening melodies is reported using models with maximal random effects (Barr et al., 2013). These models include random slopes for participants over the number of intervening items; that is, all random effects that could possibly play a role given the experimental design. For generalized linear models, p-values as calculated by the lme4 package are reported. For linear models, conservative Kenward-Roger approximations were used to adjust the degrees of freedom (Kenward & Roger, 1997) using the R package pbkrtest (Halekoh & Højsgaard, 2014a, 2014b).

Results

Figure 2.2 shows mean false alarm rates and hit rates for every participant and melody, and bias-corrected hit rates are depicted in *Figure 2.3*. In summary, participants performed significantly above chance ($Z = 24.387$, $p < .0001$) and the number of intervening melodies did not influence recognition performance beyond the first intervening item.

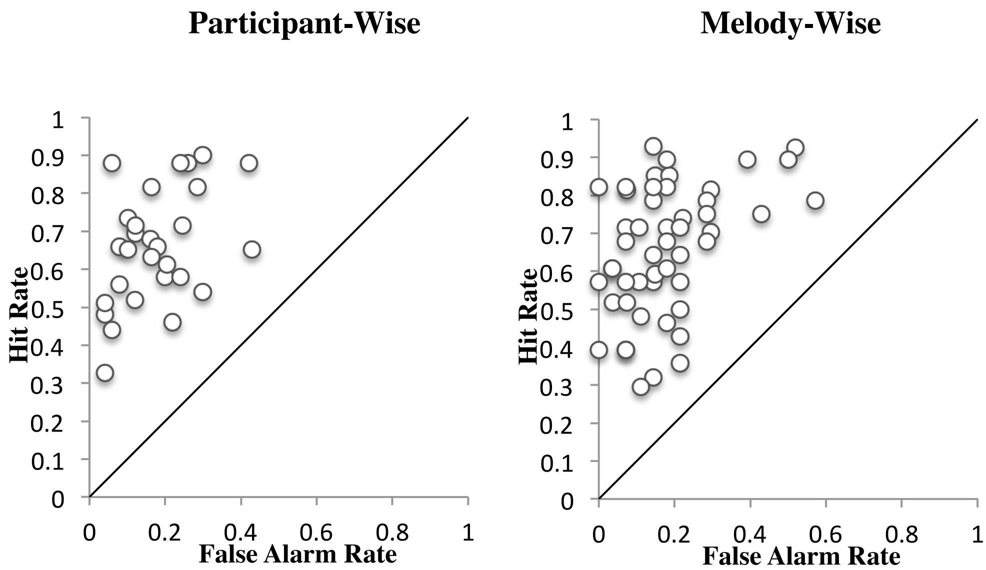


Figure 2.2: Hit rates and false alarm rates in Experiment 1. The left panel shows mean data participant-wise, and the right panel shows mean data melody-wise. The reference line represents chance level. Mixed effects models were built to take inter-melody and inter-participant variation into account.

Specifically, a generalized linear mixed effects model was constructed to investigate how the number of intervening items influences melody recognition. The model is built to predict 'old' responses on melody repetitions. The base model included the systematic factor for *Dynamic Response Tendency* and random intercepts for *Subject* and *Melody* (*BIC* = 1659.7; *LogLik* = -815.38), and improved significantly when provided with information about the *Number of Intervening Items* (*BIC* = 1655.8; *LogLik* = -791.75, $p < .0001$, ΔBIC = 3.9). Coefficient estimation for each number of intervening melodies confirmed a significantly lower recognition performance for all numbers of intervening items beyond immediate repetition (all *p*-values < .0001). A base model excluding immediate repetitions (*BIC* = 1442.7; *LogLik* = -707.17) did not improve when provided with the number of intervening items (*BIC* = 1474.7; *LogLik* = -705.48, $p > .64$). This shows that the number of intervening items beyond the first and up to six does not carry predictive value when it comes to melody recognition performance. This result is illustrated in *Figure 2.3*.

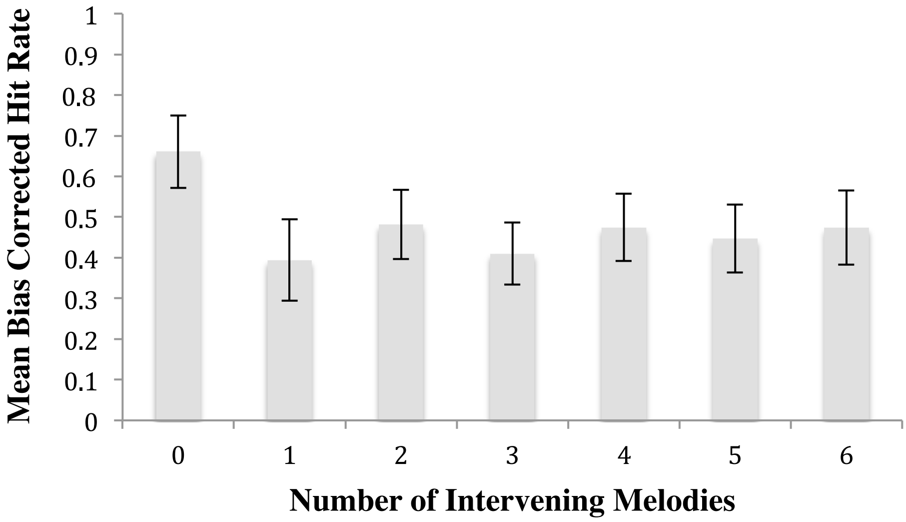


Figure 2.3: Mean bias corrected hit rate for all seven conditions of intervening melodies in Experiment 1. Note that zero represents chance recognition level. The zero intervening melodies condition produced significantly higher recognition performance than all other conditions. Performance was statistically identical in one through to six intervening melodies. Error bars show 95% confidence intervals.

A final statistical assessment of the effect of number of intervening items was made using the maximal random effect structure. The number of intervening melodies did not improve models that excluded data relating to immediate repetition ($p > .95$). As hypothesized, significant differences between melodies were observed, as shown by a significant decrease in model performance when the random intercept for melody was removed ($p < .0001$, ΔBIC = 68.1). For comparison, *Figure C.0.4* in

Appendix C displays uncorrected hit-rates for each of the different intervening item conditions in Experiment 1.

2.4.2 Discussion

Experiment 1 showed that beyond the effect of the first intervening melody, up to 6 additional intervening melodies did not significantly decrease recognition performance. This finding cannot be explained by a floor effect that would prevent further decreases in performance after the first intervening melody, as recognition performance was above chance in all six intervening melody conditions (see *Figure 2.2*). The results are consistent with Schellenberg and Habashi's (2015) findings of a lack of disruptive effects of intervening time on melody recognition. Using relatively long (12 s) melodies, Experiment 1 also supports the hypothesis of underlying processes involved in melody recognition that bypass the interference by intervening items found in other memory domains (Dowling et al., 1995).

Recent research using five four-part chord progressions followed by three-note arpeggiated continuations has shown that it can take approximately 20 s to reappraise a prior melody as a whole; that is, 20 s to integrate a melody into long term memory (Bailes, Dean, & Pearce, 2013). Buchsbaum, Padmanabhan, and Berman (2011) investigated this issue in auditory-verbal stimuli by combining a continuous recognition paradigm and fMRI. They found systematically different activation patterns in response to conditions that presented ≥ 4 intervening items. So it could be that the disruptive effect of the number of intervening melodies only manifests after the critical period of 20-30 s. In this case, the few conditions in Experiment 1 (4, 5, and 6 intervening melodies) that extended up to and beyond this critical period may not be sufficient to properly rule out this possibility. Therefore, there may still be an interference effect from the number of intervening melodies on melody recognition that occurs beyond those used in Experiment 1. This possibility was investigated in Experiment 2.

2.5 Experiment 2: Melody Recognition with Four to Thirteen Intervening Melodies

Experiment 2 aimed to replicate and extend the results of Experiment 1 with a larger number of intervening items. Experiment 2 systematically investigated four to thirteen intervening melodies. This ensured that intervening melodies were always presented beyond the potentially critical period of 20 to 30 seconds after stimulus presentation. Following from the results of Experiment 1, no differences in recognition performance were hypothesized for any number of intervening melodies.

2.5.1 Method

Participants. Experiment 2 tested 20 participants from the Western Sydney University (M_{age} = 21.1 SD_{age} = 3.6, five male) who provided informed consent and were reimbursed with chocolate. The data of three participants were excluded because two reported a high level of musical expertise (active musicians), and one made the same response (identical key press) in each trial. All participants reported normal hearing and did not participate in Experiment 1.

Stimuli and Equipment. Stimuli and equipment were identical to Experiment 1; the only difference was a reduction from 55 to 50 randomly chosen melodies from the set of 60 melodies described in Experiment 1 in order to decrease the experiment's duration. Musical scores relating to representative examples of melodies presented in Experiment 2 can be found in Appendix A (p. 249). The online supplement S1-Stimuli.zip contains the stimuli of all Experiments as well as a musical feature analysis of all melodies.

Procedure. The procedure of Experiment 2 closely followed that of Experiment 1. However, instead of zero to six intervening melodies, Experiment 2 presented four to thirteen intervening melodies. For each participant, a list of 100 trials was randomly populated with melodies that reoccurred once after four to thirteen intervening melodies. Each number of intervening melodies occurred at least four times for each participant. This ensured at least 80 analyzable trials with 40 different melodies for each participant. The remaining 20 trials were filled without restraints on the number of intervening melodies in order to allow sufficient permutations of list order. Similar to the procedure in Experiment 1, dummy melodies randomly filled any remaining item list slots and were not included in the analyses. Instructions were identical to Experiment 1.

2.5.2 Results

Figure 2.4 shows mean false alarm rates and hit rates for every participant and melody, and bias-corrected hit rates are depicted in *Figure 2.5*. Again, participants performed significantly above chance ($Z = 15.97$, $p < .0001$), with results that replicated and extended those reported in Experiment 1. The number of intervening melodies did not influence recognition performance when up to thirteen intervening melodies are presented.

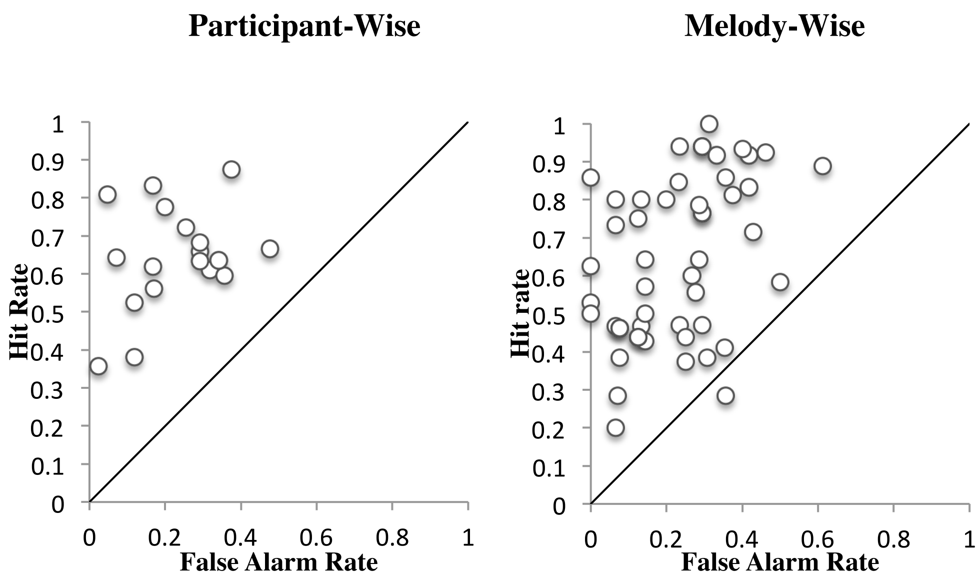


Figure 2.4: Hit rates and false alarm rates in Experiment 2. The left panel shows mean data participant-wise, and the right panel shows mean data melody-wise. The reference line represents chance level.

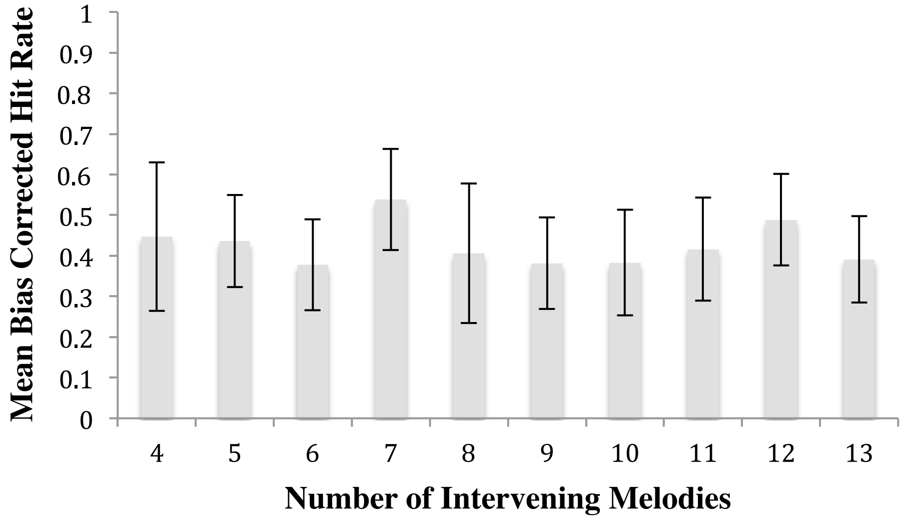


Figure 2.5: Mean bias corrected hit rate for all 10 conditions of intervening melodies in Experiment 2. Note that zero represents chance level. Recognition performance was statistically identical in all intervening melody conditions. Error bars show 95% confidence intervals.

In statistical terms, a mixed effects model predicting 'old' responses on repetition trials (*BIC* = 929.83; *LogLik* = -451.67) with a random intercept of *Subject* and *Melody* and a systematic factor for *Dynamic Response Tendency* did not improve when provided with the information of the *Number of Intervening Melodies* (*BIC* = 936.13; *LogLik* = -451.51; $p > .56$). This shows that the number of intervening items cannot be used to predict recognition performance in memory for melody between 4 and 13 intervening melodies. This effect is also illustrated in *Figure 2.5*.

A final statistical assessment of the effect of number of intervening items was made using the maximal random effect structure. *Number of Intervening Melodies* did not yield a significant result ($p > .50$). As hypothesized, the model performed significantly worse without a random intercept for *Melody* ($p < .0001$, ΔBIC = 47.72). This result is consistent with Experiment 1.

2.5.3 Discussion

Experiment 2 extends the results of Experiment 1, with no significant differences observed when the number of intervening melodies increased to thirteen. Thirteen intervening melodies equates to a temporal delay of about two and a half minutes. Thus, no differences in recognition performance between four and thirteen intervening melodies indicate that the passing of time has no effect on melody recognition performance within the first few minutes of melody recognition. This is consistent with previous research in the musical domain (Schellenberg & Habashi, 2015). However, this result is somewhat surprising from non-musical memory research, because a variety of domains do show interference effects of time (Bui et al., 2014; Campeanu et al., 2014; Hockley, 1992; Konkle et al., 2010; Olson, 1969; Sadeh et al., 2014) that include other auditory stimuli such as words (Buchsbaum et al., 2011; Campeanu et al., 2014).

The second presentation of each melody in Experiment 1 and Experiment 2 was physically identical to the first presentation. Therefore, both absolute and relative pitch (or frequency) information were available in the task (Bartlett & Dowling, 1980; Dowling & Fujitani, 1971; Krumhansl, 2000; Levitin, 1994; Schellenberg et al., 2014). Though relative pitch information seems to be predominantly used in long-term memory for melodies (Krumhansl, 2000), recent studies show that absolute pitch information is also retained (Schellenberg et al., 2014). One possible explanation for our findings in Experiments 1 and 2 could be that absolute frequency information compensated for an interference effect of the number of intervening melodies. It may be that an effect of intervening melodies will be observed when only relative frequency information is available. To test this possibility, two additional experiments were designed with melodies that were transposed on their repetition.

The results of Experiment 1 and Experiment 2 could also be limited to the specific corpus of melodies that resembled advertisement jingles or movie themes. To investigate whether the results can be replicated in other melody corpora, Experiments 3 and 4 also used a new set of melodies taken from a large corpus of European folk songs. Furthermore, prior research in the visual domain has shown cumulative disruptive interference for complex and meaningful visual stimuli (photographs) between 40, 80, 120, 160, and up to 200 intervening items (Nickerson, 1965). Potentially, the slope of the cumulative disruptive interference in memory for melodies may be too shallow to be detected with only a span of 13 intervening items. Similar to complex and meaningful visual stimuli, cumulative disruptive interference may emerge when investigating larger scale differences in the number of intervening items. While Experiments 1 and 2 investigated a relatively early time course of memory for melody, Experiments 3 and 4 were designed to investigate long-term effects of intervening melodies with the use of up to 195 intervening melodies.

2.6 Experiment 3: Melody Recognition with up to 195 Intervening Melodies

Experiment 3 was designed to investigate the influence of large numbers of intervening items when absolute frequency information is removed through melody transposition. Furthermore, the experiment investigated potential long-term effects of the number of intervening melodies on melody recognition with a new stimulus set derived from a large corpus of European folk songs.

2.6.1 Method

Participants. Thirty-two undergraduate students from the Western Sydney University (M_{age} = 21.03 SD_{age} = 5.64, six male) participated in this experiment. Participants were required to have received less than two years of musical training (five participants had musical training M = 3.4 years, SD = .90). Participants reported normal hearing and did not participate in the previous experiments. Participation was reimbursed with course credit as part of university course requirements.

Stimuli and Material. Experiment 3 used an exhaustive stimulus selection procedure to create the final set of stimuli. A detailed description of this procedure can be found in Appendix B (p. 250). A European folk song corpus of 8,397 monophonic melodies was analyzed (CCARH; Sapp, 2005). All melodies were deconstructed into their underlying musical features using the MIDI Toolbox (Eerola & Toiviainen, 2004a; 2004b, p. 96), FANTASTIC (Müllensiefen, 2009, p. 37), as well as several self-implemented routines to measure tonality (Dean, Bailes, & Drummond, 2014), autocorrelation between pitch values (Dean, Bailes, & Dunsmuir, 2014; Dean & Dunsmuir, 2015), and pitch as well as rhythmic balance and evenness (Milne, Bulger, Herff, & Sethares, 2015). Deconstructing melodies into their underlying musical features was necessary to ensure in a later step that the final subsample of melodies adequately represented the underlying corpus in respect to various musical features. An in-depth description of the musical features can be found in the MIDI Toolbox and FANTASTIC manuals (Eerola & Toiviainen, 2004a; 2004b, p. 96; Müllensiefen, 2009, p. 37). Due to the vast number of musical features (116), a principal component analysis was used to reduce the dimensions. Twenty-one significant underlying components were identified using a permutation based Monte Carlo Simulation (Parallel analysis) with a 95% confidence level (Clarkson & Jennrich, 1988; O'Connor, 2000). The score on every principal component was calculated for every melody in order to cluster the melodies in the dimension-reduced space. A hierarchical cluster analysis using Euclidean distances resulted in one large cluster. A cluster analysis using reduction in log-likelihood was implemented as a distance measurement and a nine-cluster solution emerged from the single large cluster using Bayesian Information Criteria (average silhouette measure of cohesion and separation = .01 with a smallest to biggest cluster ratio of 2.587; (Rousseeuw, 1987). The nine-cluster solution was accepted.

In total, 110 melodies were randomly drawn, and the specific number of melodies from each cluster was determined by the relative cluster size. These 110 melodies were then subjected to a perceptual pilot study in order to identify melodies that, despite being European folk songs, evoke high degrees of familiarity from Australian listeners. Twelve ($M_{age} = 33$, $SD_{age} = 12.5$) participants provided familiarity ratings on a 100-point visual analog scale. The familiarity response distributions for every individual melody were compared to the response distribution of all other melodies using non-parametric Kolmogorov-Smirnov tests. Outlier melodies that showed significantly different response distributions were removed because they could not be classified as ‘novel’ for the purpose of the experiment. This procedure resulted in 98 melodies that were mathematically derived from a large corpus of European folksong melodies and perceptually tested to be novel to Australian listeners. These 98 melodies had a mean duration of 10.86 s and a mean pitch range of 7.93 semitones. Musical scores comprising representative examples of melodies presented in Experiment 3 can be found in Appendix A (p. 249). Stimuli from all experiments can be found in the online supplement S1-Stimuli.zip.

Procedure. The procedure of Experiment 3 closely followed the procedure of the previous experiments and is visualized in *Figure 2.6*. Participants provided informed consent and instructions were presented in a standardized format on the computer. A demographic questionnaire was also digitally administered. Participants were instructed to *"carefully listen to the melodies and, after each melody, please respond whether you have heard it before in this experiment"*. After a melody was presented, two buttons appeared on the screen, one labelled "*New*" and the other one "*Old"*. Responses were made using the computer mouse[2]. The 98 melodies were presented in 196 trials. Every participant received a unique randomization list that was automatically generated at the beginning of the experiment. Melodies occurred twice throughout the experiment and, similar to previous studies (Schellenberg et al., 2014), were transposed on their repetition by six semitones up (to the key most distant from the original)[3].

---

[2] Note that in addition to the above-described procedure, remember/ know judgments (Tulving, 1985; Yonelinas, 2002) and confidence ratings were also measured after each participants' recognition response (data not shown).

[3] Data from a pilot study as well as Experiments 3 and 4 show that participants did not use the pitch height of melodies as the basis of their judgments. Several mixed effects models were built to evaluate this. The melodies varied in average pitch height. If pitch height had been used as a systematic cue to respond 'old' then melodies with higher average pitch height would show an increased number of 'old' responses. However, average pitch height variation (as calculated by FANTASTIC) between the melodies did not carry predictive value for familiarity or recognition judgments during the second

Up to three participants were tested simultaneously, but each participant could neither hear the stimuli presented to others, nor see the other participants as they completed the task. The experiment was programmed in Max MSP 6.0 and executed in Max Runtime 6.0 (Cycling74, 2014) and took approximately 45 minutes to complete.

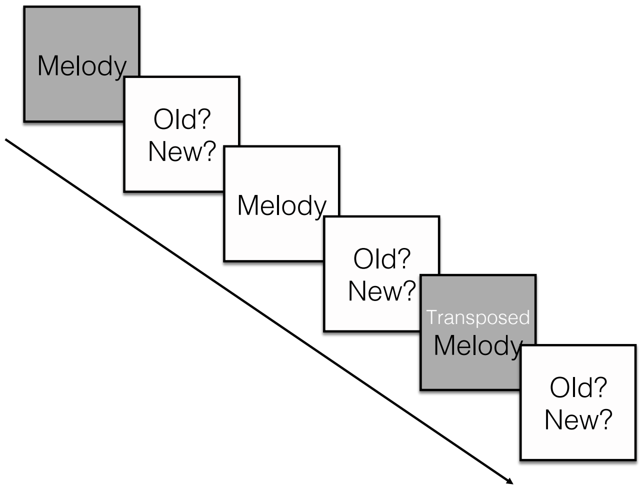


Figure 2.6: Schematic example of a sequence of trials in Experiment 3. The two grey fields represent the same melody; however, the second presentation of each melody (the 'grey' melody in this figure) in Experiment 3 was transposed. Therefore, this shows an example of one intervening melody between the first presentation of the 'grey' melody and its transposed repetition.

2.6.2 Results

Figure 2.7 shows mean false alarm rates and mean hit rates for every participant and melody. *Figure 2.8* shows the probability of Bias-Corrected Hits with increasing number of intervening items. Overall, participants performed significantly

---

presentations (all p-values > .18). This means that higher average pitch height does not systematically shift response tendencies toward old-responses.

above chance ($Z = 12.77$, $p < .0001$) with results that replicated and extended those reported in Experiments 1 and 2: the number of intervening melodies did not influence recognition performance when up to 195 intervening melodies were presented with a different melody corpus and with transposition on their second presentation.

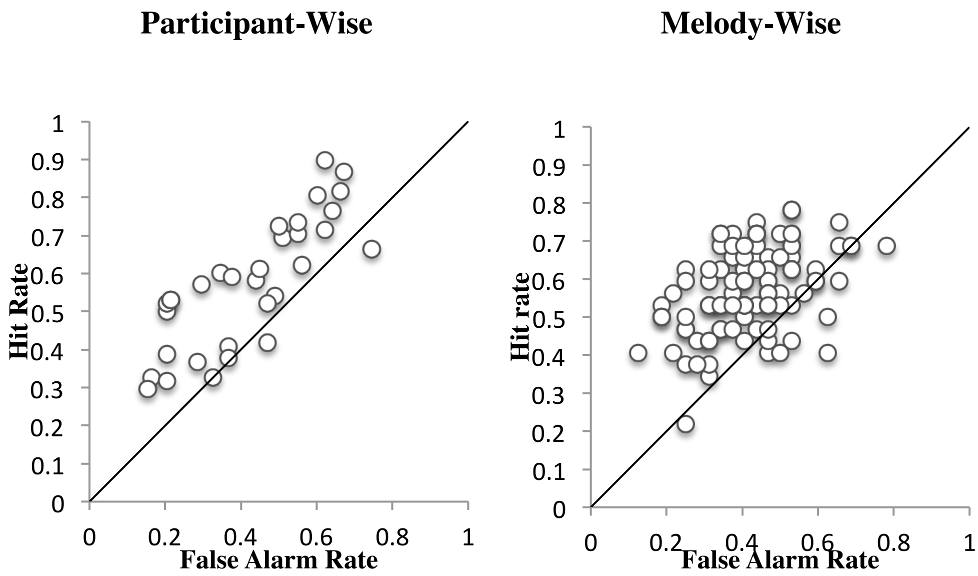


Figure 2.7: Hit rates and false alarm rates in Experiment 3. The left panel shows the data participant-wise, and the right panel melody-wise. The reference line represents chance level.

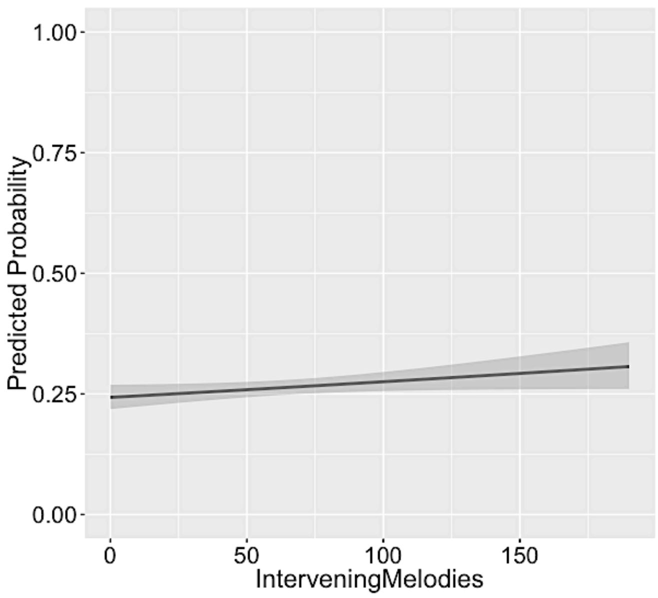


Figure 2.8: Prediction line of a generalized mixed effects model that predicts the probability of bias corrected recognition (y-axis). Performance was significantly above chance. However, the number of intervening items did not carry predictive value. The grey area around the prediction line represents a 95% confidence interval.

As before, mixed effects models addressed the predictive value of the number of intervening items on melody recognition. A base model with a systematic factor for *Dynamic Response Tendency* and random intercepts for *Participant* and *Melody* (*BIC* = 3975.0; *LogLik* = -1969.8) did not significantly improve when provided with information about the *Number of Intervening Melodies* (*BIC* = 3979.8; *LogLik* = -1969.8, p >.07). This result again shows that the number of intervening items cannot be used to predict recognition performance in memory for melody when up to 195 intervening melodies are presented. This is illustrated in *Figure 2.8*.

A final model assessment using maximal random effect structure confirms that the number of intervening items does not carry systematic predictive value in the context of memory for melody (p > .60). Similar to the previous experiments, without a random intercept for *Melody* ($p < .0001$, ΔBIC = 14.8) the model performed significantly worse, showing significant differences between melodies. For a comparison, *Figure C.0.5* in Appendix C displays uncorrected hit-rates for each of the different intervening item conditions in Experiment 1.

2.6.3 Discussion

Descriptively, performance in Experiment 3 was worse than in Experiments 1 and 2, reflecting the impact of transposition. This observation is consistent with recent findings that surface features in music, such as absolute frequencies, play a significant role in memory for melodies even with relatively long delays (Schellenberg & Habashi, 2015). Interestingly, melodies were still resistant to interference by intervening items, and that resistance was demonstrated here with up to 195 intervening items. Thus, relative pitch information alone sufficed to support the resilience. This is consistent with previous research showing that relative pitch information is important for long-term memory for melodies (Dowling & Bartlett, 1981). The results so far might be specific to tasks that focus on the direct *recognition* of sequences. Experiment 4 addressed the possibility that cumulative disruptive interference arises when participants are not directly aware that their memory is being tested. This hypothesis is tested here by measuring changes in perceived *familiarity* between first and second presentations of novel melodies.

2.7 Experiment 4: Perceived Familiarity

Experiment 4 was designed to investigate the effect of the number of intervening items in a different assessment of memory than melody recognition. Rather than measuring recognition performance, participants were instructed to rate perceived familiarity for every melody without being told that melodies reoccur. If the number of intervening melodies has a significant influence on perceived familiarity, then it may be that the results observed in Experiments 1, 2, and 3 are underpinned by a mechanism specific to direct memory task instructions. In a direct memory task such as the recognition paradigm implemented in Experiments 1, 2, and 3, participants are aware that their memory is being tested. Prior research has shown that the degree to which participants are aware of the nature of a memory tasks can influence fundamental memory phenomena (Fleischman et al., 2004; Gaudreau & Peretz, 1999; Halpern & Bartlett, 2010; Halpern & O'Connor, 2000). Specifically, some researchers posit that conscious recollection and a general feeling of familiarity are underpinned by different neurological processes (Yonelinas, 2002). If there is a pattern of results supporting no significant effect of the number of intervening melodies on perceived familiarity, then the findings reported in Experiment 1, 2, and 3 are not specific to a direct melody recognition paradigm.

Although in Experiment 4 participants were not informed that melodies may be repeated throughout the experiment, it was hypothesized that perceived familiarity should increase from the first to the second presentation of each melody. Furthermore, it was hypothesized that the number of intervening melodies does not affect this change in perceived familiarity between the first and second presentation of a melody.

2.7.1 Method

Participants. Thirty undergraduate students from the Western Sydney University volunteered to participate (M_{age} = 23.6, SD_{age} = 6.234, eight male). Participants did not have formal musical training nor did they report any hearing disabilities or participate in the previous experiments. Participation was reimbursed with university course credit.

Stimuli and Equipment. Stimuli and equipment were identical to Experiment 3 and musical scores relating to representative examples of melodies presented in Experiment 4 are presented in Appendix A (p. 249). Identical to Experiment 3, the melodies in Experiment 4 were also transposed upon repetition. Stimuli from all experiments can be found in the online supplement S1-Stimuli.zip.

Procedure. The procedure was to identical to Experiment 3 with the following exception: Participants were asked to indicate "*how familiar you perceive each melody to be*" rather than being asked to make recognition judgments. A vertical 100-point visual analogue scale was used. The familiarity scale had a spatial extent of 10 cm on the computer display and was labeled *unfamiliar* at the bottom and *familiar* at the top. Note that participants were not informed about the reoccurring nature of the melodies.

2.7.2 Results

In summary, Experiment 4 replicated and extended the results of the previous three experiments in a task that measured perceived familiarity, rather than a binary recognition response. *Figure 2.9* shows mean perceived familiarity towards first and second presentation for each participant and each melody. *Figure 2.10* shows perceived familiarity with increasing number of intervening items. Recall that participants were not instructed that melodies reoccur. After testing, participants were invited to speculate about the purpose of this experiment. No participant suspected that their memory was being tested. As hypothesized, perceived *familiarity* increased from the first (M = 38.68, SE = 2.89) to the second occurrence (M = 47.12, SE = 3.214) of the melodies (t = 11.53, p < .0001).

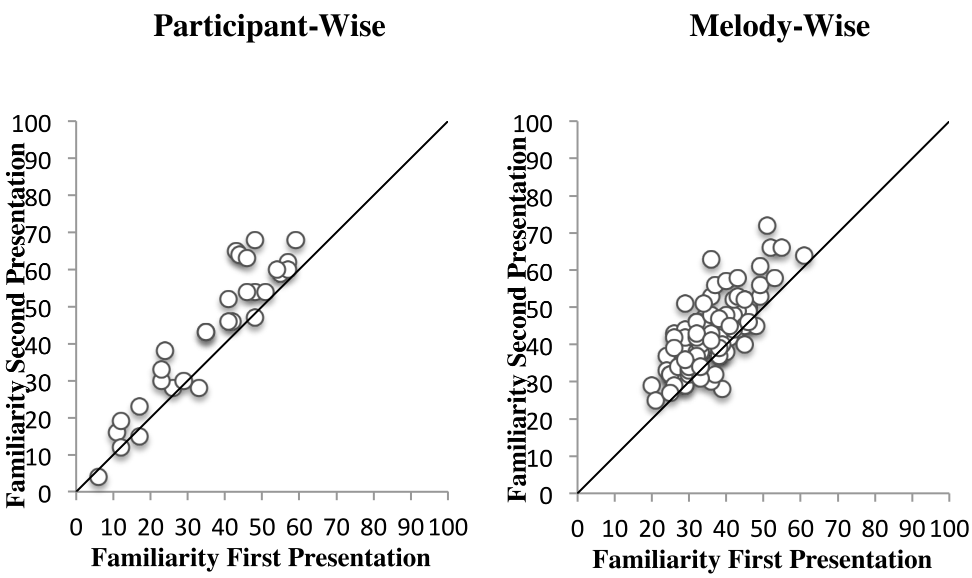


Figure 2.9: Perceived familiarity raw data for Experiment 4. The left panel shows participant-wise differences in perceived familiarity between the first and second occurrence of a melody. The right panel shows melody-wise differences. Second presentations of melodies elicit significantly higher familiarity ratings than first presentations.

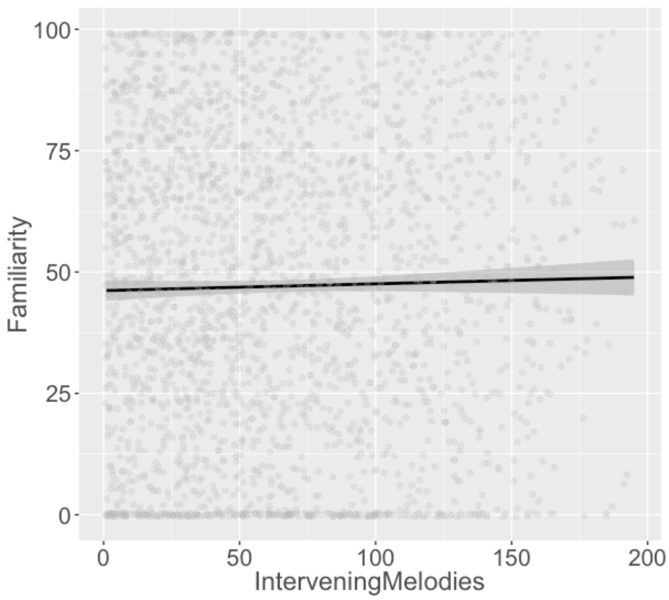


Figure 2.10: Raw data and prediction line of a linear mixed effects model predicting perceived familiarity on the second occurrence of a melody in Experiment 4, based on the number of intervening items. Familiarity increased significantly between first and second presentation of a melody. However, the number of intervening items did not carry predictive value. The grey area around the prediction line represents a 95% confidence interval.

A model predicting within-participant standardized *z*-scores of *Familiarity* with a random intercept of *Subject* and *Melody* (*BIC* = 13660.51; *LogLik* = -6813.26) performed significantly worse ($p < .001$, ΔBIC = 129.24) than the same model provided with the additional information of melody *Occurrence*; that is, the first or second occurrence of a melody (*BIC* = 13531.27; *LogLik* = -6744.40). Furthermore, a model using the maximal random effect structure also confirmed the significant effect of *Occurrence* ($p < .001$). These data further support the hypothesis that overall familiarity changes significantly between the first and second occurrence of a melody, even when taking random subject and item intercepts into account. Similar to all previous experiments, the above model performed significantly worse without a random intercept for *Melody* (*BIC* = 13788.31; *LogLik* = -6877.16, $p < .001$, ΔBIC = 257.04). This result provides evidence that there were significant differences between melodies.

The second hypothesis predicted a pattern of results that support no cumulative disruptive interference of the number of intervening melodies on the change in perceived familiarity between the first and second occurrence of a melody. This hypothesis was also supported. A mixed effects model with a random intercept for *Melody* and *Subject* and systematic factor of *Dynamic Response Tendency* predicting the change of familiarity in participant-wise *z*-scores between the first and second occurrence of a melody (*BIC* = 7797.5; *LogLik* = -3879.3) did not significantly improve ($p > .66$) when provided with information regarding the *Number of Intervening Melodies* (*BIC* = 7805.1; *LogLik* = -3879.2). This is illustrated in *Figure 2.10*. The maximal random effect structure also showed that the *Number of Intervening Melodies* was not a useful predictor of perceived familiarity ($p > .73$).

2.7.3 Discussion

The results of Experiment 4 show that the lack of interference by intervening melodies (Experiments 1-3) is not limited to direct memory task instructions, but also occurs in an indirect memory task that measures perceived familiarity. This was an important test, since previous work has suggested that conscious recollection and the feeling of familiarity are two distinct mechanisms in recognition (Yonelinas, 2002). However, the present findings suggest that the lack of a disruptive effect of the number of intervening melodies is not underpinned by a mechanism specific to direct memory task instructions or conscious recollection, but is also observed in indirect assessments of memory such as perceived familiarity.

2.8 General Discussion

The present study undertook a multi-experiment investigation of how the number of intervening melodies affects memory for melody. Over four experiments involving transposed and untransposed melodies, melody recognition and perceived familiarity assessments, and up to 195 intervening melodies in two different corpora of music, there was no indication of a disruptive effect from the number of intervening melodies on memory for melody beyond immediate repetition.

The present study provides support for the suggestion of Dowling, et al. (1995) that there may be underlying automatic processes involved in the recognition of novel melodies that are not disrupted by the presentation of intervening melodies. In the present investigation, the number of intervening items and the passing of time were closely intertwined, as additional time was required to include additional intervening items. This means that the results here also support Schellenberg and Habashi's findings (Schellenberg & Habashi, 2015) that, in contrast to other domains (Karnekull, Jonsson, Willander, Sikstrom, & Larsson, 2015), the mere passing of time does not interfere with melody recognition. The results also converge with the majority of literature showing that successful recognition of melodies is possible, even when they have only been heard a few times; in our case, only once (Bartlett, Halpern, & Dowling, 1995; Dowling, 1991; Dowling et al., 1995; Halpern & Bartlett, 2010, 2011; Halpern & Müllensiefen, 2008; McAuley et al., 2004;

Müllensiefen & Halpern, 2014; Peretz & Gaudreau, 1998; Schellenberg & Habashi, 2015; Schellenberg et al., 2014).

However, our findings are relatively surprising when one considers that a disruptive effect from the number of intervening items is reported in the context of many domains other than music (Buchsbaum et al., 2011; Bui et al., 2014; Campeanu et al., 2014; Hockley, 1992; Konkle et al., 2010; Nickerson, 1965; Olson, 1969; Sadeh et al., 2014). Note that we do not claim that memory for melody is exceptionally good compared to other stimuli. The present data allow no direct cross-domain performance comparisons. However, the lack of a disruptive effect from the number of intervening items observed here has been reported in only a few cases in other domains (Berman et al., 1991; Tillmann & Dowling, 2007).

In the following, our results will be discussed in terms of melodic transposition, melody distinctiveness, melody recognition and perceived familiarity, music's temporal organization and its relation to memory's domain specificity, as well as a novel Regenerative Multiple Representations conjecture that offers a pathway for future research designed to investigate the psychological mechanisms that may explain our findings.

2.8.1 Melodic Transposition

The findings of Experiments 1 and 2 showed no effect of the number of intervening melodies when pitch information was available to participants in the form of each note's original pitch (absolute frequency information) and the relative pitch intervals between notes (relative frequency information). Experiments 3 and 4 extended these findings, showing there is also no effect of the number of intervening melodies when only relative frequency information is available. Absolute frequency information appears to serve as additional information that is used to aid melody recognition. As indicated by our data as well as previous research, this is demonstrated in better recognition performance towards untransposed compared to transposed melodies (Bartlett & Dowling, 1980; Dowling & Fujitani, 1971; Krumhansl, 2000; Levitin, 1994; Plantinga & Trainor, 2005; Schellenberg et al., 2014). This raises the question: do changes in other important physical properties of music lead to similar results? For example, a change of musical timbre between first and second presentations of a melody would be a useful manipulation to test this hypothesis.

Here, we tested two different melody corpora representing different musical ‘genres’ (modern advertisement jingles vs. European Folk melodies). However, the melodies within each corpus were distinctly different to each other. An important future direction that is beyond the scope of the present investigation is the question regarding how similarity between intervening melodies and the target melodies affect recognition performance. Within the two melody corpora tested here, we did not observe cumulative disruptive interference from the number of intervening melodies. However, the present findings do not necessarily generalize to cases where intervening melodies are significantly more (or even identical), or less, similar to the target melodies. In single-note recognition, for example, the degree of similarity between the target pitch and intervening pitches greatly mediates interference, with greater dissimilarity leading to greater interference (D. Deutsch, 1972). Interestingly, substantially different intervening items such as spoken numbers do not cause cumulative disruptive interference (D. Deutsch, 1970). Future studies could further investigate cumulative distractors and how their similarity to a target melody influences recognition performance. A major difficulty that such studies will need to overcome is establishing the perceptual similarity of melodies such as those used in this work, as well as their cumulative effects.

2.8.2 Melody Distinctiveness

While the number of intervening melodies did not show any significant effect on melody recognition, the melodies within our two corpora did. Some melodies showed high recognition performance even after large numbers of intervening melodies, while others failed to be recognised after only one intervening melody. This suggests that some melodies were not successfully encoded in the first place. Our data suggest that once a melody is successfully encoded, the number of intervening melodies does not influence the retrieval process.

It is reasonable to expect that the specific combinations of underlying musical features in the melodies provide predictive value when it comes to melody recognition. Initial evidence for this hypothesis is found in a recent investigation using a range of musical features in melodies to predict melody recognition performance (Müllensiefen & Halpern, 2014). This study showed that less common motifs relative to a corpus could predict correct recognitions. This is analogous to the visual domain, where better long-term recognition for vivid pictures or oddities is reported (Konkle et al., 2010; Standing, 1973). The importance of musical features for prediction of successful melody recognition is a promising avenue for future investigation. The present study provides behavioural data that will aid in the development of mathematical models that use musical features as recognition predictors. The data further facilitate the endeavour of building predictive models of melody recognition by demonstrating that the number of intervening items, a predictor that carries substantial predictive power in other memory domains, does not seem to apply to melody recognition.

2.8.3 Melody Recognition and Perceived Familiarity

Unlike the situation in Experiments 1-3, participants in Experiment 4 were only instructed to report their feeling of perceived familiarity. Nevertheless, none of the experiments revealed an influence of the number of intervening items on melody recognition or familiarity. In the familiarity task, an increase in perceived familiarity was observed between the first and second presentation of each melody. This suggests that participants formed memory representations of the melodies during the first presentations that increased perceived familiarity when a melody was heard again. Interestingly, the number of intervening melodies did not influence the increase in familiarity. This shows that the lack of a disruptive effect reported here with up to 195 intervening melodies is not limited to melody recognition instructions, but can be extended to indirect measurements of memory such as perceived familiarity. Further studies can investigate whether this finding is replicated with even less direct measurements of memory, such as preference ratings in the form of mere exposure effects, or reaction time.

2.8.4 Temporal Organization and Domain Specificity

One inherent feature of music is that it continuously unfolds through time and comprises successively organized (rhythmic) events (Jackendoff & Lerdahl, 2006).

A melody develops from its first note to its last, and the stimulus as a whole is complete only when the last note has been sounded. It has long been suspected that this temporal organization of music lies at the heart of some important psychological phenomena related to music. For example, in a study by Dowling and colleagues (Dowling, Tillmann, & Ayers, 2001), participants listened to short phrases from classical minuets. After 4-5, 15, or 30 seconds, participants were required to discriminate between the initial phrase and similar lures. Discrimination performance increased with greater temporal delays, but only if the delay was filled with a continuation of the music. No such improvement was observed when the delay was filled with silence or a purely rhythmic continuation. The findings were attributed to an ongoing process of feature binding that assists in forming coherent representation of the melodies. Dowling and Tillmann (2014) suggest that this process applies as long as 'similarity', 'continuity', and 'coherence' of the stimulus is not interrupted. The authors conclude that "the important thing is not that the delay be filled, but that it be filled with musically meaningful material that engages the listener" (Dowling et al., 2001, p. 270). This condition is somewhat analogous to the continuous memory paradigm used in the present study. Specifically, any detrimental effect to memory from the number of intervening items or the passing of time may have been compensated by a domain-specific increase of performance when the 'delay' (in our case, the period of time comprising melodies) is filled with meaningful melodies. A notable difference between the Dowling, et al. (2001) study that used classical minuets and the present study is that here, delays were filled with different melodies, rather than continuations of the target stimulus. Furthermore, the continuation of the listening experience was interrupted by participants' responses after each melody.

Nevertheless, we did not observe a disruptive effect from the number of intervening items. This may indicate that findings reported in (Dowling et al., 2001) in the context of intervening delays filled with related musical content only extends to cases where the delay is filled with unrelated musical material of a similar style. Indeed, such findings have been replicated using non-classical guitar music (Dowling, Magner, & Tillmann, 2016). Similar findings have also been reported in poetry but not in prose (Tillmann & Dowling, 2007): while memory performance declines over time for prose, the mere passing of time has no effect on memory performance for poetry.

Interestingly, an effect of the number of intervening items on recognition is again absent with simple drawings, but it is observed with photographs (Berman et al., 1991; Friedman, 1990a; Konkle et al., 2010). The reasons for this discrepancy are currently unclear but we provide one possible explanation below. The present results show again that findings related to memory in one domain may not necessarily generalize to others (Fougnie et al., 2015). However, when observing several different domains that act similarly, but differently to others, one wonders whether these domains share underlying processes. In this case, a fair overarching question is as follows: what are the similarities of music, poetry, and drawings that lead to a lack of an effect from the number of intervening items and the mere passing of time? In the following, we propose a novel *Regenerative Multiple Representations* (RMR) explanation. This perspective provides clear predictions, is falsifiable, and offers future possibilities for mathematical implementation. However, the conjecture is still speculative and should only serve as pointer for potential future research.

2.9 A Novel Perspective: Regenerative Multiple Representations

The temporal structure of music is realized as a relational structure of underlying temporally spaced components. Examples of such underlying components are notes, pitch intervals, rhythms, or short note phrases within a melody. We define such relational organization as the strong interdependent connection between each element that defines and gives meaning to an object; the 'process of perceptual synthesis' that integrates the fragmented features of an auditory stimulus (D. Deutsch, 1986). In the following, we intend to emphasize the importance of the underlying relational organization, rather than the temporal organization that emerges out of the relational organization.

Relational organization is relevant for perception of most objects in our day-to-day experience. For example, an everyday object such as a chair consists of underlying geometrical shapes. While the relational organization of such an object is often clear, the relevance of different layers in this organization might depend not only on the object, but also on the observer. We have learned that a chair is important to be perceived as an integrated whole. The exact underlying components are often of little relevance. A Joiner experienced in constructing chairs, however, may have a different perception of the same chair that is likely to include additional underlying components. Thus, our prior experience can inform perceptual relevance. In formulating the RMR conjecture, we assume that perception directly influences memory; objects that are perceived as a whole will predominantly be remembered as an integrated whole, objects that are perceived as underlying components may only be remembered as those components.

Usually, we tend to perceive objects as either a whole or as their components, but not both at the same time (Goldstein, 2013, pp. 100-114). This observation has been exhaustively studied in the perceptual grouping and object recognition literature (Wolfe et al., 2012). In general, the perception of objects as a whole is favored because it is often more meaningful to the perceiver. In music, it has also long been theorized that melodies are integrated in memory as a whole (see D. Deutsch, 1986; Krumhansl, 1991, p. 295). Following the assumption that memory is guided by perception (Malmberg & Annis, 2012), the perception of an object as a whole might then lead to a representation in memory of the object as a whole. Such a representation could then be subject to interference and decay, until the representation drops below some form of recognition threshold, rendering retrieval impossible. This process is often implemented mathematically in memory models in the form of a decay parameter, an interference parameter, or both (Norman, 2013; Oberauer & Lewandowsky, 2011; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012).

However, there could plausibly be a group of objects that are simultaneously perceived as both an integrated relation *and* as two or more sets of components that create that relation. The reason this group of objects may be perceived this way is that the perceiver has learned over time that it is important to pay attention to several aspects of the object, as well as their relations. As a result, the object is perceived simultaneously on both levels and thus, multiple representations are formed. We speculate that such multiple representations may be what music, poetry, and drawings have in common. Outside of the memory domain, the hypothesis that music might be represented as a complex whole as well as its underlying components has been proposed as early as 1873 (as discussed in Schneider, 1997, p. 119). In the context of the present study, an integrated melody is an example of the object as a 'whole', whereas particular features such as notes, intervals, or short note clusters within the melody are an example of its underlying components. Stimuli that elicit multiple representations may have an advantage: if one representation fades below the threshold of retrieval, it may still be reconstructed or regenerated by cross-referencing with other representations. We suggest that *regeneration* is triggered when a single representation is accessed that is below the recognition threshold.

For example, a person may find it difficult to immediately answer whether the first and last tone of *Mary Had A Little Lamb* are the same pitch (a question that tests memory for underlying components). However, the same person may still be capable of humming the complete melody of *Mary Had a Little Lamb* (an example of retrieving the integrated whole), which results in the access to information necessary to correctly answer the initial question. In this example, if the person is prompted to answer quickly, performance may be low. However, if the person is provided with additional time to make full use of their integrated representation of the melody as a whole, then they should easily *regenerate* the specific information regarding the underlying components and perform the task with relative ease. This conjecture predicts that once encoded, such regenerative multiple representations are robust against the interference of time (decay) and intervening items, since most of the multiple representations would be theoretically required to drop below the hypothetical retrieval threshold before retrieval is impossible. The representations might be long lasting, though not necessarily permanent. Such an effect would also be independent of overall memory performance from one group of stimuli compared to another, since it only describes what might happen to objects once they are encoded. Indeed, some stimuli are likely to be harder to encode than others.

Taken together in the context of melody recognition, this conjecture speculates that: (1) melodies consist of underlying perceivable components such as notes, intervals, or short note clusters; (2) perceivers have learned that it is important to perceive the underlying components of a melody and subsequently form a memory representation of them; (3) perceivers have learned how underlying components of a melody are related; (4) perceivers are capable of integrating the relational dependency into a whole (e.g., a melody). Consequently, multiple representations specific to the underlying components of the melody and the melody as an integrated whole, are developed; (5) representations in general are subject to forgetting. However, if one representation fades, another representation that remains intact can regenerate it. 'Regenerate' here refers to information that can be retrieved by using information in other memory representations, even though the original memory representation is lost. The greater number of intact representations, the more resilient a memory is likely to be. If and how many multiple representations are formed depends on both stimulus properties and observer specifics (e.g., prior knowledge and experience). In music, most listeners have likely learned how to integrate melodies over years of exposure, but also learned that the individual underlying entities of music such as notes and pitch relations are crucial for the formation and understanding of such melodies. These melodic expectancies are a candidate-mechanism for the aforementioned regeneration process, as they help predict what comes next in a melody (Margulis, 2005; Pearce, 2014; Schellenberg, 1996) and could potentially be used to interpolate forgotten parts of a melody. As a result, the

RMR conjecture predicts that the resilience of memory for melody should continuously build up as a listener becomes more familiar with the tuning system in which it is heard. Training listeners on a new tuning system and continuously testing for cumulative disruptive interference could investigate this prediction. Other implications of the RMR conjecture are discussed in the following.

2.9.1 Implications of Regenerative Multiple Representations

The Regenerative Multiple Representations conjecture may also explain why an effect of the number of intervening items is observed in *photographs* of everyday objects (Konkle et al., 2010), but not in *drawings* of everyday objects (Berman et al., 1991). Drawings may guide attention to the underlying components of a represented object, which leads to multiple representations of the object as a whole, as well as its underlying components. However, photos of everyday objects would predominantly be perceived as the object itself. An additional component unique to the drawing, potentially contributing to regenerative events might be the brushstroke style. This leads to the testable hypothesis that memory for words written in longhand shows no effect of the number of intervening items, while memory for plain printed words does. In addition to the representation of a word as its integrated meaning, longhand might draw attention to individual strokes (underlying components) and their spatial relation to each other, forming another representation of the underlying components, whereas plain printed words would for most observers only result in a representation of the integrated meaning. Consistent with this, participants can perceive and use information related to the mode of production of words in longhand (Babcock &

Freyd, 1988; Knoblich, Seigerschmidt, Flach, & Prinz, 2002; Tse & Cavanagh, 2000).

At this stage, the RMR conjecture cannot predict *exactly* how many memory representations plain printed words have. However, the RMR conjecture can predict that words written in longhand have the same representations that plain printed words have, plus additional representations of the underlying strokes and should therefore be more resilient to intervening items effects. Interestingly, the RMR conjecture also predicts resilience towards cumulative disruptive interference in cases that have previously shown such interference, given that the observer forms more memory representations. An expert in photography composition, for example, is likely to perceive more components of photographs than average naïve observers. These additional percepts form additional memory representations, which should provide the expert with additional resilience towards cumulative disruptive interference.

The RMR conjecture also predicts the finding of an effect of the number of intervening items in words and prose, but not in poetry (Bui et al., 2014; Tillmann & Dowling, 2007). Usually, a word is predominantly represented as a whole and seldom as its actual underlying components. For example, correct letters in a word that are placed at an incorrect position are often perceptually processed as if they were in the correct position (Rayner, Pollatsek, Ashby, & Clifton Jr, 2012a). The same applies to whole sentences. Usually the meaning is represented while individual words are sometimes 'skipped' during reading (Rayner, Pollatsek, Ashby, & Clifton Jr, 2012b). In poetry, on the other hand, attention may be shifted towards the underlying components (e.g., individual words), as they carry greater importance and may be sequenced irregularly. As a result, more representations of underlying components, such as underlying words and how they interact with each other, are likely to manifest in poetry while still maintaining an overall representation for every line, stanza, and the entire poem as an integrated whole. In summary, the degree of saliency attributed to different properties of a stimulus is informed by prior experience.

Providing a task that forces participants to focus on underlying elements, the whole, or both simultaneously would be a suitable context to test the Regenerative Multiple Representations conjecture. Furthermore, one would expect to find similar results in domains that also favour multiple representations. One example of such a domain is dance. Dance consists of dynamic movement as well as underlying postures, and both have been shown to contribute to recognition of contemporary dance postures (Vicary, Robbins, Calvo-Merino, & Stevens, 2014).

2.9.2 Regenerative Multiple Representations in Relation to other Memory Models

Regenerative Multiple Representations is similar to previous memory theories in so far as it draws from multiple-trace theories (Hintzman, 1984, 1988). It also does not challenge global matching models that describe recognition as the *'match'* response of memory that reflects a familiarity distribution to a cue (Clark & Gronlund, 1996). However, even in-depth memory models such as REM, SLiM, and BCDMEM predict interference effects from lags or delays that were not observed in the present study's findings (Dennis & Humphreys, 2001; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997).

The Regenerative Multiple Representations conjecture also bears resemblance to Paivio's dual-coding theory (Paivio, 1969). It assumes two representations (words and images) that assist in retrieval of each other, therefore increasing the chance of remembering a stimulus. Regenerative Multiple Representations postulate any number of representations, and not necessarily words and images. This is important for a theory applied to music or other stimuli where multiple representations beyond two are likely. Indeed, Paivio's dual-coding theory could be described as a special case of the present Regenerative Multiple Representations conjecture, where prior experience informs our perception to focus on words and their associated images.

The notion of perceptual relevance that is informed by prior knowledge is somewhat similar to the notion of *pertinence* described in D. Deutsch (1986) (see also J. A. Deutsch and Deutsch, 1963). Pertinence describes a weighting of a perception based on a current situation, as well as long-term factors such as prior knowledge. However, while pertinence primarily influences awareness, we suggest that perceptual relevance directly influences perception and subsequent formation of future memories: a chair that is perceived as an integrated whole will be remembered as an integrated whole rather than the underlying components.

The Regenerative Multiple Representations perspective is speculative, with many open questions. For example, how can the relative strength of representations be measured? Can perceptual relevance be manipulated to form multiple representations and, as a result, facilitate learning and memory encoding? The conjecture does allow for many specific and informative predictions and mathematical implementations that can be tested in future research endeavours.

Ongoing work in our lab has tested and found support for some of the predictions made by the RMR conjecture. For example, one hypothesis concerned whether an effect of the number of intervening items will manifest if note based melodies are played in tuning systems that are completely unfamiliar to participants (e.g., microtonalities) (Herff, Olsen, Dean, & Prince, 2017). In this case, listeners should still be capable of perceiving some of the underlying component, but may fail to integrate the stimuli into perceptually coherent melodies. Based on the RMR perspective, the prediction was that the number of intervening items will have a significant negative impact on memory performance, because listeners do not have the multiple representations necessary to utilize the regeneration process. This prediction was supported in Herff, Olsen, et al. (2017). Other predictions of the RMR conjecture concerning pitch-only and rhythm-only sequences, rather than combined musical melodies, have also been tested and supported in Herff, Olsen, Prince, and Dean (2017).

2.10 Conclusion

Human memory is fallible and prone to interference. In many domains, memory performance decreases as the number of intervening items between the first and second presentation of a stimulus increases. However, this phenomenon is not universal, and music in particular has proven to possess intriguing properties when it comes to memory: these were further investigated here. Regardless of whether one or 195 intervening melodies were presented, performance in all our experiments was above chance and not affected by the number of intervening melodies. This finding was observed using two different musical corpora. Furthermore, transposition of each melody's repetition left only relative frequency information, yet the number of intervening melodies still did not affect memory for melody. In addition to direct recognition, one experiment measured memory for melody in the form of perceived familiarity, with results that were consistent with the other findings of the study.

To explain the findings, we offer a novel yet speculative R*egenerative Multiple Representations* conjecture that bears resemblance to Paivio's dual-coding theory (Paivio, 1969). The conjecture assumes that: previous experience influences our perception; perception determines which memory representations are formed; memory representations are subject to decay and interference; there are stimuli where previous experience informs us to simultaneously perceive these stimuli in multiple ways; multiple perception leads to multiple representations; multiple representations can regenerate each other, making them resilient to decay and interference; melodies in familiar tuning systems belong to the category of objects that we simultaneously perceive in multiple ways, such as underlying components (e.g., notes, intervals or phrases) as well as an integrated, coherent whole (i.e., the melody).

Future studies investigating these assumptions in the context of music (e.g., familiar tonal versus unfamiliar atonal stimuli) as well as temporally dynamic domains such as dance (e.g., dynamic movement versus underlying postures) and words (e.g., longhand versus printed) will provide empirical evidence or counterevidence regarding the *Regenerative Multiple Representations* conjecture, thus facilitating its development into an empirically informed theory of human memory that can begin to answer the question: why does the sheer number of intervening items have no influence on memory for melody, when it does for almost all other memory domains?

Chapter 3

Not-So-Resilient Memory for Melodies

Chapter 3 has been peer reviewed and accepted for publication:

Herff, S.A., Olsen, K.N., Dean, R.T., & Prince, J. (2017). Memory for melodies in unfamiliar tuning systems: Investigating effects of recency and number of intervening items. *Quarterly Journal of Experimental Psychology*.

Note: Kirk N. Olsen and Roger T. Dean are the author's supervisors. Jon Prince shared some recently collected data that is analysed here.

Author note

At the Australian Music Psychology Conference 2015, I attended Jon Prince's presentation titled: *Implicit learning of an artificial grammar structure in pitch and/or time*. To study statistical learning of artificial grammars in auditory stimuli, Dr. Prince required participants to be exposed to auditory stimuli containing unfamiliar structures. After the exposure phase, the participants went on to participate in a series of different experiments, each deepening our understanding of implicit learning of artificial grammar in the auditory domain.

I realised, however, that the initial 'exposure' phase in his experiments consisted of a memory paradigm and comprised data that was highly relevant to my investigation of memory for melody. After his talk, I approached Jon and asked him to share his data. I would like to thank Jon and his co-authors who generously agreed to share their exposure-phase data with me. Here, the raw data of Experiment 7 (Chapter 3) and Experiments 8-10 (Chapter 4) come from an investigation of implicit learning in the auditory domain, originally conducted by Jon Prince, Mari Jones, Kate Stevens, and Barbara Tillman.

3.1 Abstract

In a continuous recognition paradigm, most stimuli elicit superior recognition performance when the item to be recognised is the most recent stimulus (a recency-in-memory effect). Furthermore, increasing the number of intervening items cumulatively disrupts memory in most domains. Memory for melodies composed in familiar tuning systems also shows superior recognition for the most recent melody, but no disruptive effects from the number of intervening melodies. A possible explanation has been offered in a novel Regenerative Multiple Representations (RMR) conjecture. The RMR assumes that prior knowledge informs perception and perception influences memory representations. It postulates that melodies are perceived, thus also represented, simultaneously as integrated entities and also their components (such as pitches, pitch intervals, short phrases, and rhythm). Multiple representations of the melody components and melody as a whole can restore one another, thus providing resilience against disruptive effects from intervening items. The conjecture predicts that melodies in an unfamiliar tuning system are not perceived as integrated melodies and therefore should: a) disrupt recency-in-memory advantages; and b) facilitate disruptive effects from the number of intervening items. We test these two predictions in three experiments. Specifically, Experiments 5 and 6 of this dissertation show that no recency-in-memory effects emerge for melodies in an unfamiliar tuning system. In Experiment 7, disruptive effects occurred as the number of intervening items and unfamiliarity of the stimuli increased. Overall, results are coherent with the predictions of the RMR conjecture. Further investigation of the conjecture's predictions may lead to greater understanding of the fundamental relationships between memory, perception, and behaviour.

3.2 Memory for Melodies in Unfamiliar Tuning Systems: Investigating Effects of Recency and Number of Intervening Items

Recognition performance in most domains decreases as the number of intervening items increases between the first and second occurrence of a target stimulus. This phenomenon has been subject to much experimental investigation in a variety of tasks using a variety of stimuli, such as digits, letter trigrams, word lists and pairs, and faces (Bui et al., 2014; Donaldson & Murdock, 1968; Hockley, 1992; Konkle et al., 2010; Olson, 1969; Poon & Fozard, 1980; Rakover & Cahlon, 2001; Sadeh et al., 2014). Despite the widespread nature of this effect, increasing the number of intervening items does not always lead to decrements in recognition performance. In the visual domain, for example, disruptive effects from the number of intervening items occur for pictures of everyday objects, but not for line drawings of everyday objects (Berman et al., 1991; Friedman, 1990a; Konkle et al., 2010). In the auditory domain, disruptive effects are apparent in spoken words and recognition of single musical notes, but not for recognition of novel melodies in familiar musical tuning systems such as 12-Tone-Equal-Temperatement (12-TET) (Buchsbaum et al., 2011; Campeanu et al., 2014; D. Deutsch, 1970, 1975; Herff, Olsen, & Dean, 2017).

For example, Schellenberg and Habashi (2015) investigated how recognition memory for melodies in a familiar tuning system decays over time, and found delays of up to one week had minimal effects. Complementing Schellenberg and Habashi's results on decay, Herff, Olsen, and Dean (2017) recently investigated interference and reported no disruptive effects from the number of intervening items on melody recognition using up to 197 intervening melodies. Herff et al. (2017) did, however, find a melody recognition advantage for immediate repetition. Recognition advantages for immediate repetition, or recency effects, are commonly observed in memory literature as a memory advantage for the last stimulus presented (Berz, 1995; Dowling, 1973; Greene & Samuel, 1986; Jahnke, 1963; Roberts, 1986). However, melodies provide a special case for recency because they consist of a sequence of notes. In this case, the last stimulus encountered is always the final note, rather than the last melody. To trigger a recency effect, melodies would need to be perceived as perceptually integrated entities, rather than just individual notes. Previous literature suggests that melodies are indeed perceived as coherent integrated wholes (D. Deutsch, 1986; Dowling, 1991; Krumhansl, 1991). Herff et al. (2017) have proposed a novel Regenerative Multiple Representations (RMR) conjecture to explain why memory for melody not only exhibits recency-in-memory effects, but is also not cumulatively disrupted as the number of intervening melodies increases.

3.2.1 Regenerative Multiple Representations in Memory for Melody

The RMR conjecture combines and generalizes well-established memory theories. The conjecture firstly assumes that previous knowledge or experience directly influences our perception, and secondly assumes that perception determines the formation of future memories. In other words, we learn the most relevant way to perceive our environment, perceive objects according to this information, and form memories according to the perception. This is analogous to Deutsch's concept of *Pertinence* (D. Deutsch, 1986; see also J. A. Deutsch and Deutsch, 1965). Pertinence is described as a weighting of perception in the form of awareness based on the current situation and prior knowledge. We apply this term to memory, inferring that previously learned perceptual relevance directly influences perception itself, and therefore influences subsequent formation of new memories based on these perceptions.

Following the assumption that experience directly influences perception, and that perception determines the formation of future memories, the conjecture postulates the formation of multiple memory representations in cases where prior knowledge dictates multiple perceptual relevancies. This is, for example, demonstrated in Stroop effects where the written name of a colour and the actual colour of a word mutually interfere (Stroop, 1992). This phenomenon would not occur if the word was written in an unfamiliar language. In this case, prior knowledge cannot inform perceptual integration of the letters in the words into a meaningful word.

The third assumption of the RMR conjecture is a common one: memory representations are subject to decay and interference. Memory models often implement this process mathematically in the form of a decay parameter, an interference parameter, or both (Norman, 2013; Oberauer & Lewandowsky, 2011; Oberauer et al., 2012).

The fourth assumption of the RMR conjecture is analogous to – and a generalized form of – Paivio's dual-coding theory (Paivio, 1969). In the context of words and images, the dual coding theory describes exactly two representations that assist each other in retrieval, therefore increasing the chance of remembering a stimulus. The RMR conjecture generalizes this case and postulates any possible number of representations that are not necessarily just words and images. Prior knowledge determines the number and nature of representations. This is important for a theory applied to music, where multiple representations beyond two are highly likely due to music's perceptual reliance on multiple levels such as its underlying parts (e.g., notes, intervals, phrases), integrations of the underlying parts (e.g., coherent melody), socio-cognitive components (e.g., emotions), and motor responses evoked by its temporal organization (Barascud, Pearce, Griffiths, Friston, & Chait, 2016; D. Deutsch, 1986; Krumhansl, 1991, p. 295; Schneider, 1997, p. 119; Zatorre, Chen, & Penhune, 2007). Analogous to Paivio's dual coding theory, RMR assumes that multiple representations can regenerate each other. This would mean that once encoded, such regenerative multiple representations are robust to the interference of time (decay) and intervening items, since multiple representations would be theoretically required to drop below a hypothetical retrieval threshold before retrieval is impossible. For example, to truly forget a melody beyond recognition, the RMR conjecture assumes that not only the representation of the integrated melody, but also the representations of all melody-specific underlying components, such as short phrases within the melody, need to be inaccessible. A promising candidate mechanism for memory regeneration in the context of music is melodic expectancy. Due to familiarity with the underlying rules of a music tradition, listeners form strong melodic expectancies. These expectancies can be used to predict what comes next in a melody (Margulis, 2005; Pearce, 2014; Schellenberg, 1996) and theoretically, could also be used to interpolate forgotten parts of a melody.

3.2.2 Regenerative Multiple Representations: Implications for Unfamiliar Tuning Systems

Integrating notes, intervals, and short phrases into a coherent melody requires familiarity with the underlying rules of how individual components inter-relate (Cui, Collett, Troje, & Cuddy, 2015; Saffran, Johnson, Aslin, & Newport, 1999; Schon & Francois, 2011; Tillmann & McAdams, 2004). The regeneration process described above reduces the disruptive effect of the number of intervening items on melody recognition and requires multiple representations of the same melody; representations that code – at least partially – overlapping features of the stimulus. For example, the representations of a melody's components (e.g., notes, intervals, short phrases) as well as their integrated whole (the melody itself) can have redundancy. In order to form a representation of the integrated melody, the melodic sequence of notes needs to first be perceived as an integrated melody. This means that disrupting the perception of an integrated melody should simultaneously disrupt the formation of its memory representation. As a result, the regeneration process cannot be fully initiated. In this case, the RMR conjecture predicts stronger disruptive effects from the number of intervening items because listeners should only perceive and form representations of the melody's underlying components, and not of an integrated melody as a whole. Melodies that are sounded in an unfamiliar tuning system are a useful context to test this hypothesis. From classical to modern pop-music, the tuning system most Western listeners are familiar with is 12-TET; a tuning system that describes a division of a musical octave in 12 equal step sizes. However, there are an infinite number of other possible tuning systems.

As stated above, listeners presented with melodies in an unfamiliar tuning system should perceive and form a representation of a melody's underlying components. However, assuming the listener is unfamiliar with that tuning system's rules and relations between the components required to integrate the melodies into a coherent whole, they should not be capable of perceiving large-scale melodic structure (e.g., Castellano, Bharucha, & Krumhansl, 1984), thus not forming a representation of a melody as a whole. This should affect the regeneration process, resulting in a stronger disruptive effect from the number of intervening melodies on recognition of melodies composed in an unfamiliar tuning system, relative to melodies composed in a familiar tuning system. In other words, recognition performance in response to melodies in an unfamiliar tuning system should decline as the number of intervening melodies increases.

Previous research has observed recency-in-memory effects in recognition of melodies in the familiar 12-TET tuning system (Berz, 1995; Dowling, 1973; Herff et al., 2017). If a melody in a familiar tuning system presented at test was also the previously presented melody, then a recognition advantage emerges relative to when the tested melody was not the previously presented stimulus. This recency-in-memory effect suggests that melodies in a familiar tuning system are indeed treated as integrated wholes. This is because the recency-in-memory effect is defined as enhanced recognition for the last stimulus presented; in this case, the melody as a whole. If a melodic sequence of notes is not perceived as an integrated melody, then the 'last stimulus encountered' is no longer the melody as a whole. Rather, it is most likely the last note or at most the last phrase of the melody. As a result, no recognition enhancement for the most recent melody would be expected because the most recent stimulus is then no longer defined at the level of the melody as a whole.

If unfamiliar tuning systems disrupt integration of melodies as a whole (as suggested by the RMR conjecture), then no recency-in-memory effect should occur. This is because recognition memory in this context can only rely on representations of the melody's underlying components, rather than multiple representations that include the melody's underlying components *and* the melody as a coherent whole. Furthermore, cumulative disruptive effects of the number of intervening items should also occur. These hypotheses will be investigated here.

3.3 Aim, Design, and Hypothesis

The aim of this study is to investigate recency-in-memory and cumulative disruptive effects from an increase in the number of intervening melodies in memory for melodies using an unfamiliar tuning system. The study closely follows the design of previous work investigating these effects using stimuli composed in a familiar tuning system (Herff et al., 2017). Here, Experiment 5 tests the early time-course of recognition memory for melodies in an unfamiliar tuning system by using a direct memory task in a continuous recognition paradigm, and varying the number of intervening melodies between immediate repetition (zero) and thirteen. Experiment 6 provides a close replication of Experiment 5, however with an indirect memory task (perceived familiarity). Such a task was chosen because there is evidence that the level of task awareness influences some memory for melody phenomena (Gaudreau & Peretz, 1999; Halpern & Bartlett, 2010; Halpern & O'Connor, 2000). Experiment 7 tests whether there are cumulative effects from the number of intervening melodies on melody recognition using up to 107 intervening melodies and two repetitions of each melody.

3.4 Experiment 5 – Zero up to Thirteen Intervening Melodies in an Unfamiliar Tuning System

3.4.1 Method

Participants. Thirty-seven undergraduate students were recruited from Western Sydney University (M_{age} = 22 years, SD_{age} = 5.5; 7 male, 30 female). Participants were required to have had less than two years of musical training and no hearing impairments. Participants did not participate in any of the previous experiments (Experiments 1-4). Participants were reimbursed with course credit as part of university course requirements.

Stimuli and equipment. Stimuli from all experiments can be found in the online supplement S1-Stimuli.zip. The aim of the stimulus generation process was to generate melodies within an unfamiliar tuning system. To achieve this, a set of tonal melodies that followed the Western tonal tradition were chosen and placed in a specially designed tuning system comprising carefully defined similarities and dissimilarities to the familiar 12-TET. The original 12-TET stimulus set consisted of a mathematically and perceptually tested subset of European folk songs that has been used in a previous related study (Herff et al., 2017). Previously, a pilot study established the response pattern of a group of listeners to unfamiliar melodies and removed any of the original stimuli that showed atypical familiarity response distributions (details in Herff et al., 2017). Here we use the same stimuli but retuned to a new unfamiliar tuning system, as described below.

The construction of the unfamiliar tuning system was based on the following principles: Firstly, to ensure that potential effects from the number of intervening items can be interpreted as arising from the unfamiliarity of the tuning system, rather than contaminated by perceptual difficulties, the smallest possible step in this unfamiliar tuning system was made large enough to ensure that two pitches are easily discriminated. Pitch interval sizes are commonly measured in log-frequency units such as *cents* because pitch, particularly in a musical context, is more linearly related to log-frequency than it is to frequency. 12-TET has step sizes of 100 cents (the term *step* is here used for the interval between adjacent pitches in the tuning system). Previous research suggests that pitches separated by 40 cents or more can be discriminated approximately as well as those separated by the familiar 100 cents (Parncutt & Cohen, 1995).

Secondly, an equally tempered tuning system was sought, but we did not impose any requirement for our novel tuning systems to contain octaves. Typically, pitches an octave apart (double or half the frequency) are heard as similar (if not identical) and often have the same musical function (D. Deutsch & Boulanger, 1984; Krumhansl & Shepard, 1979). One advantage of not privileging octaves in our tuning system is that a greater number of equally tempered tuning systems can be considered in the following step. In summary, we aimed to develop an equally tempered tuning system that minimized perceived similarity to 12-TET by using step sizes between 40 and 100 cents, and without prioritization of octave relationships, because the RMR predicts larger effects the more unfamiliar a tuning system is.

Dissimilarity between tunings systems was assessed using a psychoacoustically informed and perceptually tested model of tonal affinity and similarity developed by Milne and colleagues, which was first described in Milne, Sethares, Laney, and Sharp (2011) and elaborated in Milne (2013). These papers present a related family of models of tonal similarity/affinity, one of which (*relative dyad expectation vectors*) can assess the similarity of the interval content of one scale or tuning system with the interval content of another, whilst taking into account uncertainties of pitch perception.

The model has one free parameter *sigma*, which is the standard deviation of a Gaussian smoothing kernel applied to the log-frequencies of each scale tone. Sigma models the inaccuracy of pitch perception, and a value of six cents was chosen because sigma is optimized to approximately six cents when fitted to data from three different experiments (Milne & Holland, 2016; Milne, Laney, & Sharp, 2015, 2016). When both tuning systems are smoothed in this way, their similarity is determined by their cosine similarity, which is 0 when the two systems are orthogonal (maximally dissimilar) and 1 when they are identical.

In order to find the equally tempered tuning that is maximally dissimilar from 12-TET, a MATLAB routine generated equal temperament tuning systems with steps between 40 to 100 cents, using an increment of .01 cents (so the first equally tempered tuning system has steps all of size 40 cents, the second system has steps all of size 40.01 cents, the third has steps all of size 40.02 cents, and so on until the final system with steps of 100 cents). For each tuning system, the routine calculated the cosine similarity with the 12-TET system. The tuning system with the highest dissimilarity was chosen. Appendix D (p. 265) shows the result of this computational search. This maximally dissimilar equally tempered tuning had steps of size 88.08 cents, and is denoted 88.08-CET for *cents equal temperament*. The new 88.08-CET shows a cosine similarity of .28058 with the familiar 12-TET system, as measured by the tonal affinity model (Milne, 2013; Milne et al., 2011).

The original melodies were played through the physical synthesis PianoTeq 5 using the 88.08-CET system. The scala file of the 88.08-CET system can be found in the online supplement S1-Stimuli.zip/Experiment5-6/8808cET.scl. The step numbers between notes were held constant, mapping the original 12-TET melodies to the new 88.08-CET tuning system. That means that if in the original melodies a note was two step sizes apart from an adjacent note in the sequence, it will still be two step sizes apart from the surrounding notes in the new tuning system, even though the size of the steps has changed. The result was 50 melodies retuned to an unfamiliar tuning system whose rhythms and contours are identical to the Western-tonal melodies that previously exhibited recency-in-memory effects on recognition for melodies, yet no disruptive effects from the number of intervening melodies (Herff et al., 2017). The melodies had a mean duration of 10.80 seconds (*SD* = 2.29) and consisted of 15 to 78 notes. An analysis of the melodies can be found in the online supplement S1-Stimuli.zip/ Experiment5-6/MusicalFeatures.csv, which describes every melody along with 21 musical features.[4] Melodies were presented diotically in stereo through Sennheiser HD 25 headphones using an UA-25 Edirol external USB-soundcard.

[4] Note that the musical features have been calculated on the original versions of the melodies.

Procedure. Testing took place in a sound-attenuated booth provided by the MARCS Institute, Sydney, Australia. The procedure closely followed Herff et al. (2017). Participants provided informed consent and demographic questionnaires were administered. Standardized instructions appeared on a computer screen and participants were informed that they would hear "*many different melodies one after another*" and that they were required to indicate if "*a melody has already been presented in this experiment*". Melodies were presented in random order in a continuous recognition paradigm (Shepard & Teghtsoonian, 1961). In this task, melodies are continuously presented and a response is required after each melody. Participants responded using the mouse to click one of two buttons that appeared on the screen after the presentation of each melody. One button was labeled "*New*" and the other one "*Old*". The "*New*" button was to be clicked if the first presentation of a melody was perceived, and the "*Old*" button was clicked if a melody was perceived as previously presented. Each new trial was initiated as soon as a participant gave a response. Unknown to the participant, the number of melodies intervening between the two presentations of a given melody varied between 0, 1, 2, 3, 4, 7, and 13. The order of melodies and conditions were randomized. Participants completed six practice trials, in which they adjusted the volume to their personal preference. Testing of the 100 trials took approximately 30 minutes.

Statistical approach. The statistical approach was analogous to previous approaches outlined in Herff et al. (2017). Specifically, we used generalized linear mixed effects models to investigate the influence of the number of intervening melodies on binary melody recognition data (Experiments 5 and 7) (Baayen, 2008;

Baayen et al., 2008; Judd et al., 2012; Kass & Raftery, 1995; Kruschke, 2010, 2013; Nathoo & Masson, 2016). Furthermore, we used linear mixed effects models to analyze continuous familiarity data (Experiment 6). The models were implemented in the R software platform (R-Core-Team, 2013) using the lme4 package (Bates et al., 2013) and consisted of the experimental fixed factor *Number of Intervening Melodies*. Random effects on *Participant* and *Melody* intercepts were included in the models (Barr et al., 2013). Coefficient *p*-values are reported as calculated by lme4 for generalized mixed effects models and Kenward-Roger corrected for linear mixed effects models (Halekoh & Højsgaard, 2014a, 2014b; Kenward & Roger, 1997). Significant coefficients were also further assessed in the form of a model comparison approach (Kruschke, 2011). Each model with a significant predictor was also compared with the same model but without the significant predictor using likelihood-ratio tests (Wilks, 1938). To ensure that significance was not due to an increase in model complexity, differences in Bayes information criteria (Schwarz, 1978) are reported in the form of ΔBIC (Kass & Raftery, 1995). A ΔBIC of two or greater is assumed "positive" in favor of the model with lower BIC. A ΔBIC difference of six or greater is considered "strong" evidence (Kass & Raftery, 1995).

The possibility of 'response tendency shifts' throughout the course of recognition experiments is a significant issue that we address in the present set of analyses (Berch, 1976; Donaldson & Murdock, 1968; Snodgrass & Corwin, 1988). These shifts describe changes in the response bias as an experiment progresses; for example, changes in response tendencies due to fatigue. Similar to previous studies (Herff et al., 2017; Herff, Olsen, et al., 2017), we trained participant-wise generalized mixed effects models on 'old' responses on first presentations (False

Alarm rates) based on trial number. The fitted model was then used to predict the probability of pressing 'old' on a repetition trial, based solely on trial number. These predictions were then implemented as a fixed *Dynamic Response Tendency* factor in all models to account for individual participant response tendencies, and for how these tendencies might change over the course of the experiment. Mixed effects models also assess whether overall performance in each experiment was at chance. Consequently, coefficient *Z*-scores as well as *p*-values of the increase in 'old' responses between first and second melody presentation are reported at the beginning of each results section.

3.4.2 Results

Figure 3.1 shows melody- and participant-wise performance. Overall, participants performed significantly above chance ($Z = 15.67$, $p < .0001$). A generalized mixed effects model was constructed to investigate the first hypothesis that melodies in an unfamiliar tuning do not exhibit a recency-in-memory effect. This hypothesis was supported. Coefficient estimation for each number of intervening melodies showed that no number of intervening items produced significantly worse recognition performance than immediate repetition (all *p*-values $> .10$). Furthermore, a model predicting 'old' responses on melody repetitions using a random intercept for *Subject* and *Melody*, as well as a systematic factor for *Dynamic Response Tendency* (*BIC* = 2262.9; *LogLik* = -1116.4), did not significantly improve when provided with the *Number of Intervening Items* (*BIC* = 2267.8; *LogLik* = -1115.1, $p = .103$).

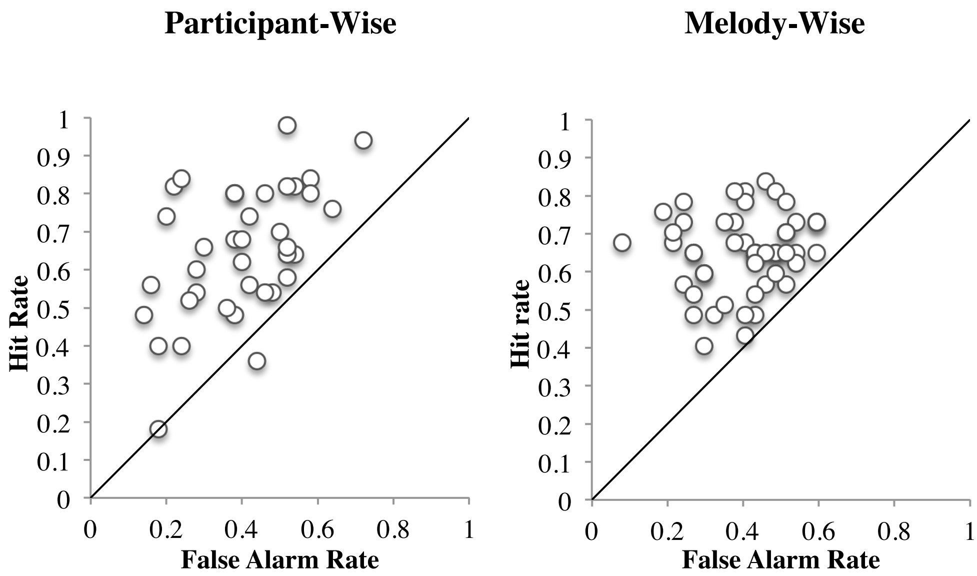


Figure 3.1: Hit rates and false alarm rates in Experiment 5. The left panel shows the data participant-wise, and the right panel melody-wise. The reference line represents chance level.

Within the tested maximum of 13 intervening melodies, the non-significant result from the intervening melodies does not support the second hypothesis that, in contrast to Western-tonal melodies, melodies in an unfamiliar tuning system show disruptive effects of the number of intervening items. *Figure 3.2* depicts recognition performance in response to each intervening item condition.

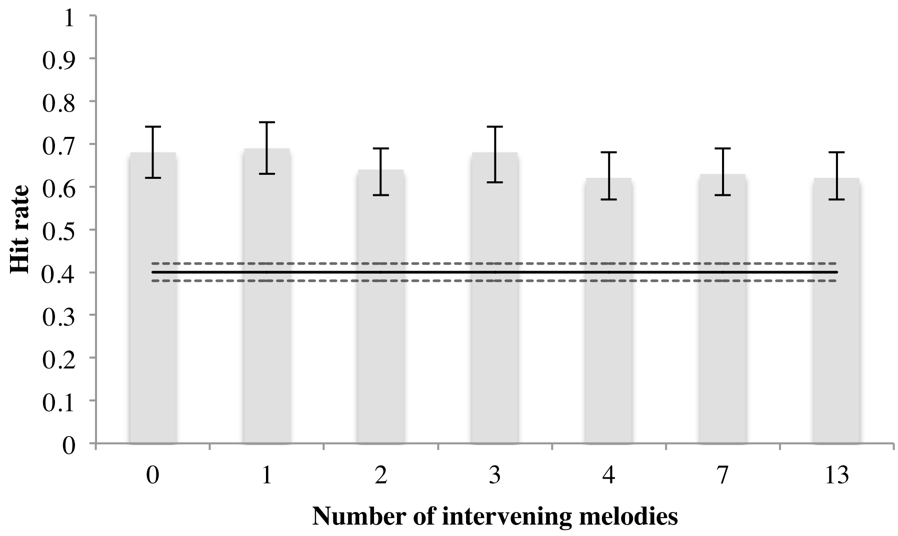


Figure 3.2: Mean hit rates for all seven conditions of intervening melodies in Experiment 5. Error bars show 95% confidence intervals. All conditions were statistically identical. The dark line depicts mean false alarm rates. The two dotted lines represent a 95% confidence interval around the false alarms. The difference between the false alarm rates and the hit rate can be interpretated as bias-corrected performance.

3.4.3 Discussion

Experiment 5 aimed to test two hypotheses generated by the RMR conjecture. Firstly, the conjecture predicts that melodies in an unfamiliar tuning system do not show recency-in-memory effects, unlike melodies in a familiar tuning system. The present results supported this hypothesis. Previously, participants showed recognition advantages for melodies in a familiar tuning system when the test item was the same melody as the previously presented item (Herff et al., 2017); in other words, no intervening melodies between the first and second occurrence. The melodies presented here in an unfamiliar tuning system did not produce such recency-in-memory effects. The second hypothesis predicted that melodies in an unfamiliar tuning system show cumulative disruptive effects from the number of intervening melodies, whereas melodies in familiar tuning systems do not (cf. Herff, et al., 2017). This hypothesis was not supported here using 13 intervening melodies. Specifically, the number of intervening melodies did not have a disruptive effect when retuned to the unfamiliar 88.08-CET system. This result is somewhat surprising, but the effect may be due to the small number of intervening melodies (up to 13).

Other auditory stimuli such as words show disruptive effects within this range of intervening items (Buchsbaum et al., 2011). However, the melodic sequence of notes used here in each melody were still similar to 12-TET in regards to pitch contour, rhythm, and the usage of an equally tempered tuning system. This similarity could account for some of the resilience in memory that has been previously observed using Western tonal melodies. Nevertheless, observing disruptive effects when using a greater number of intervening melodies and less familiar melodic material than Experiment 5 would provide a more comprehensive test of the RMR conjecture. We will address these issues in Experiment 7.

Alternatively, the findings here could be specific to the direct recognition task. Previous literature suggests that the level of task awareness might influence memory for melody (Gaudreau & Peretz, 1999; Halpern & Bartlett, 2010; Halpern & O'Connor, 2000). Participants in Experiment 5 were fully aware that their memory was being tested. It is possible that the here-observed disruption of recency-in-memory effects is specific for direct recognition and does not generalize to indirect measurements of memory. Therefore, Experiment 6 investigates the influence of intervening melodies on perceived familiarity without informing participants about the recurrent nature of the stimuli.

3.5 Experiment 6 – Perceived Familiarity using Zero to Thirteen Intervening Melodies

Rather than direct recognition performance, Experiment 6 utilised an indirect memory paradigm: perceived familiarity. Although not directly instructed to perform a memory task, participants' perceived familiarity ratings tended to be higher on the second presentation of a melody compared to the first (Herff et al., 2017). An increase in perceived familiarity can be used as a proxy of the strength of memory. Based on the results of Experiment 5 and the RMR conjecture, no recency-in-memory effect should emerge from ratings of perceived familiarity. That is, a significantly larger increase in familiarity after one intervening item, compared to all other intervening items, is not explained by a linear decrease with an overall increase in the number of intervening melodies. This hypothesis is investigated here.

3.5.1 Method

Participants. Twenty-seven undergraduate students were recruited from the Western Sydney University (M_{age} = 24 years, SD_{age} = 10.8, 5 male / 22 female) and did not participate in Experiments 1-5. Participants were required to have had less than two years of musical training and no hearing impairments. Participants were reimbursed with course credit as part of university course requirements.

Stimulus. Stimuli were identical to Experiment 5.

Procedure. The procedure closely followed that of Experiment 5. However, participants in Experiment 6 were not instructed that they were engaging in a memory task and not informed that melodies were repeated throughout the experiment. Instead, participants were prompted to indicate "*how familiar you perceive each melody to be*". Responses were made using the mouse and a vertical 100-point visual analogue scale with a spatial extent of 10 cm. The scale was labelled "familiar" at the top, and "unfamiliar" at the bottom.

3.5.2 Results

Figure 3.3 shows melody- and participant-wise perceived familiarity on first and second occurrences of melodies. Coefficient estimation revealed that participants' perceived familiarity ratings were significantly higher (t = 10.69, p < .0001) on second presentations (M = 54.20, SE = 5.85) than on first (M = 44.16, SE = 5.59).

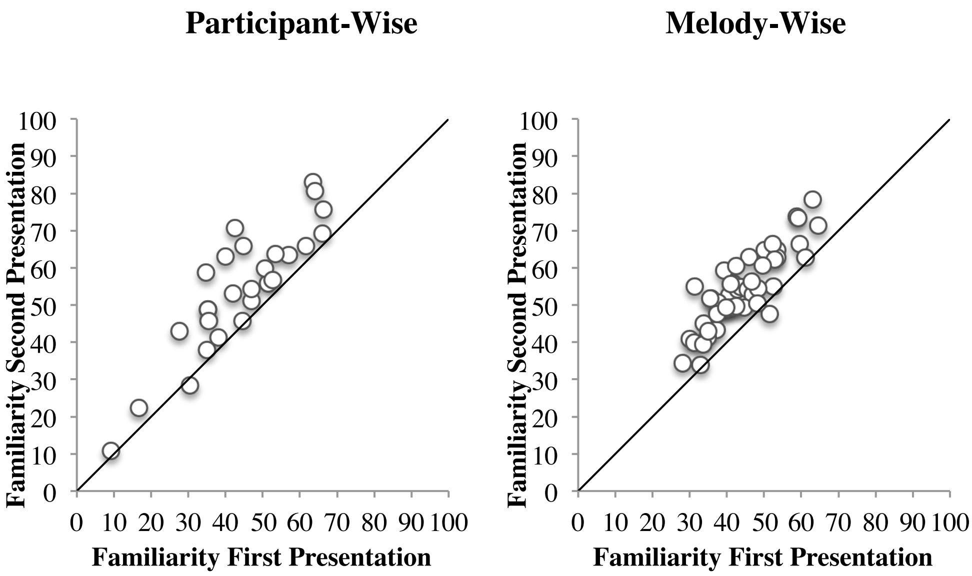


Figure 3.3: Perceived familiarity raw data for Experiment 6. The left panel shows participant-wise differences in perceived familiarity between the first and second occurrence of a melody. The right panel shows melody-wise differences. Overall, second presentations of melodies elicited significantly higher familiarity ratings than first presentations.

A linear mixed effects model investigated whether the previously observed disruption of a recency-in-memory effect with melodies in an unfamiliar tuning system also manifests using an indirect memory task. Disruption of a recency-in-memory effect was indeed observed. Coefficient assessment showed that the number of intervening melodies did not elicit statistically different changes in perceived familiarity (all p-values > .10). *Figure 3.4* shows the mean familiarity ratings for each number of intervening melodies. A model predicting the change in perceived familiarity between the first and second occurrence of the melodies using a random intercept for *Subject*, and *Melody* (*BIC* = 12937; *LogLik* = -6453.9), did not improve significantly when provided with the *Number of Intervening Items* (*BIC* = 12942; *LogLik* = -6453.2, p = .225).

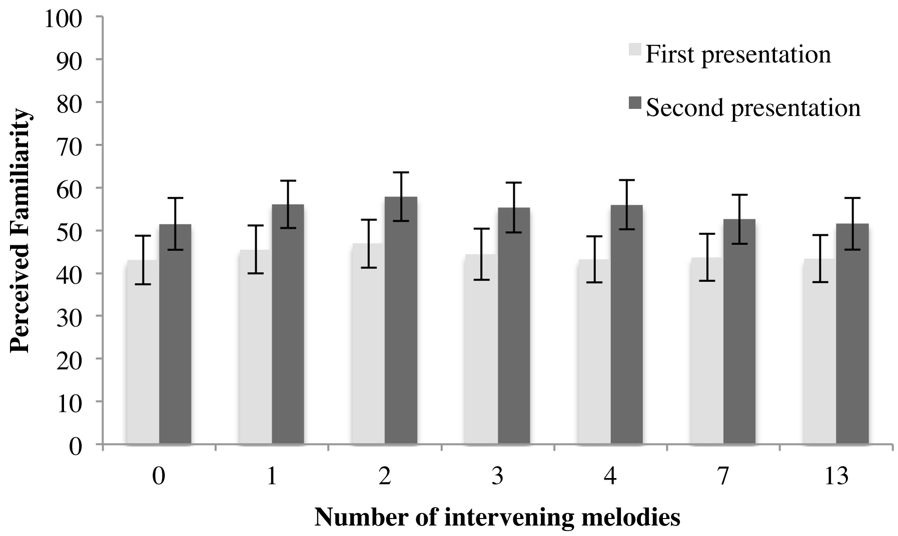


Figure 3.4: Perceived familiarity raw data for Experiment 6. The light bars show perceived familiarity on the first presentations of melodies, and the dark bars show perceived familiarity on the second presentations of the associated melodies. Error bars represent the standard error. Familiarity was significantly higher on second presentations. However, there were no statistically significant differences between the different conditions of intervening melodies.

3.5.3 Discussion

Experiment 6 replicated the results of Experiment 5 using an indirect memory task. As hypothesized by the RMR conjecture, no recency-in-memory effect for the last memory encountered was observed when using melodies in an unfamiliar tuning system. This suggests that preventing recency-in-memory effects in memory for melody by using an unfamiliar tuning system is not limited to conscious recognition task instructions, but instead generalizes to indirect memory task instructions. Consistent with Experiment 5 and not supporting the prediction of the RMR conjecture, the number of intervening items had no statistically significant disruptive effect. However, the non-significant effect could still be due to the small number of intervening melodies (up to thirteen).

It is also possible that listeners were capable of perceiving the stimuli as slight aberrations of melodies in the musical tradition they are familiar with. The stimuli used in Experiment 5 and 6 were based on European folk song melodies and therefore closely follow Western musical tradition on the basis of many musical features (e.g., contour, rhythm, equal temperament tuning). The main difference of the stimuli here, when compared with the original melodies in Herff et al. (2017) is the step-size of the tuning system. It appears this manipulation was sufficient to disrupt the recency-in-memory effect, but potentially the stimuli were too close to the familiar tuning system to disrupt integration of the melodies as a whole. Experiment 7 therefore investigates cumulative disruptive effects over a larger range of intervening melodies, using stimuli that are distinctively different to those that are typically familiar to Western listeners.

3.6 Experiment 7 – Disruptive Memory Effects at Larger Numbers of Intervening Items

Experiment 7 investigates the possibility that melodies in a distinctively unfamiliar tuning system elicit cumulative disruptive effects over large numbers of intervening melodies (beyond the 13 used in Experiments 5 and 6). The data from Experiment 7 are from a larger study conducted at Murdoch University, Australia. The study investigated statistical learning of artificial grammars in the context of music, specifically grammars applied to pitch and rhythm. During the procedure, participants were first exposed to melodies in an unfamiliar tuning system. The exposure phase in this project consisted of a continuous recognition paradigm. This is the same paradigm that Experiment 5, Experiment 6, and previous studies (e.g., Herff et al., 2017) used in the context of melody recognition. After the exposure phase, participants completed various follow-up experiments (data not reported here). The present investigation analyzed only the recognition data of the exposure phase.

During the continuous recognition paradigm of the exposure phase, participants listened to melodies in an unfamiliar tuning system presented three times throughout the experiment and provided recognition responses to each melody. The large number of intervening melodies (up to 107) and the two repetitions of each melody allows us to assess potential cumulative effects from the number of intervening items between both first and second, as well as second and third presentations of a melody.

3.6.1 Method

Participants. One-hundred-and-five participants (largely undergraduate students) were recruited from the Murdoch University community (M_{age} = 25 years, SD_{age} = 7.5, 36 male / 68 female). Participants' musical training ranged from 0 to 17 years ($M_{training}$ = 2.1 years, $SD_{training}$ = 3.3, 32 participants with more than 2 years of musical training). Participants did not previously participate in Experiments 1-6. Participation was either volunteered or reimbursed with course credit.

Stimuli and equipment. The unfamiliar tuning system used here consisted of the following pitch heights: 480, 520, 560, 605, and 665Hz. Note durations of 60, 110, 550, and 920 ms, with a 100 ms silent gap between notes were used. Melodies consisted of five or six notes. All notes were synthesized pure tones with 10 ms linear onset and offset ramps. This resulted in melodies that did not conform to Western music tradition in rhythm and tonality.[5] Durations and pitch discriminability was piloted to ensure that all pitch and duration differences were clearly discriminable ($N = 9$).[6] Using the tonal affinity model (Milne, 2013; Milne et al., 2011), this tuning system shows cosine similarity of .16186 to 12-TET and .22268 to the 88.08-CET system used in Experiment 5 and 6. This means the unequally tempered tuning system used in Experiment 7 is less similar to 12-TET than the tuning system used in Experiment 5 and 6.

Procedure. Participants gave informed consent, were seated in front of a computer, and completed a short demographic questionnaire. Similar to Experiment 5, participants were instructed that they would hear many different melodies, one after another, and it was their task to indicate whether they had heard a given melody in this experiment before. Participants responded by pressing one of two keys on the keyboard. The keys were counterbalanced between participants. The presentation

---

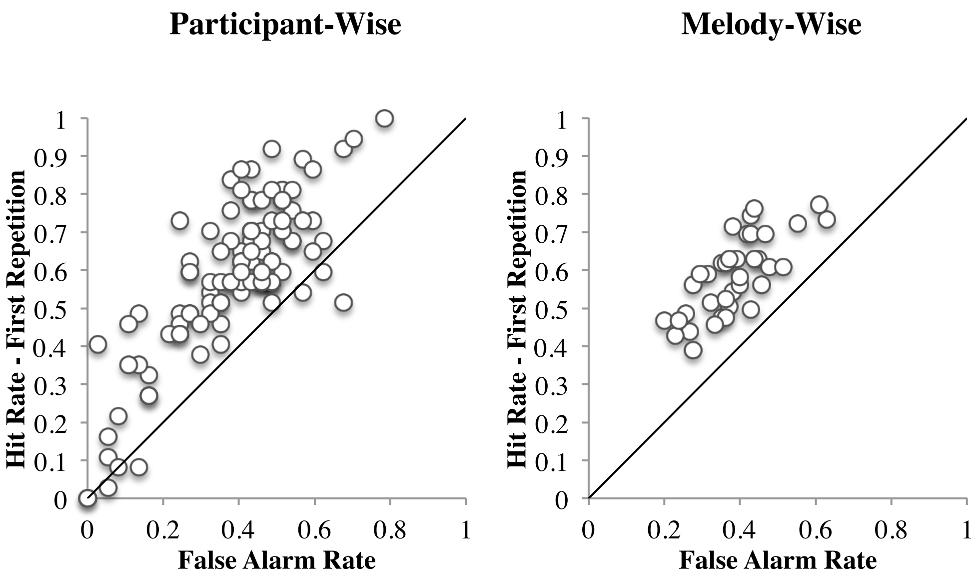
[5] The melodies were constructed in various artificial grammars, unfamiliar to listeners, which were of importance for the later investigation of statistical learning that will be reported elsewhere.

[6] In pitch perception and a standard/comparison task with a silent retention interval, a Weber fraction of .04 led to discrimination performance above 90%. The pitch Weber fractions used in the present stimuli were between .08 and .10, therefore clearly discriminable. In duration perception, Weber fractions of .30 led to discrimination performance above 98%. The duration fractions used in this study ranged from .67 to 1.27, therefore also clearly discriminable.

order of the melodies was fully randomized. Thirty-seven melodies appeared three times throughout the experiment for a total of 111 trials.

3.6.2 Results

Figure 3.5 shows melody- and participant-wise performance. Overall, participants performed significantly above chance ($Z = 24.27$, $p < .0001$).[7]



[7] Note that the here observed cumulative disruptive interference between the first and second, as well as second and third presentations of the melodies is also observed in both groups when participants with fewer or more than two years of musical training are analyzed separately (all $p < .05$).

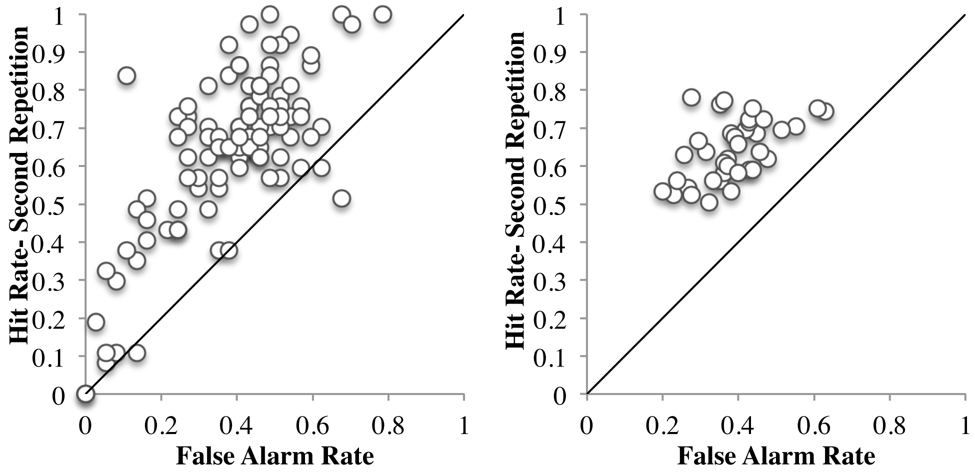


Figure 3.5: Hit rates and false alarm rates in Experiment 7. The left panels show the data participant-wise, and the right panels melody-wise. The top row shows data from the first and second presentations (i.e., first repetition). The bottom row shows data from the first and third presentation (i.e., second repetition). The reference line represents chance level. Overall, performance was significantly above chance and higher on the second repetition of the melodies.

A generalized mixed effects model investigated the hypothesis that, in contrast to Western-tonal melodies, melodies in an unfamiliar tuning system elicit cumulative disruptive effects of the number of intervening melodies. This hypothesis was supported between the first and second occurrence of a melody. A model predicting 'old' responses on melody repetitions using a random intercept for *Subject* and *Melody*, as well as a systematic factor for *Dynamic Response Tendency* (*BIC* = 4777.2; *LogLik* = -2372.1), showed a significant improvement when provided with the *Number of Intervening Items* (*BIC* = 4759.4; *LogLik* = -2359, $p < .0001$, ΔBIC = 17.8).

The hypothesis was also supported between the second and third occurrence of a melody. A model predicting 'old' responses on melody repetitions using a random intercept for *Subject*, and *Melody*, as well as a systematic factor for *Dynamic Response Tendency* (*BIC* = 4542.6; *LogLik* = -2254.8), significantly improved when provided with the *Number of Intervening Items* (*BIC* = 4532.1; *LogLik* = -2245.4, $p < .0001$, ΔBIC = 10.5). *Figure 3.6* depicts the bias free probability of recognition responses as the number of intervening items increased between the first and second, as well as second and third presentation of a melody.

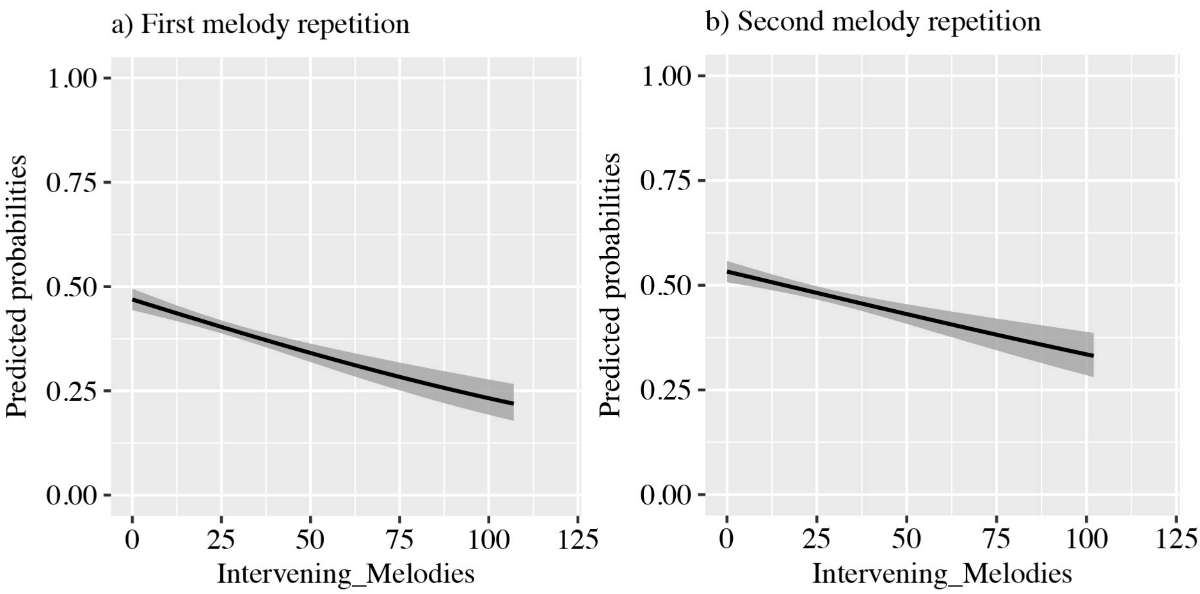


Figure 3.6: Prediction lines of generalized mixed effects models that predict the probability of bias corrected recognition (y-axis) in Experiment 7. The left panel shows the effect of the number of intervening melodies between the first and second presentations of the melodies. The right panel shows the effect between the second and third presentations. A significant disruptive effect of the number of intervening items on bias corrected recognition performance is observed in both periods. The grey area around the prediction line represents a 95% confidence interval.

Delay, Number of Intervening Melodies, or Stimuli. We have shown cumulative disruptive interference in Experiment 7 but not in Experiment 5 and 6.

However, it is not clear whether the effect in Experiment 7 was driven by the larger number of intervening melodies, the larger temporal delay between first and second presentation of the melodies, or the greater unfamiliarity of the musical system behind the stimuli. To disentangle these possible explanations, we conducted a secondary analysis of the 1 to 13 intervening melody conditions in Experiment 7. Observing cumulative disruptive interference in these conditions would show that the effect is due to the stimuli, rather than the larger temporal delay or a greater number of intervening items. This is because the stimuli in Experiment 7 are shorter compared to those used in Experiment 5 and 6 (thus providing the opportunity to exclude greater temporal delay as an explanation). Furthermore, this analysis used the same number of intervening melodies (thus providing the opportunity to exclude greater number of intervening melodies as an explanation).

Between the first and second presentation of the melodies, a model predicting 'old' responses on melody repetitions using a random intercept for *Subject* and *Melody*, as well as a systematic factor for *Dynamic Response Tendency* (*BIC* = 1588.6; *LogLik* = -779.94), showed a significant improvement when provided with the *Number of Intervening Items (Coef*: -.05*,* $p = .0013$*)*(*BIC* = 1585.4; *LogLik* = -774.77, $p = .0013$, ΔBIC = 3.2).

The same was observed between the second and third presentation of the melodies, where a model predicting 'old' responses on melody repetitions using a random intercept for *Subject* and *Melody*, as well as a systematic factor for *Dynamic Response Tendency* (*BIC* = 1625.5 *LogLik* = -798.4), showed a significant improvement when provided with the *Number of Intervening Items (Coef*: -.01*,* $p < .0001$*)*(*BIC* = 1617.2; *LogLik* = -790.68, $p < .0001$, ΔBIC = 8.3).

3.6.3 Discussion

Experiment 7 aimed to investigate cumulative effects at relatively large numbers of intervening melodies using an unfamiliar tuning system. It was hypothesized that with up to 107 intervening melodies tested here, a cumulative disruptive effect of intervening melodies should be observed. Such a disruptive effect is common in recognition in general (Bui et al., 2014; Donaldson & Murdock, 1968; Herff et al., 2017; Hockley, 1992; Konkle et al., 2010; Olson, 1969; Poon & Fozard, 1980; Rakover & Cahlon, 2001; Sadeh et al., 2014), but has not previously been observed in the context of melody recognition when a familiar tuning system with up to 195 intervening melodies is presented (Herff et al., 2017). Experiment 7 supported this hypothesis. Participants could perform the task, but a significant disruptive effect on recognition performance unfolded as the number of intervening melodies increased. This effect was still present when participants' shifts in response tendencies and random effects of melody were accounted for. A secondary analysis of the data suggests that the cumulative disruptive effects in Experiment 7 are predominantly driven by the stimuli rather than the larger temporal delay or greater number of intervening items.

3.7 General Discussion

In general, a disruptive effect on memory from the number of intervening items between first and second presentation of a target stimulus has been observed in stimuli from a variety of domains (Bui et al., 2014; Donaldson & Murdock, 1968; Hockley, 1992; Konkle et al., 2010; Olson, 1969; Poon & Fozard, 1980; Rakover & Cahlon, 2001; Sadeh et al., 2014). However, recognition of melodies in a familiar tuning system does not appear to be affected by commonly reported interference from intervening items (Herff et al., 2017). In three experiments, the present study further investigated this phenomenon by manipulating the number of intervening items between first and second presentation, as well as between second and third presentation of target melodies, sounded in an unfamiliar tuning system. A recent RMR conjecture (Herff et al., 2017) predicts a disruptive effect of the number of intervening items when melodies are presented in an unfamiliar, rather than familiar, tuning system. Furthermore, the conjecture predicts no recency-in-memory advantage for immediate melody repetition. The results here showed no recency-in-memory effect for recognition of melodies presented in unfamiliar tuning systems. A disruptive effect from the number of intervening melodies was only observed in Experiment 7 (up to 107 intervening melodies) and not in Experiment 5 and 6 (up to 13 intervening melodies). The present findings will now be discussed in the context of recency-in-memory and effects from the number of intervening melodies.

3.7.1 Effects from Recency-in-Memory

Recency-in-memory phenomena describe memory advantages for the last encountered item (Jahnke, 1963; Roberts, 1986). In the context of music, such an advantage for the last item has been previously demonstrated in recognition of single notes as well as melodies in a familiar tuning system (Berz, 1995; D. Deutsch, 1970, 1975; Dowling, 1973; Greene & Samuel, 1986; Herff et al., 2017). In order to explain recognition advantages for the last encountered melody with recency-in-memory effects, it is necessary to assume that melodies are perceived and represented as integrated entities (see Krumhansl, 1991; D. Deutsch, 1986). The

RMR conjecture predicts that if a melody is not perceived and integrated as a whole, then only a recency-in-memory effect for the last underlying part of a melody remains (e.g., single notes or phrases). This was tested here using melodies in an unfamiliar tuning system designed to disrupt perception and integration of melodies as a whole.

Specifically, the RMR conjecture assumes that prior experience influences perception and that perception directly influences formation of new memories. If prior experience does not exist (e.g., of pitch structure in an unfamiliar tuning system) and therefore does not inform the process of integrating a melodic sequence of notes into a perceptually coherent melody, then no memory representation of a perceptually integrated melody will be formed. Instead, memory is based only on representations of the melody's underlying components. As a result of the missing representation of an integrated melody, a recency-in-memory effect for the last encountered melody should not emerge. The present data support this hypothesis. Recognition performance in Experiment 5 was statistically similar between immediate melody repetition and up to thirteen intervening melodies. In Experiment 6, the change in perceived familiarity was also statistically similar. These results also support the findings of previous studies that show recency-in-memory effects and effects from the number of intervening items on memory for melody are similarly captured in direct recognition tasks and indirect perceived familiarity tasks (Herff et al., 2017).

Of relevance to future investigations is an unpublished pilot study conducted by the first author that tested the design of Experiment 5 in a small sample ($N = 14$) of high-expertise musicians (> 10 years of formal training), in contrast to the less experienced participants in Experiment 5. Interestingly, the sample of musicians also did not show a recency-in-memory effect (see Appendix E, p. 266). Empirical testing of populations with varying degrees of musical expertise or exposure to specific tuning systems is an important avenue for further research. The RMR conjecture predicts that regardless of tuning system, listeners familiar with the system should show a recency-in-memory effect, whereas listeners with no experience in the system should show no recency-in-memory advantage. This could provide a phenomenon-driven approach to explore cultural differences in the perception of auditory stimuli.

3.7.2 Effects from the Number of Intervening Melodies

As predicted by the RMR conjecture, the hypothesized disruptive effect from the number of intervening items was observed in Experiment 7. This experiment used up to 107 intervening melodies. In contrast, Experiments 5 and 6 investigated intervening gap sizes of only up to 13 melodies, and neither produced significant disruptive effects from the number of intervening items. Interestingly, in Experiment 7 disruptive effects from the number of intervening items were observed between the first and second and between the second and third presentations of the melodies. These results show that, once observed, the effect is fairly robust even when repetition strengthens representations.

The methodological differences between Experiment 7 and the previous experiments provide alternative explanations for our findings. First, Experiment 7 used large numbers of intervening items, whereas Experiment 5 and 6 investigated a relatively small number. Second, Experiment 7 used a different melody corpus. The melodies in Experiments 5 and 6 were played with a piano timbre and used pitch contours and rhythms common to melodies in familiar music traditions such as Western tonal music. In Experiment 7, the melodies were played using pure tones rather than a musical timbre. However, this difference in timbre most likely did not drive the effects because, so far, only differences in vocal timbre have previously been shown to impact memory when compared to other musical timbres (Weiss, Vanzella, Schellenberg, & Trehub, 2015). Nevertheless, the effect of timbre on memory for melodies is a topic that deserves greater attention in future studies.

The unfamiliar tuning system utilized in Experiments 5 and 6 was the most dissimilar equal-tempered tuning system when compared to 12-TET (within the given range specified in 3.4.1). However, it is still an equal tempered tuning system and therefore carries some perceptual similarity to 12-TET. Indeed, the cosine similarity measures reveal that the non-equal tempered tuning system used in Experiment 7 is decidedly more dissimilar to 12-TET than the tuning system used in Experiments 5 and 6. Experiment 7 also used a far less musical timbre (pure tones), and rhythms were not matched directly to common and familiar rhythms in Western tonal music as they were in Experiments 5 and 6. This means that they did not adhere to regular metrical structures and overall pitch contour. The stimuli in Experiment 7 can therefore be defined as far more extreme in terms of unfamiliarity than the stimuli in Experiments 5 and 6. It could be that the tuning system in Experiments 5 and 6 was unfamiliar enough to disrupt recency-in-memory effects, but not unfamiliar enough to induce effects from the number of intervening melodies (at least when the number was ≤ 13 melodies). The melodies in Experiment 7, being far less familiar than those in Experiment 5 and 6, were likely unfamiliar enough to disrupt integration entirely.

To further investigate which factors may explain the effects reported in Experiment 7, we performed a secondary analysis of the 1 to 13 intervening melody conditions in Experiment 7. We found cumulative disruptive interference in these conditions, even though the stimuli in Experiment 7 were shorter than those in Experiments 5 and 6. Intuitively, one could predict that shorter temporal delay associated with shorter melodies should lead to less interference, however, another explanation is that shorter stimuli are more vulnerable to inference. In general, this finding shows that the cumulative disruptive effect in Experiment 7 is indeed stimulus-driven, rather than driven by the larger number of intervening melodies and the associated greater temporal delay.

These findings suggest that the issue of listeners' unfamiliarity with different musical features and its effect on recency-in-memory and disruptive effects from intervening items is a fruitful future research endeavor. In general, the RMR conjecture predicts smaller effects from recency-in-memory and a larger disruptive effect from the number of intervening items as a tuning system's unfamiliarity increases. A future systematic manipulation of individual musical features while measuring the aforementioned memory phenomena has the potential to shed light on fundamental memory processing in general, and how previous experience influences perception and formation of subsequent memories in particular. This line of investigation has begun by investigating the relative influence of pitch information and rhythm information on memory for melodies (Herff, Olsen, et al., 2017), and will continue by systematically manipulating individual musical features and their associated impact on memory.

3.8 Conclusion

The aim of this study was to deepen our understanding of memory for melody by investigating the Regenerative Multiple Representations (RMR) conjecture using recognition of melodies in an unfamiliar tuning system. Previously, recognition of melodies in a familiar tuning system has shown recency-in-memory effects, but not cumulative disruptive effects from the number of intervening melodies (Herff et al., 2017). The RMR conjecture predicts opposite findings for melodies in an unfamiliar tuning system. The present data supports these predictions using direct as well as indirect recognition tasks: no recency-in-memory effects were observed, but cumulative disruptive effects from the number of intervening melodies in memory for melodies in an unfamiliar tuning system were observed only when the number of intervening items increased from 13 up to 107.

The findings of the present study help advance our understanding of memory's stimulus specificity (Fougnie et al., 2015) by exploring why some domains show clear disruptive effects from the number of intervening items, whereas others do not. However, future research is needed to investigate precise degrees of prior experience and its effect on both recency-in-memory and interference from the number of intervening items. Much research is also needed to extend the present findings to stimuli beyond the domain of music. Finally, the RMR conjecture appears to be a useful conceptual tool to guide future research investigating the critical link between prior experience, perception, and subsequent formation of memory representations. An investigation of how prior experience influences perception, how this perception determines formation of representations, and how these representations are then affected by basic memory phenomena has the potential to shed important light on the fundamental relationship between human memory, perception, and ultimately, day-to-day behavior.

Chapter 4

Pitch-only and Rhythm-only Sequences

Chapter 4 has been peer reviewed and published:

Herff, S.A., Olsen, K.N., Prince, J., & Dean, R.T. (2017). Interference in memory for pitch-only and rhythm-only sequences. *Musicae Scientiae*, Advance online publication. doi:10.1177/1029864917695654

Note: Kirk N. Olsen and Roger T. Dean are the author's supervisors. Jon Prince shared some recently collected data that is analysed here.

4.1 Abstract

In human memory, the ability to recognize a previously encountered stimulus often undergoes cumulative interference when the number of intervening items between its first and second presentation increases. Although this is a common effect in many domains, melodies composed in tuning systems familiar to participants (e.g., Western tonal music) do not seem to suffer such cumulative decrements in recognition performance. Interestingly, melodies in unfamiliar tuning systems do show cumulative decrements. This finding has been predicted by a novel Regenerative Multiple Representations (RMR) conjecture. The present study further explores this phenomenon and the conjecture by investigating pitch-only (isochronous rhythm) and rhythm-only (monotone pitch) sequences of melodies in an unfamiliar tuning system that previously showed cumulative disruptive effects. Experiment 8 replicated previous studies reporting significant interference effects from the number of intervening items when melodies use uncommon rhythms and are composed in an unfamiliar tuning system. Furthermore, as predicted by the RMR conjecture, when rhythmic information was neutralized (Experiment 9), the cumulative interference related to the number of intervening items was retained. This was also the case when the original pitch information of each melody was neutralized, leaving variation only in the rhythmic information (Experiment 10). Results are discussed in the context of the RMR conjecture: given converse results, the conjecture would have been falsified. However, it currently remains plausible and appears to be a useful tool for precise predictions about the link between prior experience, perception, and formation of new memories.

4.2 Interference in Memory for Pitch-only and Rhythm-only Sequences

In a variety of stimulus domains, recognition of a previously encountered stimulus commonly undergoes cumulative disruption as the number of intervening items between its first and second presentation increases (Buchsbaum et al., 2011; Bui et al., 2014; Campeanu et al., 2014; D. Deutsch, 1970, 1975; Donaldson & Murdock, 1968; Hockley, 1992; Konkle et al., 2010; Olson, 1969; Poon & Fozard, 1980; Rakover & Cahlon, 2001; Sadeh et al., 2014). In the musical domain, however, melodies composed in culturally familiar tuning systems (e.g., Western tonal music) do not show cumulative decrements in recognition performance as time elapses (Schellenberg & Habashi, 2015), nor systematic and cumulative decrements as the number of intervening items increases (Herff et al., 2017; Herff, Olsen, et al., 2017). Interestingly, interference effects that are due to the number of intervening items *are* observed if melodies are sounded in unfamiliar tuning systems (Herff, Olsen, Dean, & Prince, 2017). A Regenerative Multiple Representations (RMR) conjecture has been developed to explain this disparity (see below for more detail).

In the present study, three experiments were designed to further investigate the predictions of the RMR conjecture in the context of memory for melodies in unfamiliar tuning systems. This was achieved by separating the pitch and rhythmic components of the melodies from Herff, Olsen, et al. (2017) and testing for cumulative disruptive effects. Here, the components were separated by using a constant pitch in the rhythmic-only sequences (i.e., monotone pitch), and a constant note duration (i.e., isochronous) in the pitch-only sequences. The RMR conjecture would be fundamentally refuted if pitch-only and rhythm-only sequences did not show cumulative disruptive interference when the combined melodies do. However, in the present investigation, both sequences independently did show cumulative disruptive effects. Before discussing the present experimental manipulations in more detail, an overview of the conjecture and its relevance to memory for melody will first be presented.

4.2.1 Regenerative Multiple Representations Conjecture

The RMR conjecture describes a crucial link between prior experience, perception, and subsequent formation of memories. Prior experience provides information about the most useful way of perceiving our environment (similar to D. Deutsch, 1986, see also J. A. Deutsch and Deutsch, 1963). In turn, perception influences the formation of memories. In other words, we first learn the most relevant way of perceiving our environment. We then perceive objects according to this information and form memories according to the perception. Therefore, the conjecture assumes that if useful, information from prior experience can change a single percept into multiple percepts, depending on the stimulus that is being perceived. For example, an empty coffee cup is most usually perceived as a useful device to satisfy the need for a refreshing beverage. However, if for some reason an individual is in need of a projectile, the coffee cup may be perceived as a potential candidate to throw. Affordance theory (Gibson, 1977, 1978) describes the fact that humans not only perceive objects as their underlying components such as object shape, but also perceive possible actions that can be performed with an object. According to the RMR conjecture, however, additional percepts also lead to additional memory representations. In the example above, the coffee cup would be perceived as a useful drinking device *as well as* a potential projectile, leading to multiple memory representations of the same coffee cup. Interestingly, the two representations would code partially overlapping information, for example the weight of the coffee cup. These memory representations are then subject to decay (time) and interference (e.g., number of intervening items) (Norman, 2013; Oberauer & Lewandowsky, 2011; Oberauer et al., 2012). However, in cases where multiple representations code partially overlapping information, the RMR conjecture posits that representations can regenerate each other, thus providing resilience to decay and interference in the context of memory (similar to the Dual-Coding theory of Paivio, 1969).

4.2.2 Regenerative Multiple Representations in the Context of Music

Music in general and melodies in particular provide a rich context in which to investigate the RMR conjecture. This is because melodies can be perceived as integrated whole melodies, as well as underlying components such as notes, intervals, and short phrases (see D. Deutsch, 1986; Krumhansl, 1991, p. 295; Margulis, 2012; Schneider, 1997, p. 119). In this case, the RMR conjecture predicts formation of multiple representations of a target melody. These multiple representations of a melody share some information. For example, representations of intervals, short phrases, or a rhythm co-relate to a representation of an integrated, coherent melody. Once encoded, the multiple memory representations that co-relate in the form of partially overlapping information provide a melody representation with resilience towards decay and interference. This resilience could be realized in form of strong melody expectancies that listeners form when they are familiar with the underlying rules of a music tradition. Melodic or rhythmic expectancies can then be used to interpolate forgotten parts of a melody, similar to how they are used to predict what comes next in a melody (Margulis, 2005; Pearce, 2014; Schellenberg, 1996).

For listeners encultured in Western music, melodies composed in this tradition do not show memory decay effects even after delays of up to a week (Schellenberg & Habashi, 2015). In terms of interference, neither recognition nor perceived familiarity of such melodies show cumulatively disrupted interference effects from intervening items, even with multiple corpora of music and up to nearly 200 intervening melodies (Herff et al., 2017). In four experiments, Herff et al. (2017) presented novel melodies in a familiar tuning system to participants. After each melody presentation, participants indicated whether or not they had heard the melody in the experiment before. The number of intervening melodies until a target melody reappeared was manipulated between zero and up to nearly 200. Participants consistently performed significantly above chance. Furthermore, the probability of producing a correct recognition judgment was statistically identical between 1 and up to nearly 200 intervening melodies. The only exception was zero intervening melodies (immediate repetition). However, this pattern does not hold for all melodies.

Take the case of novel melodies composed in an unfamiliar tuning system (i.e., incompatible with the Western tonal tradition). As described above, the RMR conjecture assumes that knowledge is acquired from past experience. Without this experience, listeners cannot use such information to integrate notes, rhythm, intervals, and short phrases into coherent musical melodies. As a result, this information cannot influence their perception and subsequent memory representation, and no memory representation of the stimulus as an integrated musical melody will be formed.

In this context, the RMR conjecture predicts that melodies in unfamiliar tuning systems should elicit cumulative disruption of recognition memory from the number of intervening melodies, because they are not integrated into coherent memory representations. Indeed, such a cumulative disruption of recognition has been observed in a recent study as the number of intervening melodies increased up to ~100 intervening melodies (Herff, Olsen, et al., 2017). Three experiments tested two unfamiliar tuning systems, and similar to Western tonal melodies, recognition performance was ubiquitously above chance. Dissimilar to Western tonal melodies, performance decreased as the number of intervening melodies increased.

This finding provided preliminary evidence for the RMR conjecture and, at the same time, motivated further empirical testing of the conjecture that we report here. Specifically, we examine cumulative disruption effects for separate components of melodies composed in an unfamiliar tuning system. The most obvious components to test are the pitch and rhythmic patterns that comprise a melody, as these are the primary dimensions of melodic sequences. In an unfamiliar tuning system, pitch and rhythmic information should not be integrated into coherent melodies and thus should elicit the cumulative disruptive effects described earlier. The RMR conjecture predicts that this cumulative disruptive effect should also emerge when either pitch or rhythmic information is learned (from pitch-only sequences and rhythm-only sequences, respectively: see *Figure 4.1* for examples of such sequences).

To create *pitch-only sequences*, we present each note in a set of melodies for the same duration, keeping all inter-note durations identical and effectively neutralizing rhythm information. Note that participants may still perceive rhythm due to perceived segmentation that may be evoked by the pitch-sequences' melodic contours (Brochard, Abecasis, Potter, Ragot, & Drake, 2003; D. Deutsch, 1986). Nevertheless, the actual temporal placement of notes in all pitch-only sequences was identical.

To create *rhythm-only sequences,* we present each note in a set of melodies at the same pitch (frequency) while retaining its original rhythmic structure, thus effectively neutralizing the pitch information. In this context, the pitch- and rhythm-only sequences may show more interference than the original combined versions because there are fewer possible representations for the listeners. However, a lack of cumulative disruption on recognition of pitch- and rhythm-only sequences would provide strong evidence against the RMR conjecture.

Figure 4.1: Example of the stimulus manipulations used in the study. Experiment 8 used melodies that consist of a combined melodic and rhythmic sequence. Experiment 9 used the pitch-only sequence of the original stimuli. Experiment 10 used the rhythm-only sequence of the original stimuli. Note that this figure is only an example of how the stimuli were manipulated. The actual stimuli presented in the study were melodies in an unfamiliar tuning system, and with more irregular note IOIs, as described in the Methods section.

4.2.3 Memory for Pitch-Only Sequences

When melodies in a familiar tuning system are separated into their component pitch and rhythm sequences, listeners are better at recognizing when they hear a pitch-only sequence compared to a rhythm-only sequence (Hebert & Peretz, 1997; White, 1960). Dowling et al. (1995) conducted a thorough investigation of recognition for pitch-only melodies. They used seven-note isochronic melodies but no cumulative disruptive effect on recognition was observed with up to 8 intervening items. However, recent studies have shown that relatively large numbers (~100) of intervening items are required to appropriately assess cumulative disruptive effects on recognition of musical stimuli (Herff et al., 2017; Herff, Olsen, et al., 2017). Here we investigate relatively large numbers of intervening items as well as melodies in an unfamiliar tuning system.

Melodies containing their original pitch *and* rhythm information, composed in an unfamiliar tuning system, elicit cumulative disruptive effects on recognition (Herff, Olsen, et al., 2017). The RMR conjecture predicts that removing rhythmic information from these melodies means listeners will perceive fewer underlying components of the melodies upon first presentation, compared to listeners who hear melodies with both pitch and rhythm information. Therefore, listeners that hear the pitch-only version form fewer representations of the melodies. Consequently, the pitch-only sequences should elicit cumulative disruptive effects on recognition from intervening melodies. It is this prediction that is tested here in Experiment 9.

4.2.4 Memory for Rhythm-Only sequences

As mentioned above, pitch information in a melody tends to be a better cue for melody recognition than rhythmic information (Dowling et al., 1995; Hebert & Peretz, 1997; White, 1960). Nevertheless, rhythms do significantly contribute to recognition of melodies. This is demonstrated in higher recognition performance when rhythm and melody are combined, compared to conditions where only one varies (Hebert & Peretz, 1997). In terms of memory decay, recognition performance in response to rhythms decreases in a 'same-different' task as the time interval between first and second presentation of rhythms increases from 1,000 to 7,250 ms (Collier & Logan, 2000). However, literature is sparse in terms of systematic investigations of the effect of the number of intervening rhythms on memory for rhythms.

In Experiment 10, we investigate memory for rhythm-only sequences. In the context of the RMR conjecture, we test whether there are cumulative disruptive effects of the number of intervening rhythms on memory for rhythm-only sequences. The RMR conjecture predicts such disruptive effects by considering that similar to pitch-only sequences, rhythm-only sequences should lead to fewer memory representations upon first encounter compared to the original melodies.

Before investigating the influence of pitch (Experiment 9) and rhythm (Experiment 10) on cumulative disruptive effects in memory for melody, we first conducted an experiment to establish baseline interference effects for the present set of melodies composed in an unfamiliar tuning system when the melodies include all of their original pitch and rhythm information.

4.3 Experiment 8 – Recognition of Melodies in an Unfamiliar Tuning System

4.3.1 Method

Participants. Thirty-seven students were recruited from Murdoch University (M_{age} = 25.3 years, SD_{age} = 8.4, male) and received course credit for their participation. Participants did not previously participate in Experiments 1-7. The mean years of musical training was 1.9 (SD = 3.7).[8]

Stimuli. In total, thirty-seven melodies were composed in an unfamiliar tuning system, consisting of variations of the following pitches: 480, 520, 560, 605, and 665 Hz (pure tones with 10ms linear on/off ramps). Each melody had five or six

---

[8] The experiments reported here are part of a larger investigation of distributional learning of artificial grammars in the context of music. Participants in the present study were subject to various follow up experiments that will be detailed elsewhere.

notes with variations of note durations of 60, 100, 550, and 920 ms, and constant inter-note silent intervals of 100 ms. Neither the rhythm or melodic sequence of notes in each melody conformed to the Western tonal tradition. The pitch and duration discrimination of adjacent values of the pitch and duration was piloted (N = 9) and discrimination performance was above 90% for both. The stimuli of all three experiments can be found in the Supplemental Material Online section (S1–Stimuli.zip).

Procedure. Participants provided informed consent and were instructed that they would hear different melodies one after another. Melodies were presented in a continuous recognition paradigm in random order (Shepard & Teghtsoonian, 1961). Participants were required to indicate which of the melodies have been sounded before in the experiment using two keys on a computer keyboard. The keys were counterbalanced between participants. One key was always associated with 'Old' and one with 'New'. A response of 'New' indicated that a melody was heard for the first time, and a response of 'Old' indicated that the participant believed they had previously heard the melody in the experiment. In total, thirty-seven different melodies were presented three times each. The number of intervening melodies varied between one and 100.

Statistical approach. We used generalized linear mixed effects models to investigate the effect of the number of intervening melodies on binary recognition data (Baayen, 2008; Baayen et al., 2008; Judd et al., 2012; Kass & Raftery, 1995; Kruschke, 2010, 2013; Nathoo & Masson, 2016). The models were implemented in the R software platform (R-Core-Team, 2013) using the lme4 package (Bates et al., 2013) and consisted of the experimental fixed factor *Number of Intervening Melodies*. Random *Participan*t and *Melody* variation was taken into account in the form of random intercepts. Coefficient and *p*-values for *Number of Intervening Melodies* are reported for each experiment. Models that show statistically significant effects from the *Number of Intervening Melodies* are further assessed with a model comparison approach. We use log-likelihood tests to compare models with the experimental factor *Number of Intervening Melodies* against a similar models without the experimental factors of *Number of Intervening Melodies* (Wilks, 1938).

In each experiment, mixed effects models assess overall performance by comparing recognition in response to the first presentation of each melody with recognition in response to their second presentation (*Melody Presentation)*, while controlling for random effects of melody and participant. *Z*-score and coefficient *p*-values for *Melody Presentation* are reported at the beginning of each results section to report whether performance in each experiment was above chance.

Similar to previous research, response tendency shifts were taken into account in the form of conservative *Dynamic Response Tendency* models (Herff et al., 2017; Herff, Olsen, et al., 2017). These models account for any shift in participant response tendencies as the recognition experiment progresses (e.g., Berch, 1976; Donaldson & Murdock, 1968; Snodgrass & Corwin, 1988). For example, some participants may always respond ‘old’ when in doubt in the beginning of the experiment, however, over the course of the experiment their response tendency may change to always respond ‘new’ when in doubt.

To take this into account, we use generalized mixed effects models that were trained on 'old' responses to first melody presentations (False Alarms) as a function of trial number. The fitted models were then used to predict the probability of pressing 'old' on the second presentation of each melody (Hits) as a function of trial number. These predictions were then implemented as a fixed factor in the statistical assessment of the data.

4.3.2 Results

Figure 4.2 shows melody- and participant-wise performance. Overall, participants performed significantly above chance ($Z = 13.85$, $p < .001$).

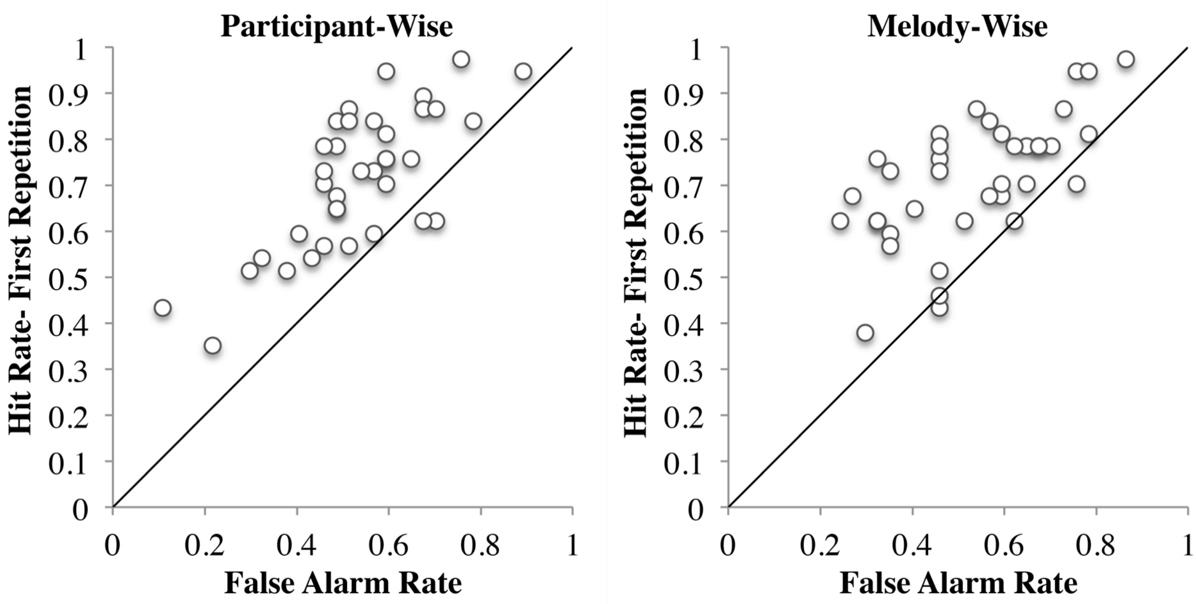


Figure 4.2: Hit rates and false alarm rates in response to melodies in an unfamiliar tuning system used in Experiment 8. The left panel shows the data participant-wise, and the right panel melody-wise. The reference line represents chance level. Overall, participants performed significantly above chance (see text for more detail).

As expected, the number of intervening items between the first and second presentation of each melody had a cumulative disruptive effect on melody recognition when melodies were sounded in an unfamiliar tuning system. A model predicting 'old' responses on melody repetitions using a random intercept for *Participant*, *Melody*, and a systematic factor for *Dynamic Response Tendency* (*LogLik* = -759.03) improved significantly when provided with the *Number of Intervening Items* (*LogLik* = -757.00, *p* = .041). In other words, intervening items showed significant cumulative disruption on participants' melody recognition performance (*coef* = -.008, *p* = .044). *Figure 4.3*a shows the modelled probability of producing bias-corrected recognition as the number of intervening melodies increases between first and second presentation of a melody.

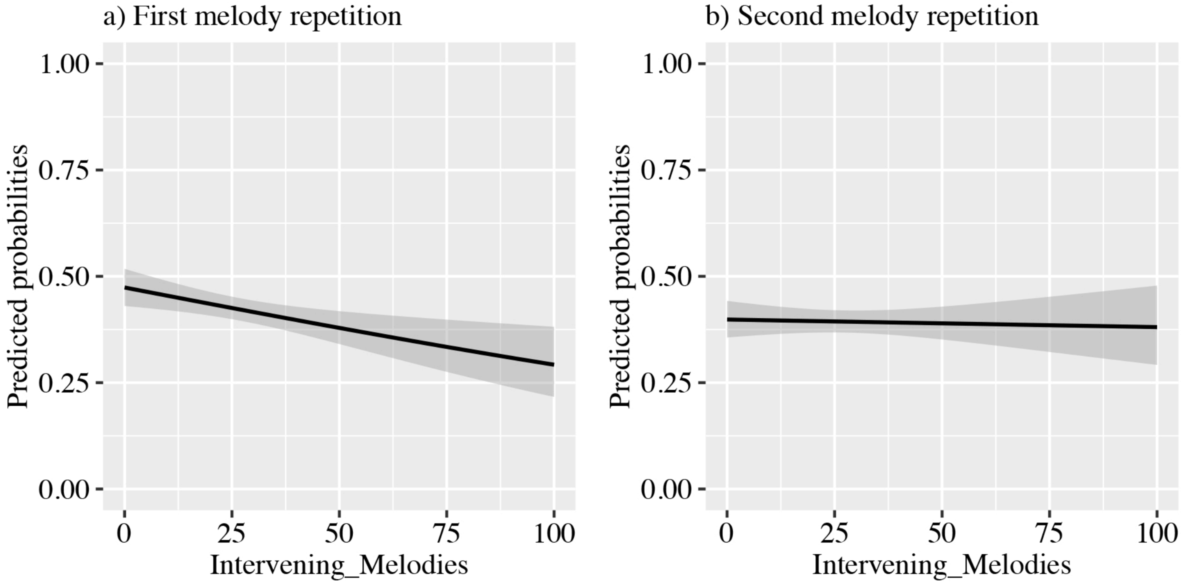


Figure 4.3: Prediction lines of generalized mixed effects models that model the probability of bias corrected recognition (y-axis) of melodies in an unfamiliar tuning system in Experiment 8. The left panel shows the effect of the number of intervening melodies between the first and second presentations of the melodies. The right panel shows the effect between the second and third presentations. A downward slope indicates cumulative disruptive effects. A statistically significant disruptive effect of the number of intervening items on bias corrected recognition performance was only observed between the first and second presentation of the melodies. The grey area around the prediction line represents a 95% confidence interval.

Interestingly, a disruptive effect of the number of intervening melodies did not reach significance between the second and third presentation of the melodies ($coef$ = -.0008, p > .05). *Figure 4.3*b shows the modelled probability of producing bias correct recognition as the number of intervening melodies increases between second and third presentation of the melodies.

4.3.3 Discussion

The goal of Experiment 8 was to establish baseline interference effects when melodies composed in an unfamiliar tuning system include all of their original pitch and rhythm information. This design also enabled a replication of findings observed in Herff, Olsen, et al. (2017), where cumulative disruption of memory for melodies in an unfamiliar tuning system was observed by varying the number of intervening melodies. Interestingly, only a partial replication was achieved in the present study. Specifically, the predicted disruptive effect on recognition from intervening melodies emerged only between the first and second presentation of the melodies, but not between the second and third presentation.

A possible explanation is that the previous studies investigating cumulative effects of the number of intervening items incorporated a greater number of participants (105 compared to 37, Herff, Olsen, Dean, et at., 2017) or a greater number of melodies (110 compared to 37, Herff, Olsen, & Dean, 2017). Thus it could be that the present experiment did not provide enough statistical power to replicate both effects. It is also worth noting that the *Dynamic Response Tendencies* models that corrected for shifts in response tendencies over the course of the experiment resulted in a conservative approach that will have decreased the probability of finding significant effects. Nevertheless, Experiment 8 was successful in its design to provide a baseline to further test melody recognition and the RMR conjecture by using a similar number of participants, melodies, and statistical analysis, but with pitch-only sequences (Experiment 9) and rhythm-only sequences (Experiment 10). The RMR conjecture predicts that cumulative disruptive effects on recognition from the number of intervening melodies is likely to be stronger (or at least equivalent) in sequences where only rhythm or pitch information is available, relative to sequences that retain both rhythm and pitch information together. It is this prediction that is tested in the following experiments.

4.4 Experiment 9 – Recognition of Pitch-only Melodies in an Unfamiliar Tuning System

Experiment 9 investigated disruptive effects from the number of intervening melodies on pitch-only versions of the melodies used in Experiment 8. A disruptive effect from the number of intervening melodies was hypothesized between the first and second presentation of the melodies. If this hypothesis is supported, then the findings will provide support for the predictions afforded by the RMR conjecture. The absence of a disruptive effect would serve as evidence against the RMR conjecture. No clear predictions can be made for disruptive effects of the number of intervening items between the second and third presentations. This is because no such effect was observed in Experiment 8. Only a condition that showed cumulative disruption effects on recognition in Experiment 8 can be used in Experiment 9 and 10 to test the RMR conjecture.

4.4.1 Method

Participants. Thirty-four undergraduate students were recruited from Murdoch University (M_{age} = 23.3 years, SD_{age} = 6.2). The mean duration of musical training was 2.6 years (*SD* = 3.0). Participants were not involved in Experiments 1-8.

Stimuli. The melodies of Experiment 8 were used, however, the original rhythms were removed. All notes in each melody lasted 400 ms with inter-note intervals of 100 ms of silence.

Procedure. The procedure was identical to Experiment 8.

4.4.2 Results

Figure 4.4 shows melody and participant wise performance. Overall, participants performed significantly above chance ($Z = 11.66, p < .001$).

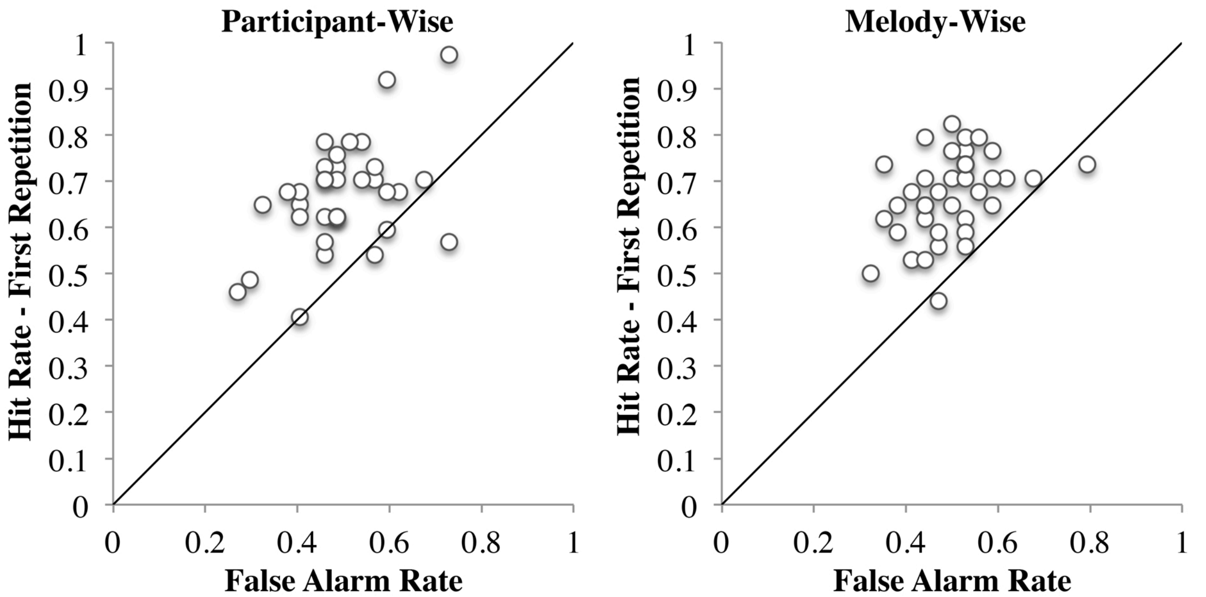


Figure 4.4: Hit rates and false alarm rates in response to pitch-only sequences used in Experiment 9. The left panel shows the data participant-wise, and the right panel melody-wise. The reference line represents chance level. Overall, participants performed significantly above chance (see text for more detail).

As anticipated, the number of intervening items between the first and second presentation of the pitch-only sequences had a cumulative disruptive effect on recognition performance. A model predicting 'old' responses on melody repetitions using a random intercept for *Participant, Melody*, and a systematic factor for *Dynamic Response Tendency* (*LogLik* = -792.37) improved significantly when provided with the *Number of Intervening Items* (*LogLik* = -789.22, p = .012). In other words, the *Number of Intervening Melodies* showed significant cumulative disruption (*coef* = -.009, p = .012) of participants' melody recognition. *Figure 4.5*a shows the modelled probability of producing bias-corrected recognition as the number of intervening melodies increases between first and second presentation of a melody.

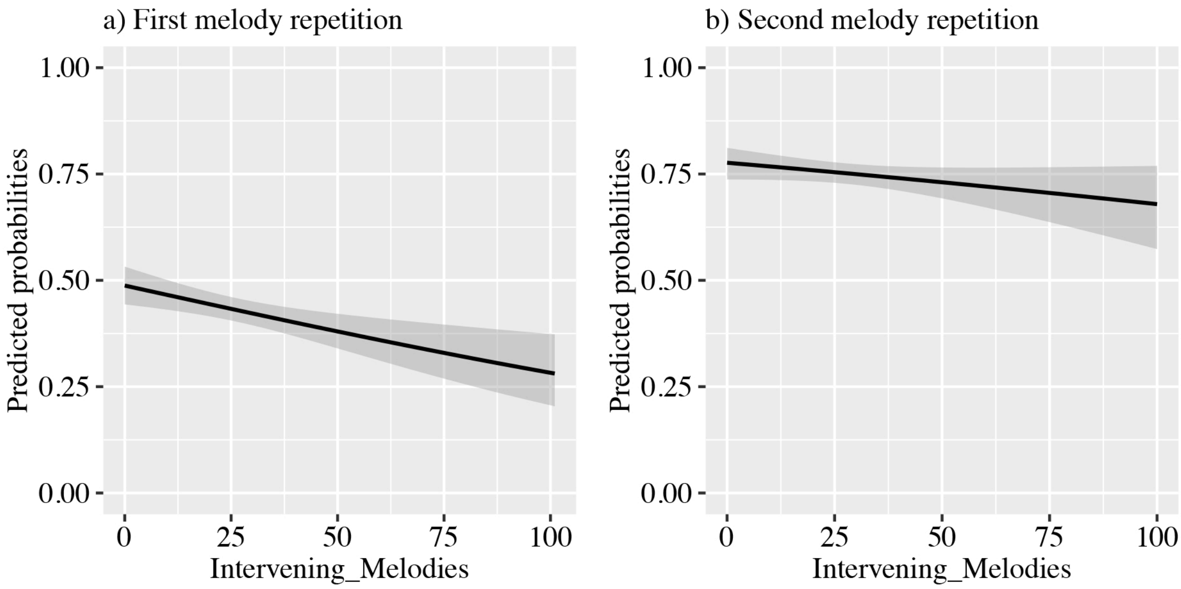


Figure 4.5: Prediction lines of generalized mixed effects models that model the bias corrected probability of recognition (y-axis) of pitch-only sequences. The left panel shows the effect of the number of intervening melodies between the first and second presentations of the melodies. The right panel shows the effect between the second and third presentations. A statistically significant disruptive effect of the number of intervening items on bias corrected recognition performance was only observed between the first and second presentation of the melodies. The grey area around the prediction line represents a 95% confidence interval.

Similar to Experiment 8, no significant disruptive effect on recognition performance with increasing *Number of Intervening Melodie*s was observed between the second and third presentations of the melodies (*coef* = -.005, $p > .05$), as depicted in *Figure 4.5*b.

4.4.3 Discussion

Experiment 9 aimed to further test the RMR conjecture by specifically investigating possible cumulative disruptive effects from intervening items on recognition of pitch-only sequences in an unfamiliar tuning system. In Experiment 8, a disruptive effect from the number of intervening items between the first and second presentation of the melodies was found when both the original melodic and rhythmic information was retained in each melody. A similar result was observed in Experiment 9, where the same melodies were used as in Experiment 8, but in pitch-only versions. This result shows that a disruptive effect on memory for melodies composed in an unfamiliar tuning system is still evident when rhythmic information is removed (which means less opportunity for multiple representations). This result was predicted by the RMR conjecture and shows that reducing the number of perceptual experiences does not reduce cumulative disruptive effects. Or put differently, less perceptible information of a stimulus (pitch and rhythm combined vs. pitch-only) does not reduce cumulative interference from intervening items.

Similar to Experiment 8, no disruptive effect from the number of intervening items was found between the second and third presentation of the melodies. Having established the effect of intervening items when melodies in an unfamiliar tuning system are presented without their original rhythmic information, we now turn to melodies presented without their original pitch information.

4.5 Experiment 10 – Recognition of Rhythm-Only Sequences

Experiment 10 investigates cumulative disruptive effects from the number of intervening items in rhythm-only sequences. A cumulative disruptive effect from the number of intervening rhythms on rhythm recognition is hypothesized between the first and second presentation.

4.5.1 Method

Participants. Thirty-six undergraduate students were recruited from the Murdoch University (M_{age} = 24.8 years, SD_{age} = 7.7). Average years of musical training was 1.2 (SD = 2.8). Participants did not previously participant in Experiments 1-9.

Stimuli. The same stimuli from Experiment 8 were used. However, the melodies were transformed into rhythm-only sequences by sounding all notes at 566 Hz (the average pitch of the entire set of melodies).

Procedure. The procedure was identical to Experiments 8 and 9.

4.5.2 Results

Figure 4.6 shows melody- and participant-wise performance. Overall, participants performed significantly above chance (Z = 12.12, p < .001) in recognizing rhythm-only sequences.

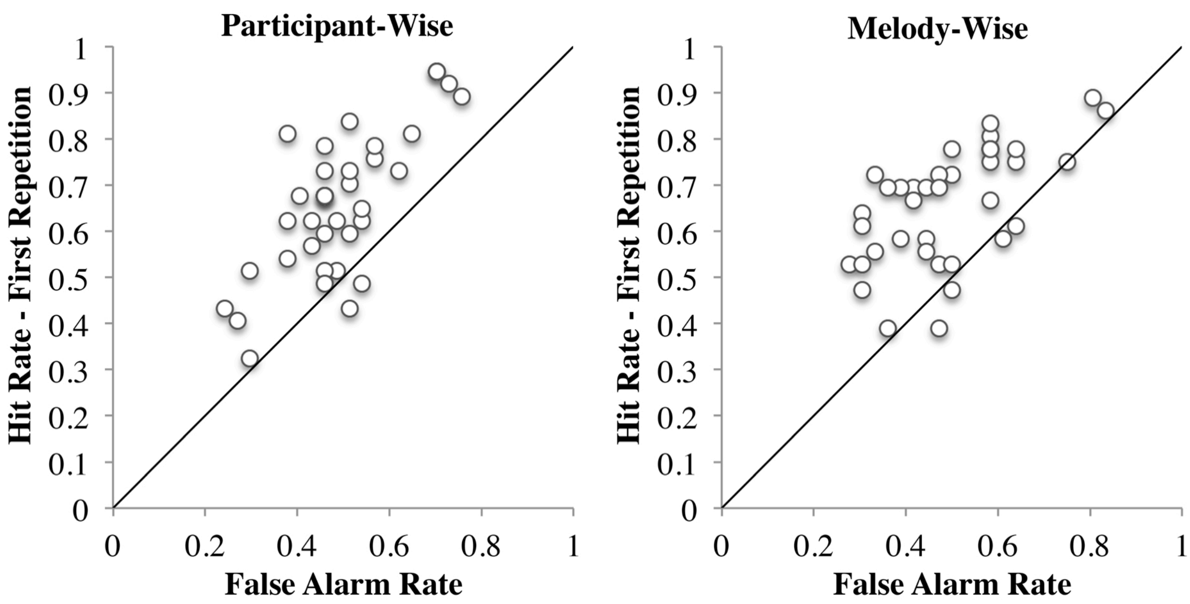


Figure 4.6: Hit rates and false alarm rates in response to rhythm-only sequences used in Experiment 10. The left panel shows the data participant-wise, and the right panel melody-wise. The reference line represents chance level. Overall, performance was significantly above chance (see text for more detail).

Consistent with Experiments 8 and 9, the number of intervening items between the first and second presentation of the rhythms had a disruptive effect on recognition performance. A model predicting 'old' responses on rhythm repetitions using a random intercept for *Participant*, *Melody*, and a systematic factor for *Dynamic Response Tendency* (*LogLik* = -811.02) improved significantly when provided with the *Number of Intervening Items* (*LogLik* = -807.98, p = .013). As with the previous two experiments, this result shows that intervening items cumulatively disrupt participants' rhythm recognition performance (*coef* = -.009, p = .013). *Figure 4.7*a shows the bias-corrected modelled probability of recognition as the number of intervening items increases between first and second presentation of a rhythm.

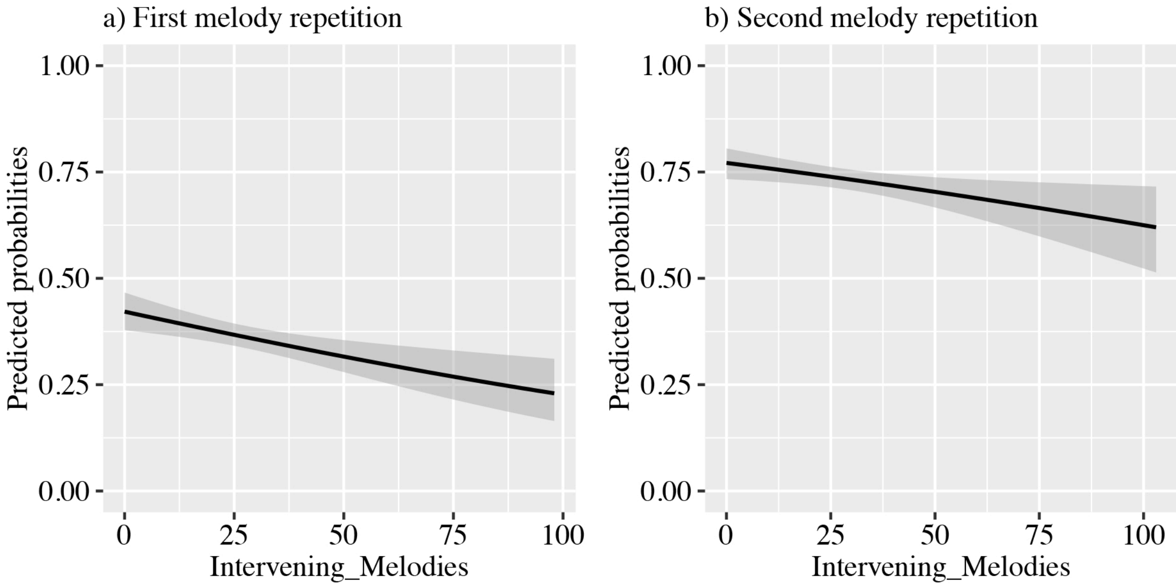


Figure 4.7: Prediction lines of generalized mixed effects models that model the bias corrected probability of recognition (y-axis) of rhythms. The left panel shows the effect of the number of intervening melodies between the first and second presentations of the melodies. The right panel shows the effect between the second and third presentations. Significant disruptive effects of the number of intervening items on bias corrected recognition performance were observed in both of these comparisons. The grey area around the prediction line represents a 95% confidence interval.

In contrast to Experiments 8 and 9, a disruptive effect on recognition performance with increasing number of intervening items was observed between the second and third presentations of the rhythms in Experiment 10 (*coef* = -.007, *p* = .037). A model predicting ‘old’ responses on third presentations of the rhythms (*LogLik* = -726.79) improved significantly when provided with the *Number of Intervening Items* between the second and third presentation (*LogLik* = -724.67, *p* = .040). *Figure 4.7*b shows the modelled probability of producing bias corrected recognition as the number of intervening items increases between second and third presentations of a rhythm.

4.5.3 Discussion

Experiment 10 aimed to provide a further test of the RMR conjecture by specifically looking at possible cumulative disruptive effects from intervening items on recognition of rhythm-only sequences. As hypothesized, a cumulative disruptive effect from the number of intervening rhythms on recognition was observed between the first and second presentation of the rhythms. Unlike Experiments 8 and 9, this effect was also observed between the second and third presentation of the rhythms. Observing such an effect between the second and third presentation using rhythm-only sequences suggests that memory representations of rhythms alone may be less resilient to interference from the number of intervening items than the combined rhythmic and melodic sequence.

4.6 General Discussion

Cumulative disruptive effects from the number of intervening items have been observed using a variety of stimuli (Bui et al., 2014; Donaldson & Murdock, 1968; Hockley, 1992; Konkle et al., 2010; Olson, 1969; Poon & Fozard, 1980; Rakover & Cahlon, 2001; Sadeh et al., 2014). Memory for melodies is also affected by these disruptive effects. However, this seems to be the case only when melodies are sounded in an unfamiliar tuning system (Herff et al., 2017; Herff, Olsen, et al., 2017). This finding has been previously predicted by a Regenerative Multiple Representations (RMR) conjecture (Herff et al., 2017). The RMR conjecture describes an important link between prior knowledge, perception, and subsequent formation of memories. The present study aimed to elucidate further the influence of prior knowledge and perception on recognition in the context of the RMR conjecture. This was achieved by first establishing baseline interference effects when melodies composed in an unfamiliar tuning system included all original pitch and rhythm information. We then took these stimuli and separated the original musical content into pitch-only sequences and rhythm-only sequences. The RMR conjecture predicts cumulative disruptive effects from the number of intervening items for melodies in an unfamiliar tuning system (Experiment 8), for their pitch-only sequences (Experiment 9), and for their rhythm-only sequences (Experiment 10).

The main prediction of the conjecture was supported by showing disruptive effects from the number of intervening items between the first and second presentation of stimuli presented in Experiment 8, 9, and 10. These results will be discussed and interpreted in the light of the RMR conjecture. Experiment 9 and 10 had the potential to falsify the RMR conjecture if no cumulative disruptive effects were observed. However, those experiments did show cumulative disruptive effects, thus providing preliminary support for the RMR conjecture.

4.6.1 Recognition of Melodies in Unfamiliar Tuning Systems

Experiment 8 partially replicated the results of the study in Herff, Olsen, et al. (2017). In that study, a robust and strong disruptive effect from the number of intervening melodies was found between the first and second, as well as second and third presentation of melodies in an unfamiliar tuning system. In Experiment 8, we observed the same patterns in the results, however only the number of intervening melodies between the first and second presentation of the melodies had a statistically significant effect on melody recognition. In the original work, the disruptive effect was weaker between the second and third presentations compared to the first and second; however, it was still significant. A likely explanation for this quantitative discrepancy is that the original study used nearly three times as many participants (N = 105) than Experiment 8 here (N = 37). This suggests the importance of large sample sizes in studies that aim to investigate cognitive processes measured over relatively large numbers of continuous conditions, such as the number of intervening items presented here. Nevertheless, the partial replication of the original study, namely the significant disruptive effect of the number of intervening melodies between the first and second presentation, served its original purpose here of providing a baseline to further test melody recognition and the RMR conjecture with pitch-only sequences (Experiment 9) and rhythm-only sequences (Experiment 10).

4.6.2 Memory for Pitch-Only Melodies in an Unfamiliar Tuning System

Experiment 9 used pitch-only versions of the melodies presented in Experiment 8. The stimuli were taken from the melodies of Experiment 8 but modified to comprise note durations and inter-note onsets that were identical between each note. The RMR conjecture predicts cumulative disruptive effects from the number intervening melodies for pitch-only sequences in instances where they were observed using the original rhythmical melodies. This is because the RMR conjecture first assumes that prior knowledge informs perception and perception influences formation of memory representations. If prior knowledge informs multiple ways of perceiving the same stimulus, then the conjecture suggests that the formation of multiple representations can support or regenerate each other if they code at least partially overlapping information. In the context of music, this means that prior experience informs perceptual relevance of notes, intervals, short musical phrases, as well as an integrated melody as a whole.

Our results support the RMR conjecture: melodies in an unfamiliar tuning system and with uniform rhythmic structure are not integrated as a whole, at least not to the extent that is required to recover from disruptive effects from intervening melodies. Specifically, the results in Experiment 8 showed a significant decrease in recognition performance with increasing number of intervening melodies between the first and second presentation of the melodies. Secondly, a cumulative disruption between the first and second melody presentations was observed in Experiment 9.

Therefore, we now have evidence together with that reported in Herff, Olsen, et al. (2017) that the use of melodies in an unfamiliar tuning system disrupts formation of a coherent representation of an integrated, musical melody. This might be because information on how to integrate notes, intervals, and short phrases into coherent melodies has not been acquired because of a lack of exposure to the unfamiliar tuning system. This finding suggests interesting follow up experiments. For example, future research could investigate cumulative disruptive effects in atonal melodies. These are melodies that use a familiar pitch set but unfamiliar arrangements of these pitches. Due to the unfamiliarity with the tonal-grammar, the RMR conjecture also predicts cumulative disruptive effects for atonal melodies, even if they incorporate a familiar pitch set.

4.6.3 Memory for Rhythm-Only Sequences

Experiment 10 used rhythm-only versions of the stimuli presented in Experiment 8. This means that the stimuli in Experiment 10 were identical to Experiment 8, but with the original pitch information removed. Similar to Experiment 9, results from Experiment 10 provide support for the RMR conjecture. As with the pitch-only sequences used in Experiment 9, the conjecture also predicts cumulative disruptive effects on recognition from the number of intervening items in instances where they were observed using the original melodies. This prediction was supported.

However, in contrast to Experiments 8 and 9, Experiment 10 also found a significant disruptive effect from the number of intervening items between the second and third presentation of the rhythms, and not just between their first and second presentation. The rhythm-only sequences in Experiment 10 provide less perceptible information than the original melodies with both rhythmic and melodic information in Experiment 8. This is also the case for the pitch-only sequences in Experiment 9. Therefore, the statistically significant disruptive effect on recognition of rhythm-only but not pitch-only sequences between the second and third presentation of the melodies can be explained by general performance differences between memory for rhythms and memory for melodic sequences. That is, memory for rhythms in general tends to be worse than memory for pitches (Hebert & Peretz, 1997; White, 1960). Nevertheless, the findings of Experiment 10 again provide further preliminary support for the RMR conjecture.

Memory for rhythm-only sequences has been thoroughly investigated using relatively short intervals of time (~ 10 sec) (Collier & Logan, 2000; Schaal, Banissy,

& Lange, 2015). The present study is the first empirical work investigating the effects from the number of intervening rhythms on recognition performance for rhythm-only sequences over relatively large numbers of intervening rhythms (up to 100). It is clear from the present study that memory for rhythm-only sequences is similar to other stimulus domains that do elicit cumulative disruptive effects from intervening items (Bui et al., 2014; Donaldson & Murdock, 1968; Hockley, 1992; Konkle et al., 2010; Olson, 1969; Poon & Fozard, 1980; Rakover & Cahlon, 2001; Sadeh et al., 2014), at least when the rhythmic structure is relatively unfamiliar to the listener.

4.7 Conclusion

Previous research has shown that melodies in a familiar tuning system show no systematic cumulative disruptive effects from the number of intervening items. However, melodies in an unfamiliar tuning system do show systematic disruptive effects, similar to those observed using many other non-musical stimuli. Here, we replicated the previous finding that melody recognition in an unfamiliar tuning system is possible with up to 100 intervening melodies, but is still susceptible to cumulative disruptive effects as the number of intervening items increases. We further extended this finding to rhythm-only sequences and pitch-only sequences in an unfamiliar tuning system. The overall pattern of results observed here support the predictions of a new and novel RMR conjecture (Herff et al., 2017).

An important next step for the development of the RMR conjecture into a complete theory is to test its applicability with stimuli outside of the musical domain. The original rationale of the conjecture was motivated by findings in the domain of music, language, and vision (Herff et al., 2017). So far, only music has been investigated from the perspective of the RMR conjecture. Nevertheless, the conjecture appears to be a useful tool in which to make precise predictions about the link between prior experience, perception, and formation of new memories.

Chapter 5

Memory as a Window into Music Perception

Chapter 5 is currently under peer review:

Herff, S.A., Dean, R.T., & Olsen, K.N. (submitted). Inter-rater agreement in memory for melody as a measure of listeners' similarity in music perception. *Psychomusicology: Music, Mind & Brain*.

Note: Kirk N. Olsen and Roger T. Dean are the author's supervisors.

5.1 Abstract

Music is a cultural universal, yet the individual experience of music can strongly differ between listeners. Here, we investigate the similarity of listeners' response patterns in the context of memory for melody and argue that memory can serve as a proxy to perception. If music perception is similar across listeners, then this similarity should be reflected in similar memory response patterns towards a specific melody corpus. We used inter-rater agreement in melody recognition tasks as a window into how 'similarly' listeners perceive music, and in particular, melodies. Specifically, inter-rater agreement of 10 melody recognition experiments was analysed and findings indicate inter-rater agreement of up to r ~ .70. However, inter-rater agreement was strongly dependent on whether direct recognition or indirect recognition in the form of perceived familiarity was measured, with direct recognition showing higher agreement among listeners. Furthermore, the specific melody corpus and tuning system played a significant role, as did whether melodies consisted of pitch-only, rhythm-only, or both pitch and rhythm information. Results are interpreted in light of their practical implications for computational models of memory for melody. We argue that these findings provide strong evidence that mathematical models designed to predict human memory for melody should focus on musical features that combine rather than separate components of rhythm and melody, and with greater emphasis on musical features that are independent of tuning system.

5.2 Inter-rater Agreement in Memory for Melody as a Measure of Listeners' Similarity in Music Perception

Music is an integral part of all cultures (Cross, 1999, 2001, 2014), yet differences in music traditions around the world are likely to lead to profoundly different ways in which listeners perceive music (Nettl, 1977). Even within one music tradition, listeners may perceive music differently; for example, due to differences in musical expertise (Dean, Bailes, & Dunsmuir, 2014; Herff & Czernochowski, in press). The tritone paradox presents an impressive example of how music perception can differ between listeners (D. Deutsch, 1991). Listeners are presented with alternating notes, half an octave apart. Some listeners perceive this as a descending pattern, whereas others perceive it as an ascending pattern.

However, previous research has shown that there is also considerable similarity in the ways listeners perceive music, even across cultures. A study by Balkwill and Thompson (1999) tested if Western listeners can identify the intended emotion in music from Hindustani raga excerpts. Along the emotional dimensions of Joy, Sadness, and Anger, Western listeners could correctly identify the intended emotion of these unfamiliar excerpts. Another study using similar labels of perceived emotion found that individual differences in music-perceived emotions can be linked to personality constructs such as alexithymia (Taruffi, Allen, Downing, & Heaton, 2017). Focusing on musical features in particular, Prince (2014) found that rhythm is one of the largest contributors to perceived similarity in melodies. In that study, listeners provided direct feedback on their perception of similarity using a rating scale. Other domains such as vision often use indirect measurements such as reaction times as a window into perception of observers (e.g., Wiley, Wilson, & Rapp, 2016). However, it is sometimes difficult to take meaningful discrete reaction time measurements in response to music, since music is continuous and unfolds through time. Here, we suggest an alternative method by using human memory as a window into listeners' perception of music.

Most theories of human memory assume that memory representations are based on perceptual experiences (Dennis & Humphreys, 2001; Hintzman, 1984, 1988; McClelland & Chappell, 1998; Paivio, 1969; Shiffrin & Steyvers, 1997; Tulving, 1972). For example, the first component in the Atkinson-Shiffrin memory model is a stimulus input that leads to the sensory register that detects and holds sensory information (i.e., a perception) (Shiffrin & Atkinson, 1969). Another example is MINERVA 2, which is based on a first experience or event (e.g., a perception) (Hintzman, 1984). A recent Regenerative Multiple Representations conjecture describes a crucial link between prior experience, perception, and the subsequent formation of memories (Herff et al., 2017; Herff, Olsen, et al., 2017; Herff et al., 2017). The conjecture suggests that differences in perceptual experience should translate to differences in memory representations. Consequently, similarity in listeners' memory response patterns to a specific set of melodies can be seen as a window into similarities between listeners' perception of that specific set of melodies. The present study was designed to further investigate the question of how similarly listeners perceive music by assessing inter-rater agreement in the context of memory for melody. In doing so, one of the primary aims was to quantify perceptual similarity between music listeners using a measurement that does not rely on emotional labels or semantic descriptions. Specifically, we addressed how inter-rater agreement varies: (1) between melody corpora within the same tuning system; (2) between melody corpora with different tuning systems; (3) between direct and indirect measurements of memory; and (4) between melodies that consist of pitch-only sequences, rhythm-only sequences, or both pitch and rhythm information combined. These manipulations are discussed in the following. As discussed later, we aim for the present results to provide practical information for computational models designed to predict memory for melody based on musical features.

5.2.1 Key Manipulations

Melody corpora with the same tuning system. Even within a given tuning and tonal system, there are still substantial differences in musical materials. A simple example of this would be the large variety of different musical genres that are popular and commonly heard in the Western music tradition. It is possible that the degree of similarity in listeners' perception varies between different music styles, even if the tuning system is identical. Here, we address this by analysing memory in response to two corpora of music that are both in the tuning system familiar to Western listeners, yet are distinctly different in their genre (see *Stimulus* section).

Melody corpora with different tuning systems. Testing melody corpora with different tuning systems is a compelling way of approximating the influence of musical enculturation on music perception and cognition (Stevens, 2012). This is because dissimilarities in tuning systems are readily apparent between musical cultures and most likely influence music perception. Potentially, the degree of similarity in music perception between listeners will vary depending on whether or not the melodies use a set of pitch and interval rules that are familiar to listeners. Indeed, music traditions come with underlying rules and expectations (D. Deutsch, 1986; Krumhansl, 1991, p. 295). The degree of similarity in listeners' perception of melodies may not only depend on the exact pitches used, but may also depend on whether or not listeners are familiar with these underlying rules and expectations of melodies.

Here, we analyse inter-rater agreement in participants' memory response patterns to melody corpora in three different tuning systems (see *Stimulus* section). Besides the tuning system familiar to Western listeners, we also use a novel tuning system that shares many of the underlying rules with the familiar tuning system; however, it uses a different pitch set. The third tuning system is a novel tuning system that uses unfamiliar rules as well as an unfamiliar pitch set.

Direct and indirect measurements of memory. Explicit information about the nature of a memory task can lead to different performances when compared to indirect memory tasks (Fleischman et al., 2004; Gaudreau & Peretz, 1999; Halpern & O'Connor, 2000). This is also observed in musical memory tasks (Halpern & Bartlett, 2010). Here, we analyse inter-rater agreement in direct *and* indirect memory tasks. In a direct task, participants are aware that their memory is being tested. In an indirect memory task, participants are not directly informed about the nature of the memory task. Considering that indirect memory tasks entail more uncertainty for participants than direct task instructions that are often implemented to effectively homogenise listeners, we predict that direct memory tasks show higher inter-rater agreements than indirect memory tasks.

Pitch-only sequences, rhythm-only sequences, or both combined. Melodies usually consist of a pitch sequence combined with a rhythmic sequence. Generally, it is easier to recognise a melody when hearing its pitch-only sequence rather than its rhythm-only sequence (Hebert & Peretz, 1997; Herff et al., 2017; White, 1960). However, recognition performance of a melody is best when the original combined pitch and rhythmic sequence is presented (Hebert & Peretz, 1997; Herff et al., 2017), yet it is important to note that recognition performance and inter-rater agreement are not necessarily correlated. Indeed, it is possible that multiple listeners produce a similar memory-response profile towards a set of melodies, yet show low actual recognition performance. The present analysis compares inter-rater agreement in memory tasks for melodies that consist of combined pitch and rhythmic sequences with inter-rater agreement in memory tasks that test pitch-only sequences and rhythm-only sequences.[9]

5.2.2 Predictive Models of Memory

In the visual domain, 'memorability' of a picture appears to be a stable property across human observers. This means that pictures that are more memorable for one person are also most likely more memorable for another person (Isola et al., 2011). In other words, observers show high inter-rater agreement in their memory response patterns towards a set of pictures. In Isola et al. (2011), participants were split into two groups and the response pattern from one half were correlated with those from the other half (r = .75). The correlation coefficient can be squared to

[9] Note that the present investigation is concerned with similarity in listeners' responses (for more detail on recognition performance for pitch-only and rhythm-only sequences see (Herff et al., 2017).

obtain the proportion of variance in the response pattern from one half of participants that can be explained by the response pattern from the other half (J. Cohen, 1988).

Recently, Flexer and Grill (2016) measured listeners' inter-rater agreement in a music similarity task in an attempt to validate computational models of music similarity (based solely on acoustic features). Similar to Isola et al. (2011), they also correlated the response patterns from one half of participants with the other half (see p. 244). They found average split-half correlations of r = .40. This means that ~16% of the variance in the pattern of average responses from one half of participants can be explained by the response pattern of the other half. The authors argue that this result has implications for computational models of music perception that predict average similarity responses to musical material. Specifically, as the inter-rater agreement decreases, so does the possible performance of a predictive model. In other words, the higher the inter-rater agreement between listeners for a category of perceptual response, the more precise a predictive model of that response could potentially be.

Predicting Memory for Melodies. Recognition of monosyllabic words can be well predicted (up to 45% of the variance in hit rates) using underlying features such as word frequency (Cortese, Khanna, & Hacker, 2010). In music, underlying melodic features can relate to popularity (Kopiez & Müllensiefen, 2011) and a recent study using a blocked indirect recognition design explained between 9.6% and 25.3% of the variance in the recognition responses using musical features (Müllensiefen & Halpern, 2014). Müllensiefen and Halpern (2014) developed computational models that predict average recognition performance of melodies based on musical features of the stimuli. Music is a complex stimulus that comprises many underlying musical features. These features can be as simple as average pitch height or as sophisticated as rhythmic complexity values (Müllensiefen & Halpern, 2014). Understanding which of these features carry predictive value can shed light on which features listeners may use to selectively base recognition judgements, both consciously or non-consciously. In the following we will discuss the implications of the present work for computational models that, similar to Müllensiefen & Halpern, 2014, aim to predict average memory performance for melodies based on musical features of the melodies.

One problem with models that use stimulus features to predict memory responses is that there are non-stimulus related processes involved in memory. For example, variables such as decay over time, emotional associations, attention lapses, lack of motivation, fatigue, expertise, or repetition all influence memory for melody (Cuddy et al., 2012; Gardiner, Kaminska, Dixon, & Java, 1996; Herff & Czernochowski, submitted; McAuley et al., 2004; Samson, Dellacherie, & Platel, 2009). As a result, it is not always clear how much variance a model using stimulus features could potentially explain, and how much variance is due to non-stimulus related processes.

The present investigation aims to shed light on the average proportion of variance that a pattern of responses from one group of participants can explain the pattern of responses in another group. In general, the higher the proportion of explained variance between participant groups, the more promising the predictive model that is based on stimulus features. This is because the similarity in the average response pattern between participants towards a set of melodies can largely be attributed to stimulus features, or some other constant feature of the experimental condition (e.g., this could be the environment, distractors, social circumstances experienced by participants, etc.), whereases dissimilarities in the response patterns can largely be attributed to non-stimulus related inter-individual differences.

5.2.3 Methods of Testing Memory

In their investigation of memory for melody Müllensiefen and Halpern (2014) used a blocked recognition design. In a blocked design, participants first hear a large number of melodies in a learning phase and then, in a test phase, hear new melodies mixed with old melodies. Participants then have to decide which melodies have been previously presented. This paradigm is a useful tool for memory research as it provides a clear distinction between encoding and retrieval. However, everyday music recognition does not usually provide a clear separation between encoding and retrieval. Rather, for every stimulus encountered, the same question of whether this stimulus has been heard before is assessed without specific focus on encoding or retrieval. A useful alternative to a blocked design is a continuous recognition paradigm (Shepard & Teghtsoonian, 1961). In a continuous recognition paradigm, participants are presented with one melody after another throughout the experiment and judge if each melody has previously been presented in the experiment. Sometimes, a melody is presented that has indeed been presented before. This paradigm has the advantage that is does not provide participants with information about whether they have to focus on encoding or retrieval (Dowling, 1991).

In a series of studies using continuous recognition paradigms, Herff and colleagues (Herff et al., 2017; Herff, Olsen, et al., 2017; Herff et al., 2017) investigated context variables that might influence melody recognition and might blur the predictive power of musical features. Surprisingly, with up to 195 intervening melodies, these studies demonstrated that the number of intervening melodies has effectively no disruptive impact on melody recognition performance (Herff et al., 2017). Similar results in the context of temporal decay rather than interference have been obtained, showing that temporal delay of up to a week has minimal to no disruptive effect on melody recognition (Schellenberg & Habashi, 2015).

These findings provide additional motivation for the investigation of the predictive power of musical features. This is because decay and interference are traditionally two of the major influences on memory (Norman, 2013; Oberauer & Lewandowsky, 2011; Oberauer et al., 2012). With the impact of these two variables shown to be minimal in memory for melodies, the proportion of variance explainable by musical features might be larger than using other stimuli where these two effects are important (e.g., words, sentences, prose, pictures, faces, numbers etc.).

5.3 Experiments

5.3.1 Aim and Rationale

This paper is an analysis of the combined data obtained by Herff and colleagues' investigations of memory for melody originally published or submitted in (Herff et al., 2017; Herff, Olsen, et al., 2017; Herff et al., 2017). The aim of the present analysis is twofold. Firstly, we aim to shed light on the question 'how similar is music perception between listeners?' We use memory as a proxy to address this question. We assume that if multiple listeners' perception of music is similar, then response patterns (in this case, memory) to particular melody corpora will also be similar. Secondly, we aim to investigate how the average proportion of variance that a pattern of responses from one group of participants can explain the pattern of responses in another group (in other words, inter-rater agreement). The overall goal is that the findings will inform future computational models that aim to use musical features to predict memory for melody.

5.3.2 Method

Participants. All participants were recruited from the Western Sydney University (Experiments 1, 2, 3, 4, 5, 6) or the Murdoch University (Experiments 7, 8, 9, 10), Australia. The first six experiments exclusively recruited participants with fewer than 2 years of formal musical training and who were not actively participating in any form of music. Experiments 7, 8, 9, 10 recruited participants with mixed musical background, predominately consisting of non-musicians. The vast majority of participants comprised undergraduate university students enrolled in psychology courses.

Procedure. All studies analysed here used the same basic continuous recognition paradigm. After each melody was presented, participants were required to make a response before the next melody started. In eight of the experiments, participants were informed that after each melody they should judge whether they have heard this melody in this experiment before by pressing a button labelled ‘old’, or whether this is the first time they have heard this melody, by pressing a button labelled ‘new’. In two experiments, instead of direct memory task instructions, participants were only instructed to indicate perceived familiarity on a 100-point visual analogue scale. The increase in perceived familiarity between first and second presentation of the melodies provides an indirect approach to measure recognition performance. In six experiments, every melody was presented twice throughout the experiment (i.e., one repetition). In four experiments, every melody occurred three times (i.e., two repetitions). In these cases, results are reported separately for both melody repetitions.

Stimuli. A summary of the stimuli characteristics is provided in Table 1. The stimuli of all Experiments can be found in the online supplement S1-Stimuli.zip.

Table 5.1: A Summary of Stimulus Characteristics

Exp	N_{Melodies}	Duration	Sound	Tuning system	Stimulus Notes
1	60	12 sec	Piano	12-TET, Western tonal	New melodies that resemble advertisement jingles.
2	55	12 sec	Piano	12-TET, Western tonal	From Exp .1
3	98	10.86 sec	Piano	12-TET, Western tonal	European folk songs,
4	98	10.86 sec	Piano	12-TET, Western tonal	Same as Exp. 3
5	50	10.86 sec	Piano	Novel 88.08-CET	Unfamiliar in pitch, but based on Exp. 3 and 4

6	50	10.86 sec	Piano	Novel 88.08-CET	Same as Exp 5
7	37	2.695 sec	Pure tones	Novel artificial tuning	Unfamiliar in pitch and rhythm
8	37	2.695 sec	Pure tones	Novel artificial tuning	Same as Exp. 7
9	37	2.695 sec	Pure tones	Novel artificial tuning	Pitch, based on Exp. 7, 8
10	37	2.695 sec	Pure tones	Novel artificial tuning	Rhythm, based on Exp. 7, 8

Note. From left to right, Experiment index, number of different melodies used, average melody duration, sound the melodies were realised in, the tuning system used, and additional notes about the stimuli.12-TET refers to the 12-tone-equal-temperament tuning system, the tuning system most dominant in Western cultures. 88 CET refers to a novel, artificial tuning system that uses a 88 cent step equal temperament, rather than the 100 cent in 12-TET. The novel tuning system of Experiments 7-10 is detailed in the text.

Two experiments used melody corpora that resembled modern advertisement jingles, and another two corpora used European folks songs (Herff et al., 2017). This means that a total of four experiments used stimuli in 12-tone-equal-temperament (12-TET), the tuning system familiar to Western listeners. The other six experiments used melodies in unfamiliar tuning systems. Some of the melodies in an unfamiliar tuning system were based on the melodies in a familiar tuning system. These experiments realised the melodies in a new 88-Cent-Equal-Termparment tuning system. The 88-CET tuning system is the equally tempered tuning system most dissimilar to the familiar Western tonal system within all 40 to 100-CET tuning system systems (Herff, Olsen, et al., 2017) based on the tonal affinity model of Milne et al. (Milne, 2013; Milne & Holland, 2016; Milne, Laney, et al., 2015; Milne et al., 2016; Milne et al., 2011). This means that adjacent pitches in the tuning system are 88 cents apart, rather than the 100 in the Western tonal tuning system. As a result, this tuning system uses a different pitch-set, however, the melodies have the same contour and rhythm than familiar melodies; only the pitch of each note is adjusted to the new tuning system.

The melodies in Experiment 7 to 10 did not conform to Western music tradition in rhythm and tonality (Herff, Olsen, et al., 2017; Herff et al., 2017). The melodies used pitch heights of 480, 520, 560, 605, and 665Hz and note durations of 60, 110, 550, and 920 ms, with a 100 ms silent gap between notes.[10] Some experiments presented a given set of stimuli either in pitch-only, rhythm-only, or combined versions with both pitch and rhythm information. This allows the present analysis to compare the contribution of rhythm and pitch sequences to the recognition and perceived familiarity response similarity between participants.

Statistical approach. We assess inter-rater agreement (or similarity) in recognition by samples of participants towards sets of melodies. To this end, in each experiment the average recognition performance for each melody of half the

---

[10] In pitch perception and a standard/comparison task with a silent retention interval, a Weber fraction of .04 led to discrimination performance above 90%. The pitch Weber fractions used in the present stimuli were between .08 and .10, therefore clearly discriminable. In duration perception, Weber fractions of .30 led to discrimination performance above 98%. The duration fractions used in this study ranged from .67 to 1.27, therefore also clearly discriminable. The melodies were constructed in various artificial grammars, unfamiliar to listeners, which were of importance for the later investigation of statistical learning that will be reported elsewhere.

participant sample was correlated with the average recognition performance of the other half towards the same melodies. Specifically, the participant sample was split into two-halves and the average hit-rate for each melody was calculated. These average hit-rates for each melody were then correlated with the average hit-rates for the same melodies by the other half of the sample. To increase estimation precision of the average split-half correlation, we repeated this process 1000 times with random splits and average split-half correlations with *p*-values and standard deviations reported for each experiment. The *p*-values and Cohen's d (J. Cohen, 1988) were obtained by comparing the vector of actual split-half correlations with a vector of split-half correlations in which melody names were shuffled for one of the halves. Note that due to the large number of (split-half) observations (1000), most effects will show significance. Therefore, we use correlation values as interpretable effect sizes and report Cohen's d as a measure of the effect size in standard deviations (J. Cohen, 1988).

Furthermore, the proportion of variance that the response pattern of one group of participants predicts from the response pattern of another can inform predictive models of memory that use stimulus features as predictors. Specifically, the higher the proportion of explainable variance, the more promising the predictive model will be (Flexer & Grill, 2016). Here, we use average split-half r^2 as an indicator of the average proportion of variance that the response pattern of one group of participants predicts in the response pattern of another (J. Cohen, 1988; Flexer & Grill, 2016). Average split-half r^2 values were calculated by squaring each of the 1000 split-half correlations and then averaging them (the result is thus normally slightly different from simply squaring the r-values in *Table 5.2*).

For each of the 10 experiments, ANOVAs were conducted on the average split-half correlations of to investigate potential influences of the factors *Task Instructions* (direct recognition vs. indirect recognition in the form of perceived familiarity), *Tuning System* (familiar vs. unfamiliar), *Repetition* (first vs. second), and *Sequence Type* (pitch-only vs. rhythm-only vs. both pitch and rhythm combined) on inter-rater reliability in the form of average split-half correlations. As noted above, due to the large number of split-half correlations, most effects will be found significant and effect sizes should therefore be predominantly used for data interpretation. Similarly, figures will depict Standard Deviations instead of Standard Errors and confidence intervals. This is because with the large number of split-half correlations, Standard Errors become too small to be informative in visual representations of the data.

5.3.3 Results

All Experiments yield significant average split-half correlations of melody specific recognition with 1000 splits (all $p < .0001$). A summary of the results can be found in Table 2. The table reports average split-half correlations for all 10 experiments. Note that Experiments 5 and 6 used the same melodies previously used in Experiments 3 and 4, however retuned to an unfamiliar tuning system. Experiments 4 and 6 used the identical stimuli compared to Experiments 3 and 5, respectively, but in an indirect task where participants were instructed to report perceived familiarity, rather than direct recognition. Figure 1 shows the split-half correlation distributions for each experiment. A wide distribution shows large differences between the 1000 split-halves, whereas a narrow distribution represents similar split-half correlations in all 1000 split-halves. Distributions on the far left suggest overall low split-half correlations, whereas distributions to the far right suggest high split-half correlations.

Table 5.2: Summary of the Results

Exp	**N**	**Stimuli**	**Design**	**r**	**SD**	**d**	**r^2**
1	28	60, familiar tuning	Recognition	.631	.065	5.787	.409
2	20	55, from Exp .1	Recognition	.561	.071	4.977	.3161
3	32	98, European folk songs, familiar tuning	Recognition	.322	.067	3.665	.106
4	30	98, same as Exp. 3	Familiarity	.135	.076	1.506	.023
5	37	50, unfamiliar tuning, based on Exp. 3 and 4	Recognition	.340	.093	2.829	.134
6	27	50, same as Exp 5	Familiarity	– .056	.104	.391	.015
7	105	37, unfamiliar tuning	Recognition	1st .710	.063	5.372	1st .510
				2nd .594	.081	4.630	2nd .360
8	36	37, same as Exp. 7	Recognition	1st .613	.079	4.770	1st .384
				2nd .539	.087	4.102	2nd .294
9	34	37, pitch-only, based on Exp. 7 and 8	Recognition	1st .146	.121	1.024	1st .037
				2nd .180	.120	1.192	2nd .049
10	36	37, rhythm-only, based on Exp. 7 and 8	Recognition	1st .503	.094	3.673	1st .258
				2nd .499	.097	3.512	2nd .256

Note. The table depicts from left to right, Experiment index, number of participants, a short description of the stimuli, whether participants were instructed explicitly to make a recognition judgement or indirectly by reporting perceived familiarity, average split-half correlations, standard deviation of the average split-half correlations, Cohen's d as effect size (standardised distance to the correlations observed when melody names are shuffles in one half of the participants), and the average of all squared split-half correlations. Note that Experiments 7, 8, 9, and 10 repeated melodies twice throughout the experiments. In these cases, results for both repetitions are reported separately.

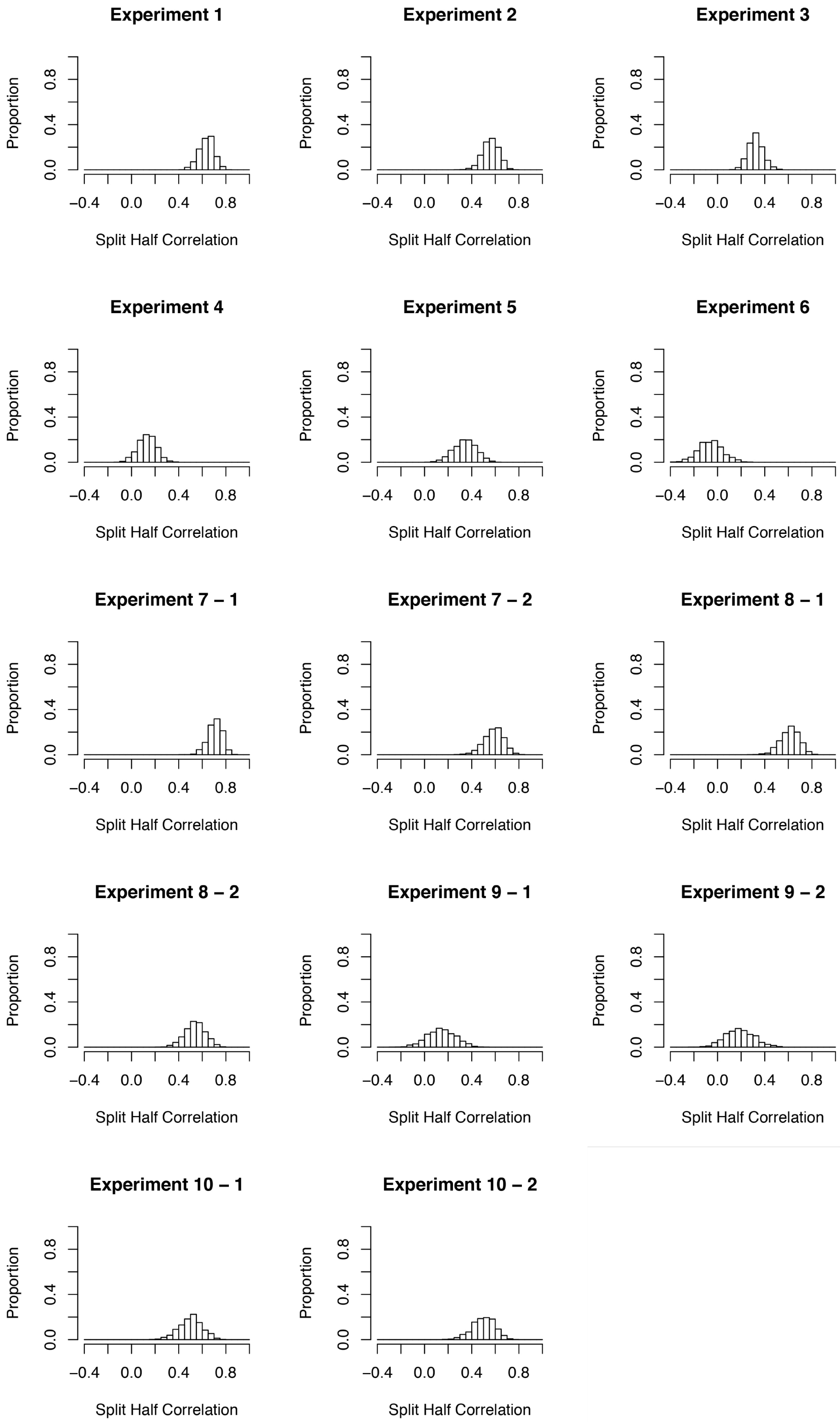

Experiment 1
Experiment 2
Experiment 3
Experiment 4
Experiment 5
Experiment 6
Experiment 7 – 1
Experiment 7 – 2
Experiment 8 – 1
Experiment 8 – 2
Experiment 9 – 1
Experiment 9 – 2
Experiment 10 – 1
Experiment 10 – 2
Proportion
0.0 0.4 0.8
−0.4 0.0 0.4 0.8
Split Half Correlation

Figure 5.1: Distributions of split-half correlations in each Experiment. For Experiments 7-10, the dashed numbers '1' and '2' shown after the title of each experiment index to the first and second repetition of the melodies.

Task Instructions and Tuning Systems. An ANOVA analysed the split-half correlations of Experiments 3, 4, 5, and 6 revealing a significant main effect for *Task instructions* (direct recognition vs. indirect familiarity), $F(1,3996) = 11375.75$, $p < .0001$, $\eta_p^2 = .740$, *Tuning system,* $F(1, 3996) = 1003.98$, $p < .0001$, $\eta_p^2 = .201$, as well as their significant interaction, $F(1, 3996) = 1471.02$, $p < .0001$, $\eta_p^2 = .269$. Figure 2 depicts means and standard deviations of the average split-half correlations and visualises the two main findings: (1) the indirect familiarity tasks show lower split-half correlations compared to the recognition tasks; and (2) the unfamiliar tuning system showed slightly higher split-half correlations compared to the familiar tuning system. Furthermore, Figure 2 depicts the interaction between *Tuning system* and *Task instructions*, showing that the familiar tuning system produces higher inter-rater agreement in the indirect perceived familiarity task compared to the familiar tuning system, but this is not observed when a direct recognition task is used.

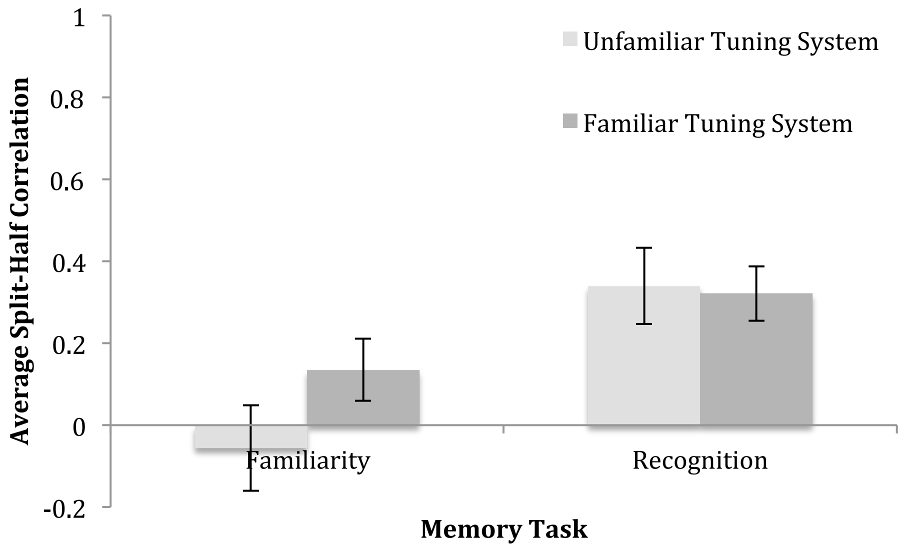


Figure 5.2: Average split-half correlations for the indirect perceived familiarity tasks, the direct recognition task and both unfamiliar and familiar tuning systems. Data of Experiments 3, 4, 5 and 6. Error bars show standard deviation because standard errors (and confidence intervals) are too small to depict due to the large number of split-half correlations.

Repetition, Pitch-only, Rhythm-only, and Combined Sequences.

Experiments 7 and 8 used melodies in an unfamiliar tuning system presented in a direct recognition task. Experiments 9 and 10 used the same stimuli that Experiments 7 and 8 used, however, Experiment 9 used the pitch-only versions and Experiment 10 used the rhythm-only versions of the original stimuli. An ANOVA analysed the data of Experiments 7, 8, 9, and 10 showing significant effects of *Stimulus type,* $F(2, 7994) = 14105.06$, $p < .0001$, $\eta_p^2 = .779$, *Repetition* ($F(1, 7994) = 88.88$, $p < .0001$, $\eta_p^2 = .011$, and their significant interaction, $F(2, 7994) = 328.88$, $p < .0001$, $\eta_p^2 = .076$. The small effect sizes observed for the main effect of *Repetition* and the interaction make the significant *Stimulus type* factor the most relevant finding from this analysis. Figure 3 shows average split-half correlations in regards to the different types of stimuli after the first and second repetition. Overall, rhythm-only sequences show higher average split-half correlations (M_r = .501, SD_r = .096) than the pitch-only sequences (M_r = .163, SD_r = .122). The combined melodies show more explainable variance (M_r = .614, SD_r = .099) than did their underlying pitch and rhythmic sequences separately.

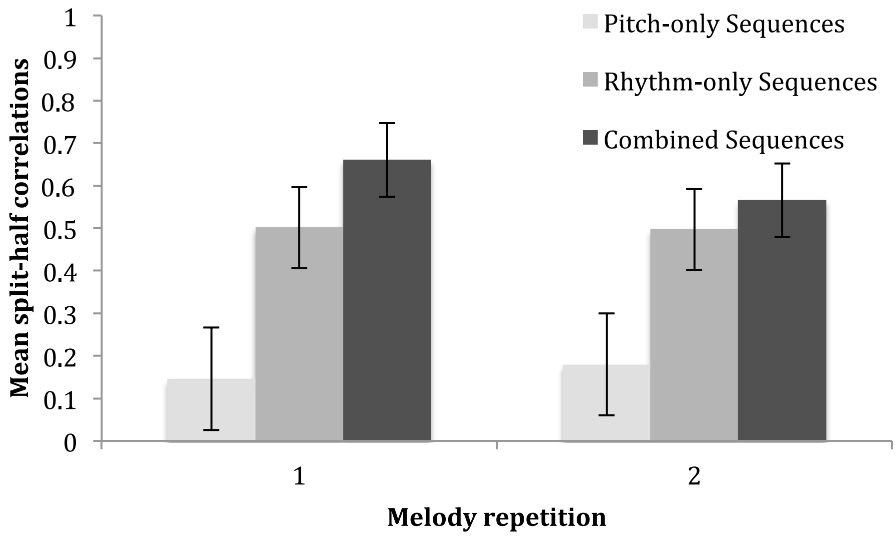


Figure 5.3: Average split-half correlations for both melody repetitions and all three kinds of stimuli; the original melodies in an unfamiliar tuning system that consisted of a combined pitch and rhythmic sequence, as well as underlying pitch and rhythmic sequences tested separately in a new sample. Data of Experiments 7, 8, 9, and 10. Error bars show standard deviation because standard errors (and confidence intervals) are too small to depict due to the large number of split-half correlations.

5.3.4 Discussion

The present study analysed the memory response patterns of listeners to various melody corpora from 10 different experiments. We calculated the inter-rater agreement of memory recognition between participants. Assuming that listeners' similarity in memory response patterns is primarily driven by perceptual similarities between listeners, we investigated the extent to which listeners' music perception converges. We observed some striking similarities between listeners with inter-rater agreement of up to r = 71. The highest average inter-rater agreement we observed is similar to those in other domains (e.g., monosyllabic recognition, r = .70 in Cortese et al. (2010), and picture recognition, r = .75 in Isola et al. (2011). However, average inter-rater agreement significantly varied based on memory task instructions, tuning system, and the nature of the stimuli (pitch-only, rhythm-only, both combined).

In light of these main findings, the degree of listeners' similarity in music perception as well as the implications for predictive models of melody recognition will now be discussed in the context of the influence of melody corpora, the influence of memory task instructions, the influence of tuning system, and the relative contribution of rhythm and pitch information to stability of memorability of melodies.

The Influence of Melody Corpora. Within the Western-tonal experiments analysed in the present study, there was a large similarity between listeners' recognition response patterns. However, the degree of agreement varied between corpora, even though these corpora used melodies in the same tuning system. A new corpus resembling modern advertisement or movie themes (Experiments 1 and 2) showed higher inter-rater agreement than a corpus of European folk songs (Experiment 3). This suggests that similarity in listeners' perception of music changes as a function of the precise auditory material. Depending on the melody corpus, listeners are therefore likely to be very similar in the way they perceive the music. More precisely, it suggests that similarity in music perception decreases and memory response patterns diverge as the musical style of a melody corpus becomes less familiar. This is because music from advertisements and movie themes are likely to be more familiar or more easily remembered for Australian listeners than European folk songs.

The similarity between participants' recognition judgements also informs computational models that aim to predict average memory responses using musical features as predictors (similar to Müllensiefen & Halpern). By squaring the split-half correlation coefficients and then averaging them, we calculated the proportion of variance that the response pattern from one group of participants explains of the response pattern from another group (J. Cohen, 1988; Flexer & Grill, 2016). In Experiment 1, the present results show that an average proportion of up to ~40% of the variance can be explained by the response pattern of another group of participants. This suggests that the endeavour to develop musical feature models that predict recognition patterns to a substantial degree is a feasible one, as a large proportion of the variance seems to be stimulus driven. This assertion is based on the rationale that a large proportion of the unexplainable variance between two groups of participants that perform the same task is based on inter-individual differences, whereas a large proportion of the variance that can be explained is based on stimulus features. In a practical context, we argue that higher the proportion of explained variance between two participant groups, the more promising the predictive model will be when memory is predicted by stimulus features within the music.

The wide range of variance that can be explained (10% to 40% in Experiments 1,2, and 3) for responses within the Western melody corpora, also shows that the potential usefulness of predictive feature models depends on the exact melody corpus. As the corpora that approximated pop music or advertisement melodies elicited greater inter-rater agreement in memory responses when compared to traditional European Folk melodies, the present results point towards future questions. Specifically, future research could address the question of the influence of genre familiarity on similarity in listeners' perception. However, such an endeavour should be wary of differences in sample expertise. Groups with high levels of musical expertise may have a different response profile. In general, it is to be expected that the more homogenous a participant sample is in regards to musical expertise, the more similar their responses will be. As a result, this would increase the proportion of explainable variance, but simultaneously make predictive models less generalizable. In terms of the generalizability of a predictive model, it is also important to consider the precise memory task utilised.

The Influence of Memory Task. The present investigation analysed data of multiple experiments that deployed identical melody corpora but different memory tasks. All experiments analysed here used a continuous recognition paradigm, but in some experiments participants were instructed that some melodies would be repeated throughout the experiment, and therefore they were required to report which melodies had been presented before (direct memory task). In other experiments, participants were instructed to just rate their perceived familiarity towards the melodies (indirect memory task). Even though sometimes the same overall memory phenomena can be demonstrated in both the indirect and direct memory tasks (Herff et al., 2017; Herff, Olsen, et al., 2017), it is clear from the present findings that the different tasks elicit significantly different inter-rater agreement in response to the same melody corpora. Memory responses were more similar between participants in the recognition task than the perceived familiarity tasks, even though the same melodies were used. This finding is not surprising considering that indirect paradigms may introduce additional participant-wise variation that comes with the uncertainty of task instructions; nevertheless, it is worth noting for future experiments and analyses.

The substantially smaller inter-rater agreement reported here between groups of participants in the indirect memory task also indicates less potentially explainable variance for musical feature models designed to predict perceived familiarity. This suggests that modelling recognition instead of perceived familiarity may be a more effective endeavour when aiming to explain large proportions of variance. These results also suggest that future investigations utilising indirect measures of memory (such as increases in perceived familiarity between melody occurrences) to build predictive models of melody recognition might not be as fruitful as they intuitively seem. Interestingly, the effect size from comparing differences in the average split-half correlations between first and second melody repetitions (Experiments 7-10) was negligibly small ($\eta_p^2 = .011$). This means that the degree of inter-rater agreement (not the performance) is hardly affected by the additional melody repetition, suggesting that a small number of additional melody repetitions do not increase similarity between listeners' perception of a piece. More empirical attention, however, will be useful to evaluate models that utilise tuning system-independent musical features as the basis of their memory response predictions.

The Influence of Tuning System. The present investigation comprises experiments that presented melodies in the tuning system familiar to participants and tuning systems that were unfamiliar to participants. Importantly, Experiments 4 and 5 used stimuli directly based on Experiment 3, only detuned into an unfamiliar tuning system. As a result, many musical features such as rhythm and pitch contour in the stimulus set were identical between these experiments. Interestingly, inter-rater agreement was comparable between these two corpora. In fact, the agreement was slightly higher for the unfamiliar tuning. This difference is significant but comparably small ($\eta_p^2 = .201$). Consequently, we interpret the results conservatively and conclude that we did not find meaningfully greater inter-rater agreement for the unfamiliar tuning system compared to the familiar tuning system. This result suggests that the degree of similarity in listeners' music perception does not change dramatically with tuning system per se, but rather, with the musical features that are unaffected by the tuning system.[11] Consequently, it may be that a large proportion of the explainable variance in memory for melodies might be due to musical features that are independent of tuning system (such as melody contour). This observation has direct implications for future models aiming to predict melody recognition, as musical features that evoke similar perceptual responses in listeners' may also carry large proportions of predictive power. Further support that underlying musical

[11] We found further support for this interpretation in a post-hoc analysis. We used the same average correlation procedure to correlate the response patterns in Experiment 3 (Western tonal) with Experiment 5 (melodies based on Experiment 3 but played in an unfamiliar tuning system). We found similar average-split-half correlations between Experiments 3 and 5 ($r = .234$, $SD = .118$), relative to those within Experiment 3 ($r = .32$, $SD = .067$) and within Experiment 5 ($r = .34$, $SD = .09$).

features might predict melody recognition performance, even outside the domain of familiar tuning systems, can be observed from the data of Experiments 7 and 8. These experiments used a different unfamiliar tuning system and show proportions of explainable variance of ~37%. Therefore, the present findings suggest that the intrinsic predictive power of stimuli for memory for melody might not derive from familiarity with the underlying tuning system, but instead from musical features that operate independent of tuning system.

The Influence of Rhythm and Pitch Sequence. The unique composition of stimuli used in Experiments 7, 8, 9, and 10 allow some additional conclusions. Experiments 9 and 10 used the same stimuli as Experiments 7 and 8. However, Experiment 9 provided participants solely with pitch-only versions of the stimuli and Experiment 10 used rhythm-only versions of the original stimuli in Experiments 7 and 8. This means that Experiments 7 and 8 provide a baseline in which to compare responses to combined rhythm and pitch sequences with pitch (Experiment 9) or rhythm (Experiment 10) sequences separately. Interestingly, the degree of similarity in listeners' response pattern was higher in the rhythm-only sequences compared to the pitch-only sequences. Furthermore, the degree of similarity was much higher in the combined sequences than the results from the two individual sequences would suggest.

More specifically, the proportion of explainable variance in melody recognition performance in the pitch-only sequences tested here is around 4%. The rhythm-only sequences, on the other hand, show around 26% explainable variance with significantly higher inter-rater agreement between listeners. This is important from at least two perspectives. Firstly, pitch-only sequences are often reported to be more memorable than rhythm-only sequences (Hebert & Peretz, 1997; White, 1960). The present analysis does not challenge these findings, but rather suggests that participants show higher inter-rater agreement in their recognition judgements towards pure rhythmic sequences compared to pure pitch sequences. The present findings are in line with reports that rhythms are one of the main contributors to perceived similarity in melodies (Prince, 2014). Secondly, rhythm-only sequences show ~26% and pitch-only sequences ~4% explainable variance in the memory response patterns. However, combining the same rhythmic and pitch sequences into coherent melodies results in more than 38% of the variance potentially explained.

These results suggest that there are remarkable interactions between rhythm and pitch for listeners' similarity in perception. A candidate-mechanism for these interactions could be the bilateral guidance of attention and phrase perception in rhythm and pitch. For example, rhythms can guide attention to specific parts of a melody and vice versa (Jones & Boltz, 1989; Jusczyk & Krumhansl, 1993; Palmer & Krumhansl, 1987; Schmuckler & Boltz, 1994). More importantly, such interactions appear to have similar effects between perceivers. Considering that these apparent interactions appear to account for a large proportion of the explainable variance of melodies, future attempts to predict memory for melody should investigate musical features that reflect interactions of rhythm and melodies. This finding also appears intuitive in the context of the Regenerative Multiple Representations conjecture, which asserts that multiple perceptual experiences lead to multiple memory representations (Herff et al., 2017; Herff, Olsen, et al., 2017; Herff et al., 2017) and prior knowledge informs how and if these representations are integrated into a coherent whole representation. Given that our participants all derived from a similar cultural background (Australian), it can be assumed that the way in which they integrate additional information into a coherent new melody representation would be similar. In turn, this would explain why multiple representations increase perceptual similarity between observers.

5.4 Conclusion

The present study analysed the data of 10 melody recognition experiments. Here, we used inter-rater agreement in the memory response patterns of listeners as a window into similarity of music perception. The results revealed striking similarities in listeners' memory response patterns, which we conclude is evidence of similarities' in listeners' perception of music. This conclusion is grounded in a new and developing Regenerative Multiple Representations conjecture that describes a crucial link between prior experience, perception, and subsequent formation of new memories (Herff et al., 2017; Herff, Olsen, et al., 2017; Herff et al., 2017). The findings reported here also reveal factors that influence the degree of similarity in listeners' memory for melodies, such as the influences of the particular melody corpora, tuning systems, testing paradigm, and whether melodies consisted of pitch-only, rhythm-only, or both pitch and rhythm information combined.

Furthermore, by examining inter-rater agreement in memory for melodies between listeners, this study also aimed to inform predictive models of melody recognition that use musical features as predictors. Overall, results suggest that future models that aim to predict melody recognition based on musical features should carefully consider the precise task instructions given to the participants and focus on musical features that describe the interaction between rhythm and melody, as well as musical features that are tuning-system independent. A model that can predict melody recognition beyond those variables using the commonalities in the predictive power of musical features would provide a strong framework for future research endeavours that not only aim to predict memory for melody, but memory for complete musical pieces.

Chapter 6

General Discussion

6.1 General Discussion

Memory is stimulus specific (Fougnie et al., 2015). As a complex, dynamic real-world stimulus, music in particular has intriguing properties when it comes to memory (Bailes, 2007, 2015; Baird & Samson, 2014; Chazin & Neuschatz, 1990; Colley, 2016; Cuddy & Duffin, 2005; Halpern & Bartlett, 2011; Jacobsen et al., 2015; Schulkind, 2009; Smith, 1985; Williams, 2015). The results presented here support these two statements once again. We collected and analysed the data of 10 experiments to investigate cumulative disruptive effects in memory for melody, specifically in melody recognition. Commonly in a recognition task, an increase in the number of intervening items until a stimulus is presented a second time leads to cumulative disruptive effects on recognition performance (Bui et al., 2014; Donaldson & Murdock, 1968; Hockley, 1992; Konkle et al., 2010; Nickerson, 1965; Olson, 1969; Poon & Fozard, 1980; Rakover & Cahlon, 2001; Sachs, 1967; Sadeh et al., 2014). We showed that melodies in a familiar tuning system do not show cumulative interference, whereas most other stimuli do (Chapter 2 To explain this result, we proposed a novel Regenerative Multiple Representations (RMR) conjecture (in 2.9). The RMR conjecture describes a crucial link between prior experience, perception, and subsequent formation of memories and predicts that melodies in unfamiliar rather than familiar tuning systems should show the common cumulative disruptive effects. We tested, supported, and replicated this prediction: melodies in unfamiliar tuning systems do show cumulative disruptive interference (Chapter 3 as well as 4.3). The RMR conjecture also predicts that if a combined stimulus type, such as a melody consisting of a pitch and rhythmic sequence in an unfamiliar tuning system, shows cumulative interference, so must the individual components as well. We tested this assumption and as predicted, both pitch sequences and rhythmic sequences of unfamiliar melodies show cumulative disruptive effects on recognition (Chapter 4 In a final step, we analysed the data of all experiments together to assess the degree of agreement between participants' recognition judgements (Chapter 5 In the following, we discuss the main findings and implications of this dissertation.

6.1.1 Overview of Main Findings

Resilient Memory for Melodies. From the perspective of memory, melodies are remarkable. Involuntary music imagery (e.g., 'ear-worms') shows how persistent memory for melody can be (Bailes, 2007, 2015; Halpern & Bartlett, 2011; Williams, 2015). The selective sparing of memory for melody in some forms of dementia and severe brain injures (Baird & Samson, 2014; Cuddy & Duffin, 2005; Jacobsen et al., 2015; Schulkind, 2009) adds to the assertion by some researchers that memory for melody might be 'special' (Jackendoff & Lerdahl, 2006; Schulkind, 2009; Stevens, 2015). In controlled laboratory studies, memory for melodies shows resilience towards the effect of temporal decay (Schellenberg & Habashi, 2015). The findings of four experiments reported here in Chapter 2 complement past findings of music's resilience towards temporal decay (Schellenberg & Habashi, 2015) by showing that memory for melody has remarkable resilience to interference. With up to nearly 200 intervening melodies, memory for melodies in a familiar tuning system does not show the cumulative disruptive effects that are so often observed using other stimuli (Bui et al., 2014; Donaldson & Murdock, 1968; Hockley, 1992; Konkle et al., 2010; Nickerson, 1965; Olson, 1969; Poon & Fozard, 1980; Rakover & Cahlon, 2001;

Sachs, 1967; Sadeh et al., 2014). However, we also observed distinct memory phenomena in memory for melody that are identical to the majority of studies that measure memory in response to other stimuli.

For example, a recency-in-memory effect, which refers to a significant recognition advantage for the last item encountered, is widely observed with most stimuli (Berz, 1995; Dowling, 1973; Greene & Samuel, 1986; Jahnke, 1963; Roberts, 1986). Results reported here show that this specific memory phenomenon is also prevalent in memory for melodies in a familiar tuning system (Chapter 2 The last melody encountered (zero intervening melodies) shows significantly better recognition performance than other melodies that, as discussed above, all show similar recognition performance, regardless of the number of intervening melodies. Interestingly, both of these memory phenomena – that is, a lack of a cumulative disruptive effect from the number of intervening items and a recency-in-memory effect – were reversed when listeners are presented with melodies in an unfamiliar tuning system.

Not-so-Resilient Memory for Melodies. The findings discussed above are not ubiquitous for all melodies. Indeed, if melodies are played in unfamiliar tuning systems such as those in Experiment 7 (Chapter 3 then the usual cumulative disruptive effects are observed. Furthermore, no recency-in-memory advantage materialises.

Experiment 1 and Experiment 2 (melodies in a familiar tuning system) in Chapter 2 show clear advantages for immediate melody recognition, whereas Experiment 5 and Experiment 6 (melodies in an unfamiliar tuning system) in Chapter

3 show no distinct recognition performance peak for zero intervening melodies. In Chapter 3 we also demonstrate that these effects are shaped by how closely a tuning system resembles a familiar tuning system.

Experiments 5 and 6 used a tuning system that is unfamiliar, but still resembled the familiar 12-TET tuning system in so far as it also used an equal temperament (88-CETs). Furthermore, Experiments 5 and 6 used the same stimuli from Experiments 3 and 4, but adjusted to the new unfamiliar tuning system. As a result, many musical features of the stimuli, such as pitch contour, were likely to be familiar to the listeners in Experiments 5 and 6, even though the overall pitch set was unfamiliar. The unfamiliar tuning system in Experiments 5 and 6 disrupted the recency-in-memory effect that was previously observed with melodies in a familiar tuning system. However, the similarities between stimuli in Experiments 5 and 6, and Experiments 3 and 4, was close enough to still provide melodies with some resilience to cumulative disruptive effects. This can be seen by the lack of cumulative disruptive interference within the first 13 intervening melody conditions in Experiments 5 and 6. Experiment 7, however, used stimuli that were even less familiar in regards to pitches as well as rhythms when compared to those used in Experiments 5 and 6. As a result, we observed cumulative disruptive interference in Experiment 7, even within the first 13 intervening melody conditions.

On a side note that might be relevant for future research, we collected pilot data with a high musical expertise group. It appears that even musicians do not show recency-in-memory effects in response to melodies in an unfamiliar tuning system (this pilot data is shown in Appendix E, p. 266). Future research could further investigate the role of expertise in cumulative interference in memory for melodies in an unfamiliar tuning system. Instead of further investigating the role of expertise, the present project followed the path of investigating the importance of melodies' underlying components; specifically, pitch and rhythmic sequences.

Pitch and Rhythmic Sequences. Pitch and rhythmic sequences are two of the main components of melodies. As mentioned above, we found cumulative disruptive effects from the number of intervening items using melodies in an unfamiliar tuning system (Experiment 7 in Chapter 3 and Experiment 8 in Chapter 4 Using the same stimuli split into their underlying pitch-only sequences (Experiment 9 in Chapter 4 and rhythm-only sequences (Experiment 10 in Chapter 5 we observed the same effect. Both unfamiliar pitch-only sequences and uncommon rhythm-only sequences show cumulative disruptive effects on recognition performance as the number of intervening sequences increased. This also suggests that if a combined stimulus, for example a simple melody that consists of a pitch-sequence and a rhythmic sequence, shows cumulative disruptive interference, then so does its underlying components.

The present investigation was also the first to investigate long-term rhythm recognition. Rhythm recognition has only been previously investigated in-depth at relative short delays (~ 10 sec)(Collier & Logan, 2000; Schaal, Banissy, et al., 2015). Here we demonstrated that even uncommon rhythms can be recognised significantly above chance, even after large numbers of intervening items. However, memory for rhythms (uncommon ones at least) seems to be subject to cumulative disruptive effects from intervening items.

The inverted memory phenomena (no recency-in-memory effect but cumulative disruptive interference) in melodies in unfamiliar tuning systems compared with melodies in familiar tuning systems seems puzzling on the surface. Existing and well-established memory theories usually predict clear decay and/or interference curves (Dennis & Humphreys, 2001; McClelland & Chappell, 1998; Norman, 2013; Oberauer & Lewandowsky, 2011; Oberauer et al., 2012; Shiffrin & Steyvers, 1997). In circumstances where no such curves were observed, such as poetry, the temporal and rhythmic properties of the stimulus may be responsible for the lack of decay or interference (Tillmann & Dowling, 2007). This makes sense, considering that both poetry and music unfold over time and usually have a rhythmic dimension (Jackendoff & Lerdahl, 2006). Yet, this would not explain the resilience to cumulative disruptive effects in memory for drawings, relative to photographs (Berman et al., 1991; Friedman, 1990a; Konkle et al., 2010). Furthermore, the temporal-rhythmical account cannot explain why melodies in an unfamiliar tuning system show cumulative disruptive interference, whereas melodies in a familiar tuning system do not. This is especially the case when one considers that the melodies in a familiar tuning system in Experiments 3 and 4 and the melodies in an unfamiliar tuning system in Experiments 5 and 6 used the exact same rhythms. In addition, the rhythm-only sequences in Experiment 10 in Chapter 4 also show cumulative disruptive effects, which is also at odds with a purely temporal/rhythmical account of this phenomenon. In an attempt to explain these findings and provide testable predictions such as those found in 2.9.1, a novel RMR conjecture was developed.

6.1.2 Theoretical Implications of the Dissertation

The Regenerative Multiple Representations Conjecture. The RMR conjecture draws from several key experimental findings and established memory theories (e.g., the Dual-coding theory of Paivio, 1969). One key finding that inspired this conjecture is that words and photographs show cumulative disruptive interference, whereas poetry and drawings, similar to our findings with melodies, do not (Berman et al., 1991; Friedman, 1990a; Konkle et al., 2010; Tillmann & Dowling, 2007). An intriguing commonality between the stimuli that do not show cumulative disruptive effects are strong interdependent connections between underlying components that are integrated into a coherent whole in a 'process of perceptual synthesis' (D. Deutsch, 1986). Indeed, music, poetry, and drawings seem to be perceived as underlying components that are integrated into a coherent whole. Importantly, both the underlying components and the coherent whole are remembered. It is argued here that this is a crucial difference between music, poetry, and drawings, all of which show resilience towards cumulative disruptive effects from intervening items, and their counterparts of spoken words, prose, and photographs, all of which do show cumulative disruptive effects. It appears that in stimuli such as spoken words, prose, and photographs, a memory representation is formed for the integrated whole, whereas the underlying components are rapidly forgotten (Babcock & Freyd, 1988; D. Deutsch, 1986; Knoblich et al., 2002; Krumhansl, 1991, p. 295; Rayner et al., 2012a, 2012b; Schneider, 1997, p. 119; Tse & Cavanagh, 2000). So why might we remember only the coherent whole for some stimuli, but remember a coherent whole *as well as* the underlying components for other stimuli?

We suggest the reason lies in prior experience. Over the course of a lifetime we learn the most relevant way of perceiving a stimulus (Goldstone, 1998). A similar conception has previously been described as *pertinence*, a weighting of a perception based on a current situation, as well as a long-term factor such as prior knowledge (D. Deutsch (1986) (see also J. A. Deutsch and Deutsch, 1963). Indeed, experimental studies show that prior knowledge influences and guides perception (Malmberg & Annis, 2012). Returning to why this is relevant for the lack of cumulative disruptive effects in music, poetry, and drawings, we suggest that these are all stimuli in which we have learned to pay close attention, not only to the integrated whole, but to the underlying components. For example, in prose, the semantic information is by far the most important information; underlying words are of less relevance (Rayner et al., 2012b). In poetry, on the other hand, the precise underlying wording is similarly important to the overarching semantic and affective meaning. Details and further examples are discussed in 2.9. Given that prior knowledge informs multiple perception of stimuli like music, poetry, and drawings, which in turn, leads to multiple memory representations, the question remains: Why does this process lead to increased resilience towards cumulative disruptive effects? Interestingly, the suggestion that multiple representations assist in retrieval has been proposed in the past, and a large corpus of research supports that claim. This is further discussed in 2.9.2.

In summary, the RMR conjecture combines previous experimental results and memory theories to describe a crucial link between prior knowledge, perception, and formation of new memories. The conjecture is speculative but does allow some precise predictions that can guide research. Some of these predictions were tested over the course of this doctoral research.

Testing the RMR conjecture. The RMR conjecture asserts that prior knowledge informs the process of perception. For example, we perceive the underlying components of a melody and because we are familiar with the 'rules' (that is, the statistical occurrence) of how melodies' underlying components inter-relate, we form an integrated, coherent representation of the melody as a whole (Cui et al., 2015; Saffran et al., 1999; Schon & Francois, 2011; Tillmann & McAdams, 2004). However, if we are not familiar with these rules, the RMR conjecture predicts that we cannot form a representation of the melody as an integrated whole. Furthermore, the RMR conjecture postulates that it is multiple representations of the melodies that provide memory for melody with the resilience towards cumulative interference observed in Chapter 2 This means that disrupting the formation of integrated representations of a simple melody should disrupt the previously observed resilience towards cumulative interference. This also means that it should be possible to disrupt perception and thus representation of a melody as a whole, by presenting melodies in a tuning system unfamiliar to the listener. We tested this prediction, and, as previously discussed, this is precisely what we observed in Chapter 3 as well as Section 4.3. This lends preliminary support to the conjecture, motivates further empirical testing and sheds light on why memory for music is 'special'.

In Chapter 4 we reported a series of experiments that had the potential to fundamentally falsify the RMR conjecture. If melodies in an unfamiliar tuning system that consist of a combination of underlying components (pitch-sequence and rhythm-sequence) show cumulative disruptive interference, then following the RMR conjecture, so must the underlying components when tested separately. This is because the underlying components separately provide fewer possible percepts and memory representations. This hypothesis was tested and supported. Currently, the RMR conjecture is capable of making clear, falsifiable predictions. It is a useful tool to inspire further research, especially in, but not limited to, the domain of music perception. The next section provides an overview of some of the RMR conjecture's testable predictions that future research can address.

6.1.3 Future Directions

Future Directions Within the Domain of Music. Atonal melodies provide an intriguing middle ground between melodies in a familiar tuning system and melodies in an unfamiliar tuning system. This is because atonal melodies can utilise the pitch set familiar to listeners, yet may not follow the hierarchical rules that characterize the Western tonal tradition. The RMR conjecture predicts stronger cumulative interference for atonal melodies relative to melodies in a familiar tuning system. Furthermore, comparing cumulative interference in memory for melodies in an unfamiliar tuning system with such effects in atonal music would deepen our understanding of the independent contribution of prior low-level perceptual experience (familiarity with a pitch-set and specific intervals) and prior higher-level statistical experience (the hierarchical rules underlying a music tradition) to memory. This is because, in relation to melodies in a familiar tuning system, atonal melodies 'only' disrupt the benefits of higher-level prior experience with underlying rules, not necessarily prior low-level experience with a specific pitch-set and intervals.

A novel, unfamiliar tuning system like the system used in Experiment 7, however, disrupts the usefulness of both prior experience with the pitch set used in the Western music tradition, as well as the underlying rules common to this tradition. The data from Experiments 5 and 6 complements such new data with atonal melodies. This is because in Experiments 5 and 6, we utilised melodies in an unfamiliar tuning system that used an unfamiliar pitch-set, but the statistical rules were identical to those used in the Western tonal tradition.

Given that the RMR conjecture emphasises the influence of prior experience on perception, the resilience of memory for melodies to cumulative disruptive effects should arise in unfamiliar tuning system as listeners get used to that tuning system. This prediction could be tested in a cross-cultural context. For example, listeners familiar to the western tuning tradition should show cumulative disruptive effects for melodies in other music traditions. Whereas listeners that are only familiar with the music tradition of their culture should show cumulative interference for melodies in the Western tradition. Furthermore, as these listeners become more familiar with other cultures' musical traditions, the magnitude of cumulative disruptive effects should slowly decrease (some support for this can be found in Kessler, Hansen, & Shepard, 1984).

In an international collaboration we are currently planning to investigate another prediction as a result of this dissertation and our work on the RMR conjecture. This international project aims to critically investigate the crucial link between prior experience, perception, and formation of memory postulated by the RMR conjecture. In this project, we are planning to develop a novel experimental paradigm. The paradigm aims to utilise auditory perceptual illusions (e.g., virtual and spectral pitches, see Seither-Preisler et al., 2007). A stimulus set will be crafted where information about the nature of the illusion fundamentally changes perception from random sounds to a coherent melody. Participants performed a memory task where only one half of the participants were provided information about the perceptual illusion. The RMR conjecture predicts that participants who are provided with additional information about the stimuli that will help form multiple memory representations will subsequently show less cumulative disruptive interference on recognition.

Overall, the testing and development of the RMR conjecture will benefit strongly from further investigation of the precise underlying mechanisms of the regeneration process. Specifically, in the context of music, future studies could investigate which musical features are commonly remembered immediately, which features can be regenerated, and which features are irretrievably 'lost'.

Future Directions Outside the Domain of Music. The implications of the present findings and the RMR conjecture are not specific to the domain of music. One intriguing prediction of the RMR conjecture is presented in Chapter 2 words written in longhand should induce more memory representations than the same words written in plain printed letters. This is because readers track the direction as well as mode of production of words written in longhand (Babcock & Freyd, 1988; Knoblich et al., 2002; Tse & Cavanagh, 2000). As a result, the RMR conjecture predicts that readers not only form a memory representation of the integrated meaning of the word, but also of the underlying components that are unique to a specific style of longhand. These multiple representations should provide words written in longhand with resilience towards cumulative disruptive interference when compared to words written in plain print. This prediction is precise and falsifiable, and is analogous to the observations of cumulative disruptive interference in response to photographs, but not drawings (Berman et al., 1991; Friedman, 1990a; Konkle et al., 2010).

The link between prior experience, perception, and formation of new memories provides a framework to make predictions about domain-specific expertise. The RMR conjecture suggests that high domain-specific expertise is often associated with changes of perception relative to low domain-specific expertise. For example; an architect can accurately identify and perceive combinations of architectural styles that may be irrelevant for a normal perceiver; a botanist is capable of identifying differences between plant species that are perceptually invisible for uninformed observers; and an entomologist can distinguish species that would be labelled identically by normal observers. Some of these changes in perception occur rapidly after being exposed to key information, whereas others require years of training. For example, Asian and African elephants can be hard to differentiate. However, if one is informed that Asian elephants have two bumps on their heads whereas African elephants have one, you can easily distinguish the two species from one another. As a result of this newly acquired knowledge, perception changes in the future and future memory representations of elephants are enriched with the information of whether an elephant was African or Asian. This additional memory representation then aids memory formation. For example, if someone asks, 'How many bumps did the elephant that you saw earlier have on its head?', one's visual memory representation might have faded beyond the point where one can remember the exact number of bumps. However, the perceiver may still remember that they saw an Asian elephant, which can help 'regenerate' the information that the elephant had two bumps. In this example, a small piece of information changes perception and, subsequently, the formation of new memory representations.

Other changes in perception might require intense training. Learning to perceive and produce pronunciation differences in an unfamiliar language, for example, can take a long time (Browman & Goldstein, 1995; Escudero & Chládková, 2010). In the framework of the RMR conjecture, the above examples are linked by the fact that additional experience (instructed or learned over long exposure) changes perception. The RMR conjecture assumes that these additional precepts form additional memory representations that increase resilience to cumulative disruptive effects. Domain specific expertise can increase memory performance; however, it is important to note that the increased resilience toward intervening items is independent of overall memory performance.

On a general note, when investigating memory in multiple domains it is tempting to conduct cross-domain comparisons of absolute memory performance. However, such comparisons should be conducted with care. For instance, cross-domain memory performance comparisons are difficult at best, mostly because what is being remembered can be of a dramatically different form and involve different modalities (Cohen, Horowitz, & Wolfe, 2009, Cohen, Evans, Horowitz, & Wolfe, 2011). This is one advantage of investigating fundamental memory phenomena (such as recency-in-memory and cumulative disruptive interference), as they contain both qualitative and quantitative features, rather than purely quantitative as in absolute memory performance. The presence or absence of specific and fundamental memory phenomena (such as recency-in-memory, and cumulative disruptive effects) can reasonably be compared between domains.

Stepping away from the RMR conjecture and intervening item effects, we now explore a broader, more practical perspective on the implications of the current findings.

6.1.4 Practical Implications of the Dissertation

Implications for Listeners' Similarity in Music Perception. Within the visual domain, memorability of pictures appears to be a stable property across human observers. In other words, visual pictures that are more memorable for one person are also most likely more memorable for another person (Isola et al., 2011). Within the domain of music, we know that a culture without music is yet to be found (Cross, 1999, 2001, 2014). Clearly, different listeners perceive music quite differently. For example, in Chapter 3 we demonstrate the profound influence that listeners' familiarity with a tuning system can have on fundamental memory phenomena, and there is a wealth of research investigating the fascinating differences in the music traditions between different cultures (Stevens, 2012; Stevens, Tardieu, Dunbar-Hall, Best, & Tillmann, 2013). So, how similar is human perception of music? We investigated this question in the context of memory using the combined data of all participants from all experiments. Following from the RMR conjecture that assumes a crucial link between prior knowledge, perception, and memory formation, we assume that the more similar listeners' memory response patterns are, the more similar their perception of the stimulus must have been. We correlated the response patterns of the participants towards the set of melodies in each experiment. Through the procedure detailed in Chapter 5 we obtained average split-half correlations for each set of melodies. These values represent inter-rater agreement and are a proxy of how stable mean values of memorability and perception of the melodies are across listeners. Split-half correlation values for each melody set were separately obtained here. This allowed us to investigate how different properties of each melody set affects memory stability between participants' responses.

We found that a set of melodies composed in the spirit of modern advertisement melodies (Experiments 1 and 2) produced remarkably high average split-half correlations. Interestingly, another melody corpus (Experiment 3) of European folk songs in the same familiar tuning system yielded far lower average split-half correlations. This suggests that different melody corpora can lead to substantially different degrees of similarity in listeners' music perception, even if both corpora use melodies in the same familiar tuning system. Furthermore, Experiment 8 was a direct replication of Experiment 7. Both Experiments yielded similar average split-half correlations. This demonstrates the degree of similarity in listeners' response patterns is relatively stable and replicable across experiments with the same melody corpus. Interestingly, average split-half correlations did not change significantly when melodies from European folk songs (Experiment 3) in a familiar tuning system were played in an unfamiliar tuning system instead (Experiment 5). This suggests that possible contributors to the degree of agreement between human listeners may be invariant to the tuning system; for example, contour and rhythm.

Some insight into possible contributors to the degree of agreement between human listeners was found in the average split-half correlation analysis of the experiments presented in Chapter 4 Experiment 8 used a corpus of melodies in an unfamiliar tuning system consisting, as melodies usually do, of pitch-sequences and rhythm-sequences combined. When Experiment 8 and Experiment 10 tested the sequences individually, the rhythm-sequences showed far higher average split-half correlations compared to the pitch-sequence. Combined with the finding that the exact tuning system has only minimal impact on the listener's agreement, this suggests that the rhythmic sequences may be a major contributor to the similarity of melody memorability across listeners. Another interesting observation was that the summed average split-half correlations of the pitch-only and rhythm-only sequences fell well short of the split-half correlations of the melodies that consisted of both sequences combined. This suggests exciting interactions between pitch-sequences and rhythm-sequences that may increase listeners' agreements. These interactions, as well as the role of rhythmic sequences, carry further practical implications.

Practical Implications for Memory Research. One of the practical implications for memory research to emerge from this dissertation is the technique developed to account for participants' response tendencies. Accounting for participants' response biases is challenging. Static response biases are often encountered and widely discussed in the memory literature (Snodgrass & Corwin, 1988). However, participants also have the tendency to dynamically change their response biases over the course of an experiment (Berch, 1976; Donaldson & Murdock, 1968). These dynamic tendencies are far less often accounted for in the statistical analysis. This is an issue, because if we only account for each participant's static response bias then we are misrepresenting the participant's actual recognition performance as their response bias at the beginning and the end of the experiment might be dramatically different. With the dynamic response tendency models (see 2.4.1) we provide a participant-wise, simple, intuitive, and easy to implement solution to capture and account for the shifts of participants' response biases over the course of an experiment. Fundamentally, the dynamic response tendency models normalise hit-rates over false-alarm rates continuously over the course of the experiment and for each participant individually (Snodgrass & Corwin, 1988). An example of dynamic response tendency shifts can be found in Appendix C, p. 261. Importantly, the approach is not limited to our data, but is also applicable to any experiment that faces potential shifts in response biases with increasing trial number.

Practical Implications for Predictive Models. Predicting memory, specifically recognition, rather than explaining it post-hoc, is a complex endeavour. A common approach uses stimulus features to predict average recognition performance (Cortese et al., 2010; Müllensiefen & Halpern, 2014). Some of the difficulties in constructing a predictive feature model lie in the selection of precise features to focus on, as well as the fact that recognition is influenced by factors that are not inherent to the stimulus (Cuddy et al., 2012; Gardiner et al., 1996; Herff & Czernochowski, submitted; McAuley et al., 2004; Samson et al., 2009). The results of this dissertation suggest that models attempting to predict melody recognition performance based on musical features should pay special attention to features representing rhythmic information; features that model the interaction of pitch-sequences and rhythm-sequences may be particularly informative. Musical features that represent tuning system information, on the other hand, may carry less predictive power. Interestingly, in IDyOM (Pearce, 2005) – a computational model of music perception that uses a short-term memory component (e.g. a musical sequence) and a long-term memory component (e.g. an underlying corpus) – a melody in an familiar tuning system is computed to comprise the same short-term memory information content as an identical melody but sounded in an unfamiliar tuning system. Only the long-term memory component would differ between these two versions of a melody. The present results therefore suggest that when researchers are interested in predicting memory for melody, IDyOM's short-term component may be a useful tool.

Our present results may also help in approximating how useful a musical feature model would be in the first place. Supposedly, melody recognition is influenced by factors that are similar between subjects, such as musical features, and factors that vary severely between listeners, such as musical expertise. Stimulus feature models make use of the factors that are relatively invariant between observers. As a result, Flexer and Grill (2016) suggest that the degree of overlap between participants' response patterns towards a set of melodies informs models that predict listeners' responses based on musical features. In other words, the more similarly participants behave, the more likely it is that musical features carry a lot of predictive power in relation to the variation between participants, whereas the more dissimilarly participants' behave, the more likely it is that the predictive power of musical features is relatively limited. In Chapter 5 we implemented a method of measuring average split-half correlations in participants' responses for each melody corpus and investigated how these values approximate similarity between participants' response patterns. Consequently, this method also has the potential to inform the predictive power of musical feature models. Within the domain of music, we demonstrated that the potential predictive power of musical feature models varies dramatically between melody corpora. The results of Chapter 5 also suggests that future predictive models of melody recognition should focus on features that represent the interaction of pitch and rhythm, rather than treating those components independently, as well as features that are independent of tuning system.

Note that the variance of an individual's recognition performance with melodies was not included in the split-half correlations discussed, which concerned average performance patterns. However, just as a predictive model can be made of average performance patterns with a likely success that depends on the degree of inter-personal similarity in response, so a predictive model could be made of individual variances for individual melodies. Whether the present findings translate to such variance-based models awaits further investigation.

Stimulus Selection Procedure. To adequately represent a huge underlying corpus of European folk songs we used a mixed methods approach. The approach is detailed in Appendix B (p. 250) and describes the combination of feature- and cluster analysis, as well as perceptual piloting. Fortunately for us, we live in a time where tremendous numbers of stimuli are easily accessible. Where previous researchers had to manually gather melodies, we now have huge melody corpora such as KernScores and the Million Song Dataset (Bertin-Mahieux, Ellis, Whitman, & Lamere, 2011; CCARH; Sapp, 2005). Stimuli are increasingly drawn from stimulus corpora, often randomly to adequately represent the underlying corpus. Unfortunately, depending on which stimulus factors one is interested in, one would need to draw a very large number of stimuli from the underlying corpus to provide an adequate distribution within the factor that you are interested in. This is particularly true if you attempt to represent the underlying corpus in regards to multiple underlying dimensions. However, testing large numbers of stimuli in an experimental study is often not feasible. More concretely, in the present study we required a sample of ~100 European Folk melodies, shorter than 15 seconds, and unfamiliar to our Australian participants. We wanted these 100 melodies to adequately represent the underlying melody corpus to maximise generalizability of our results. To achieve this, we placed all melodies in a multi-dimensional space, where each dimension represents a potential factor of interest (musical features in our case). We excluded melodies that dropped out by a limiting criterion, in our case, longer than 15 seconds. We then performed dimension reduction and cluster analysis, all detailed in Appendix B (p. 250), and drew our actual stimuli from the underlying clusters proportional to cluster size. In the next step, we recruited a group of participants to perceptually test if the melodies are indeed perceived as unfamiliar. To this end, we identified the response distribution that a group of participants would give to an unfamiliar melody. We then compared the response distribution of each melody towards this prototype distribution and removed any melody that was significantly different. While this approach may seem lengthy and complicated, it ensures that a sample of stimuli drawn from an underlying corpus possesses the assumed properties.

We hope that the detailed description in Appendix B (p. 250) might serve as guide to help others in drawing a stimulus sample from an underlying corpus, whilst ensuring specific properties of the sample. Of course, there are other possible approaches to achieve the same outcome; however, Appendix B (p. 250) provides a relatively easy-to-follow, yet rigorous approach that can be adjusted to the specific needs of a given experiment.

Finally, one of the main possible contributions of this dissertation is the data itself. In the process of the aforementioned stimulus selection procedure, we implemented several routines to calculate various musical features. Therefore, the musical feature calculation plus the complete memory dataset, in conjunction with the practical and theoretical implications outlined above, all combine to provide a promising foundation for future in-depth exploration of mathematical models capable of using musical features to predict memory for melodies. We intend to build such models in the future using the insight and data provided here.

6.2 General Conclusion

Within the interdisciplinary space between memory and music, we investigated interference in memory and the issue of memory's domain specificity. We showed that memory for melodies in a familiar tuning system is resilient to cumulative disruptive effects from the number of intervening items, whereas most other stimuli show systematic decreases as the number of intervening items increases. We provide an explanation in the form of a novel Regenerative Multiple Representations (RMR) conjecture that emphasises a crucial link between prior experience, perception, and subsequent formation of memories. This conjecture provides clear predictions, for example: (1) that melodies in *unfamiliar* tuning systems, rather than *familia*r tuning systems should show cumulative disruptive effects; (2) that a profound memory phenomenon such as the recency-in-memory effect can be disrupted if we prevent formation of an integrated memory representation of the melody as a whole; and (3) that when an integrated stimulus shows cumulative disruptive interference, then so must its underlying parts. We tested and supported all of the above predictions, thereby demonstrating that the RMR conjecture is a useful tool to build predictions for future research to consider. Several additional testable predictions of the RMR conjecture are provided throughout this dissertation and the conjecture awaits further testing and mathematical implementation to evolve into a comprehensive theory.

Here, we also explored similarity between the memory response patterns of human listeners when performing a memory for melody task. Results reveal significant interactions between rhythmic sequences and pitch sequences of melodies that increase the similarity between listeners' responses. In addition, we provided a basic stimulus selection protocol to draw stimuli from an underlying corpus in a principled fashion, as well as implemented a procedure to reveal the likely predictive power of models using stimulus features to predict memory performance within a given stimulus corpus.

In summary, the present dissertation conducted a systematic investigation of memory for melody and in the process, has contributed to our understanding of fundamental memory phenomena, provided practical implementations, and ultimately, has shed further light on the mechanisms that explain why music may indeed be 'special'.

References

Anderson, J.R., & Bower, G.H. (2014). *Human associative memory*. New York: Psychology press.

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-195). New York: Academic Press.

Baayen, R.H. (2008). *Analyzing linguistic data: A practical introduction to statistics using r*. New York: Cambridge University Press.

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412. doi: 10.1016/J.Jml.2007.12.005

Babcock, M.K., & Freyd, J.J. (1988). Perception of dynamic information in static handwritten forms. *The American journal of psychology*, 111-130.

Bachem, A. (1954). Time factors in relative and absolute pitch determination. *The Journal of the Acoustical Society of America, 26*, 751-753.

Baddeley, A. (1997). *Human memory: Theory and practice, revised edition*. East Sussex: Psychology Press.

Baddeley, A. (2014). *Essentials of human memory (classic edition)*. East Sussex, New York: Psychology Press.

Bailes, F. (2007). The prevalence and nature of imagined music in the everyday lives of music students. *Psychology of Music, 35*, 555-570.

Bailes, F. (2015). Music in mind? An experience sampling study of what and when, towards an understanding of why. *Psychomusicology: Music, Mind, and Brain, 25*, 58-68.

Bailes, F., Dean, R.T., & Pearce, M.T. (2013). Music cognition as mental time travel. *Scientific Reports, 3*. doi: 10.1038/srep02690

Baird, A., & Samson, S. (2014). Music evoked autobiographical memory after severe acquired brain injury: Preliminary findings from a case series. *Neuropsychological Rehabilitation, 24*, 125-143. doi: 10.1080/09602011.2013.858642

Balkwill, L.L., & Thompson, W.F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception, 17*, 43-64.

Barascud, N., Pearce, M.T., Griffiths, T.D., Friston, K.J., & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences*, E616-E625. doi: 10.1073/pnas.1508523113

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278. doi: 10.1016/j.jml.2012.11.001

Bartlett, J.C., & Dowling, W.J. (1980). Recognition of transposed melodies - a key-distance effect in developmental perspective. *Journal of Experimental Psychology-Human Perception and Performance, 6*, 501-515. doi: 10.1037/0096-1523.6.3.501

Bartlett, J.C., Halpern, A.R., & Dowling, W.J. (1995). Recognition of familiar and unfamiliar melodies in normal aging and alzheimers-disease. *Memory and Cognition, 23*, 531-546. doi: 10.3758/Bf03197255

Bates, D.M., Maechler, M., Bolker, B., & Walker, S. (2013). Lme4: Linear mixed-effects models using eigen and s4. *R package version, 1*.

Berch, D.B. (1976). Criterion change in continuous recognition memory: A sequential effect. *Bulletin of the Psychonomic Society, 7*, 309-312. doi: 10.3758/BF03337199

Berman, S., Friedman, D., & Cramer, M. (1991). Erps during continuous recognition memory for words and pictures. *Bulletin of the Psychonomic Society, 29*, 113-116.

Bertin-Mahieux, T., Ellis, D.P., Whitman, B., & Lamere, P. (2011). *The million song dataset*. Paper presented at the ISMIR.

Berz, W.L. (1995). Working memory in music: A theoretical model. *Music Perception*, 353-364.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capcacities that do not depend on formal musical training. *Cognition*, 100(1), 100-130. doi: 10.1016/j.cognition.2005.11.007

Brandler, S., & Rammsayer, T.H. (2003). Differences in mental abilities between musicians and non-musicians. *Psychology of music, 31*, 123-138.

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The "ticktock" of our internal clock direct brain evidence of subjective accents in isochronous sequences. *Psychological Science, 14*, 362-366.

Browman, C.P., & Goldstein, L. (1995). Dynamics and articulatory phonology. *Mind as motion*, 175-193.

Buchsbaum, B.R., Padmanabhan, A., & Berman, K.F. (2011). The neural substrates of recognition memory for verbal information: Spanning the divide between short- and long-term memory. *Journal of Cognitive Neuroscience, 23*, 978-991. doi: 10.1162/Jocn.2010.21496

Bui, D.C., Maddox, G.B., Zou, F., & Hale, S.S. (2014). Examining the lag effect under incidental encoding: Contributions of semantic priming and reminding. *Quarterly Journal of Experimental Psychology, 67*, 2134-2148. doi: 10.1080/17470218.2014.909506

Cabeza, R., Kapur, S., Craik, F.I.M., McIntosh, A.R., Houle, S., & Tulving, E. (1997). Functional neuroanatomy of recall and recognition: A pet study of episodic memory. *Journal of Cognitive Neuroscience, 9*, 254-265. doi: Doi 10.1162/Jocn.1997.9.2.254

Campeanu, S., Craik, F.I.M., Backer, K.C., & Alain, C. (2014). Voice reinstatement modulates neural indices of continuous word recognition. *Neuropsychologia, 62*, 233-244. doi: 10.1016/J.Neuropsychologia.2014.07.022

Castellano, M.A., Bharucha, J.J., & Krumhansl, C.L. (1984). Tonal hierarchies in the music of north india. *Journal of Experimental Psychology: General, 113*, 394.

CCARH. European folk songs. retrieved 2014, from http://kern.ccarh.org/cgi-bin/browse?l=essen/europa

Chazin, S., & Neuschatz, J.S. (1990). Using a mnemonic to aid in the recall of unfamiliar information. *Perceptual and Motor Skills, 71*, 1067-1071.

Clark, S.E., & Gronlund, S.D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3*, 37-60. doi: 10.3758/Bf03210740

Clarkson, D.B., & Jennrich, R.I. (1988). Quartic rotation criteria and algorithms. *Psychometrika, 53*, 251-259. doi: 10.1007/Bf02294136

Cohen, J. (1988). *Statistical power anlaysis for the behavioural sciences* (2nd ed. ed.). Mahwah, NJ: L. Erlbaum Associates.

Cohen, M.A., Evans, K.K., Horowitz, T.S., & Wolfe, J.M. (2011). Auditory and visual memory in musicians and nonmusicians. *Psychonomic Bulletin & Review, 18*, 586-591. doi: 10.3758/s13423-011-0074-0

Cohen, MA., Horowitz, T, S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. Proceedings of the National Academy of Sciences 106.14: 6008-6010

Colley, N. (2016). *Learning styles and exam preparation: Can music facilitate the learning of abstract concepts*. (Bachelof of Science), Western Sydney University.

Collier, G.L., & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory & Cognition, 28*, 529-538. doi: 10.3758/Bf03201243

Corrigall, K.A., & Trainor, L.J. (2014). Enculturation to musical pitch structure in young children: Evidence from behavioral and electrophysiological methods. *Developmental Science, 17*, 142-158. doi: 10.1111/desc.12100

Cortese, M.J., Khanna, M.M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory, 18*, 595-609. doi: Pii 924954056 10.1080/09658211.2010.493892

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Essence of Memory, 169*, 323-338. doi: 10.1016/S0079-6123(07)00020-9

Cross, I. (1999). Is music the most important thing we ever did? Music, development and evolution. In S. W. Yi (Ed.), *Music, mind and science*. Seoul: Seoul National University Press.

Cross, I. (2001). Music, cognition, culture, and evolution. *Biological Foundations of Music, 930*, 28-42.

Cross, I. (2014). Music and communication in music psychology. *Psychology of Music, 42*, 809-819. doi: 10.1177/0305735614543968

Cuddy, L.L., & Duffin, J. (2005). Music, memory, and alzheimer's disease: Is music recognition spared in dementia, and how can it be assessed? *Medical Hypotheses, 64*, 229-235. doi: 10.1016/J.Mehy.2004.09.005

Cuddy, L.L., Duffin, J.M., Gill, S.S., Brown, C.L., Sikka, R., & Vanstone, A.D. (2012). Memory for melodies and lyrics in alzheimer's disease. *Music Perception, 29*, 479-491. doi: 10.1525/Mp.2012.29.5.479

Cui, A.X., Collett, M.J., Troje, N.F., & Cuddy, L.L. (2015). Familiarity and preference for pitch probability profiles. *Cognitive Processing, 16*, 211-218. doi: 10.1007/s10339-015-0651-7

Cycling74. (2014). Max-msp. Retrieved 19/6, 2014, from http://cycling74.com/products/max/

Dean, R.T., Bailes, F., & Drummond, J. (2014). Generative structures in improvisation: Computational segmentation of keyboard performances. *Journal of New Music Research*, 1-13. doi: 10.1080/09298215.2013.859710

Dean, R.T., Bailes, F., & Dunsmuir, W.T.M. (2014). Time series analysis of real-time music perception: Approaches to the assessment of individual and expertise differences in perception of expressed affect. *Journal of Mathematics and Music, 8*, 183-205. doi: 10.1080/17459737.2014.928752

Dean, R.T., & Dunsmuir, W.T.M. (2015). Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. *Behavior Research Methods*, 1-20. doi: 10.3758/s13428-015-0611-2

Dennis, S., & Humphreys, M.S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*, 452-478.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science, 168*, 1604-1605. doi: 10.1126/science.168.3939.1604

Deutsch, D. (1972). Mapping of interactions in the pitch memory store. *Science, 175*, 1020-1022.

Deutsch, D. (1975). The organization of short-term memory for a single acoustic attribute. In D. Deutsch & J. A. Deutsch (Eds.), *Short-term memory* (pp. 107-151). New York: Academic Press.

Deutsch, D. (1986). Auditory pattern recognition. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. II, Cognitive Processes and Performance, pp. 32-31 - 32-49). New York: Wiley.

Deutsch, D. (1991). The tritone paradox: An influence of language on music perception. *Music Perception: An Interdisciplinary Journal, 8*, 335-347.

Deutsch, D., & Boulanger, R.C. (1984). Octave equivalence and the immediate recall of pitch sequences. *Music Perception: An Interdisciplinary Journal, 2*, 40-51.

Deutsch, J.A., & Deutsch, D. (1963). Attention - some theoretical considerations. *Psychological Review, 70*, 80-90. doi: 10.1037/h0039515

Donaldson, W., & Murdock, B.B. (1968). Criterion change in continuous recognition memory. *Journal of Experimental Psychology, 76*, 325-330. doi: 10.1037/H0025510

Dowling, W.J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics, 14*, 37-40.

Dowling, W.J. (1991). Tonal strength and melody recognition after long and short delays. *Perception and Psychophysics, 50*, 305-313. doi: 10.3758/Bf03212222

Dowling, W.J., & Bartlett, J.C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology: A Journal of Research in Music Cognition, 1*, 30-49. doi: 10.1037/h0094275

Dowling, W.J., & Fujitani, D.S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49*, Suppl 2:524+.

Dowling, W.J., Kwak, S., & Andrews, M.W. (1995). The time course of recognition of novel melodies. *Perception and Psychophysics, 57*, 136-149. doi: 10.3758/BF03206500

Dowling, W.J., Magner, H., & Tillmann, B. (2016). Memory improvement with wide-awake listeners and with nonclassical guitar music. *Psychomusicology: Music, Mind, and Brain, 26*, 26-34.

Dowling, W.J., & Tillmann, B. (2014). Memory improvement while hearing music: Effects of structural continuity on feature binding. *Music Perception, 32*, 11-32. doi: 10.1525/Mp.2014.32.1.11

Dowling, W.J., Tillmann, B., & Ayers, D.F. (2001). Memory and the experience of hearing music. *Music Perception, 19*, 249-276. doi: 10.1525/Mp.2001.19.2.249

Eerola, T., & Toiviainen, P. (2004a). Midi toolbox: Matlab tools for music research. http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/miditoolbox/

Eerola, T., & Toiviainen, P. (2004b, October 10-14, 2004). *Mir in matlab: The midi toolbox*. Paper presented at the 5th International Conference on Music Information Retrieval, Universitat Pompeu Fabra Barcelona, Spain.

Erdelyi, M.H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology, 6*, 159-171.

Escudero, P., & Chládková, K. (2010). Spanish listeners' perception of american and southern british english vowels. *The Journal of the Acoustical Society of America, 128*, EL254-EL260.

Eysenck, M.W., & Keane, M.T. (2015). *Cognitive psychology: A student's handbook* (Seventh Edition ed.). London and New York: Taylor & Francis.

Fernandes, M.A., & Moscovitch, M. (2000). Divided attention and memory: Evidence of substantial interference effects at retrieval and encoding. *Journal of Experimental Psychology-General, 129*, 155-176. doi: 10.1037/0096-3445.129.2.155

Ferris, S.H., Crook, T., Clark, E., Mccarthy, M., & Rae, D. (1980). Facial recognition memory deficits in normal aging and senile dementia. *Journals of Gerontology, 35*, 707-714.

Finnigan, S., Humphreys, M.S., Dennis, S., & Geffen, G. (2002). Erp 'old/new' effects: Memory strength and decisional factor(s). *Neuropsychologia, 40*, 2288-2304. doi: 10.1016/S0028-3932(02)00113-6

Fleischman, D.A., Wilson, R.S., Gabrieli, J.D.E., Bienias, J.L., & Bennett, D.A. (2004). A longitudinal study of implicit and explicit memory in old persons. *Psychology and Aging, 19*, 617-625. doi: 10.1037/0882-7974.19.4.617

Flexer, A., & Grill, T. (2016). The problem of limited inter-rater agreement in modelling music similarity. *Journal of New Music Research, 45*, 239-251. doi: 10.1080/09298215.2016.1200631

Fougnie, D., Zughni, S., Godwin, D., & Marois, R. (2015). Working memory storage is intrinsically domain specific. *Journal of Experimental Psychology-General, 144*, 30-47. doi: 10.1037/a0038211

Franek, M., van Noorden, L., & Rezny, L. (2014). Tempo and walking speed with music in the urban context. *Frontiers in Psychology, 5*, 1361:1361-1314. doi: 10.3389/Fpsyg.2014.01361

Freyd, M. (1923). The graphic rating scale. *Journal of Educational Psychology, 14*, 83-102. doi: 10.1037/h0074329

Friedman, D. (1990a). Cognitive event-related potential components during continuous recognition memory for pictures. *Psychophysiology, 27*, 136-148.

Friedman, D. (1990b). Erps during continuous recognition memory for words. *Biological Psychology, 30*, 61-87. doi: 10.1016/0301-0511(90)90091-A

Gardiner, J.M., Kaminska, Z., Dixon, M., & Java, R.I. (1996). Repetition of previously novel melodies sometimes increases both remember and know responses in recognition memory. *Psychonomic Bulletin and Review, 3*, 366-371. doi: 10.3758/BF03210762

Gaudreau, D., & Peretz, I. (1999). Implicit and explicit memory for music in old and young adults. *Brain and Cognition, 40*, 126-129.

Gibson, J.J. (1977). The theory of affordances. In R. Shaw & B. J. (Eds.), *Perceiving, acting, and knowing: Towards an ecological psychology* (pp. 127-143). Hillsdale, New Jersey: Lawrence Erlabum Associates.

Gibson, J.J. (1978). The ecological approach to the visual perception of pictures. *Leonardo, 11*, 227-235.

Goldstein, E. (2013). *Sensation and perception* (9th Edition ed.). Belmont, CA, United States: Cengage Learning.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology,* 49, 585-612, Doi: 10.1146/annurev.psych.49.1.585

Greene, R.L., & Samuel, A.G. (1986). Recency and suffix effects in serial-recall of musical stimuli. *Journal of Experimental Psychology-Learning Memory and Cognition, 12*, 517-524.

Halekoh, U., & Højsgaard, S. (2014a). A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models–the r package pbkrtest. *Journal of Statistical Software, 59*, 1-32.

Halekoh, U., & Højsgaard, S. (2014b). Pbkrtest: Parametric bootstrap and kenward roger based methods for mixed model comparison.

(Version R package version 0.4-0,). Retrieved from URL http: //CRAN.R-project.org/package=pbkrtest.

Halpern, A.R., & Bartlett, J.C. (2010). Memory for melodies. In M. R. Jones, R. R. Fay & A. N. Popper (Eds.), *Music perception* (Vol. 36, pp. 233-258). New York: Springer. doi: 10.1007/978-1-4419-6114-3_8

Halpern, A.R., & Bartlett, J.C. (2011). The persistence of musical memories: A descriptive study of earworms. *Music Perception, 28*, 425-431. doi: 10.1525/Mp.2011.28.1.425

Halpern, A.R., & Bower, G.H. (1982). Musical expertise and melodic structure in memory for musical notation. *American Journal of Psychology, 95*, 31-50. doi: 10.2307/1422658

Halpern, A.R., & Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for music. *Quarterly Journal of Experimental Psychology, 61*, 1371-1384. doi: 10.1080/17470210701508038

Halpern, A.R., & O'Connor, M.G. (2000). Implicit memory for music in alzheimer's disease. *Neuropsychology, 14*, 391-397. doi: 10.1037//0894-4105.14.3.391

Hebert, S., & Peretz, I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory and Cognition, 25*, 518-533. doi: 10.3758/Bf03201127

Herff, S.A., & Czernochowski, D. (in press). How musical expertise shapes the influence of divided attention on melody recognition. *Musicae Scienciae*

Herff, S.A., Dean, R.T., & Olsen, K.N. (submitted). Inter-rater agreement in memory for melody as a measure of listeners’ similarity in music perception. *Psychomusicology: Music, Mind & Brain*.

Herff, S.A., Olsen, K.N., & Dean, R.T. (2017). Resilient memories for melodies: The number of intervening melodies does not influence novel melody recognition. *Quarterly Journal of Experimental Psychology*.

Herff, S.A., Olsen, K.N., Dean, R.T., & Prince, J. (2017). Memory for melodies in unfamiliar tuning systems: Investigating effects of recency and number of intervening items. *Quarterly Journal of Experimental Psychology*.

Herff, S.A., Olsen, K.N., Prince, J., & Dean, R.T. (2017). Interference in memory for pitch-only and rhythm-only sequences. *Musicae Scientiae*. Advanced online publication, doi: 10.1177/1029864917695654

Hintzman, D.L. (1984). Minerva-2 - a simulation-model of human-memory. *Behavior Research Methods Instruments & Computers, 16*, 96-101.

Hintzman, D.L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528-551. doi: 10.1037/0033-295x.95.4.528

Hockley, W.E. (1992). Item versus associative information - further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning Memory and Cognition, 18*, 1321-1330. doi: 10.1037//0278-7393.18.6.1321

Isola, P., Xiao, J., Torralba, A., & Olivia, A. (2011). What makes an image memorable? *Computer vision and pattern recognition (cvpr), 2011 ieee conference on* (pp. 145 - 152). Providence, RI: IEEE. doi: 10.1109/CVPR.2011.5995721

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition, 100*, 33-72. doi: 10.1016/J.Cognition.2005.11.005

Jacobsen, J.H., Stelzer, J., Fritz, T.H., Chetelat, G., La Joie, R., & Turner, R. (2015). Why musical memory can be preserved in advanced alzheimer's disease. *Brain*, 1-13. doi: 10.1093/brain/awv135

Jahnke, J.C. (1963). Serial position effects in immediate serial recall. *Journal of Verbal Learning and Verbal Behavior, 2*, 284–287. doi: 10.1016/S0022-5371(63)80095-X

Jones, M.R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological review, 96*, 459-491.

Judd, C.M., Westfall, J., & Kenny, D.A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*, 54-69. doi: 10.1037/A0028347

Jusczyk, P.W., & Krumhansl, C.L. (1993). Pitch and rhythmic patterns affecting infants sensitivity to musical phrase structure. *Journal of Experimental Psychology-Human Perception and Performance, 19*, 627-640. doi: 10.1037/0096-1523.19.3.627

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*, 2109-2128.

Karnekull, S.C., Jonsson, F.U., Willander, J., Sikstrom, S., & Larsson, M. (2015). Long-term memory for odors: Influences of familiarity and identification across 64 days. *Chemical Senses, 40*, 259-267. doi: 10.1093/Chemse/Bjv003

Kass, R.E., & Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773-795. doi: 10.1080/01621459.1995.10476572

Kenward, M.G., & Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983-997. doi: 10.2307/2533558

Kessler, E.J., Hansen, C., & Shepard, R.N. (1984). Tonal schemata in the perception of music in bali and in the west. *Music Perception*, 131-165.

Knoblich, G., Seigerschmidt, E., Flach, R., & Prinz, W. (2002). Authorship effects in the prediction of handwriting strokes: Evidence for action simulation during action perception. *The Quarterly Journal of Experimental Psychology: Section A, 55*, 1027-1046.

Konkle, T., Brady, T.F., Alvarez, G.A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General, 139*, 558-578. doi: 10.1037/a0019165

Kopiez, R., & Müllensiefen, D. (2011). Auf der suche nach den 'popularitätsfactoren' in den song-melodien des beatles-albums revolvers [in search of popularity factors in the tunies from the album revolver of the beatles]. In S. Meine & N. Noeske (Eds.), *Musi and popularität: Aspecte zu einer kulturgeschichte zwischen 1500 und heute* (pp. 207-225). Münster: Waxmann.

Krause, A., North, A., & Hewitt, L. (2014). Music selection behaviors in everyday listening. *Journal of Broadcasting & Electronic Media, 58*, 306-323. doi: 10.1080/08838151.2014.906437

Kroll, N.E.A., Knight, R.T., Metcalfe, J., Wolf, E.S., & Tulving, E. (1996). Cohesion failure as a source of memory illusions. *Journal of Memory and Language, 35*, 176-196. doi: 10.1006/jmla.1996.0010

Krumhansl, C.L. (1991). Music psychology - tonal structures in perception and memory. *Annual Review of Psychology, 42*, 277-303. doi: 10.1146/Annurev.Ps.42.020191.001425

Krumhansl, C.L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin, 126*, 159-179. doi: 10.1037//0033-2909.126.1.159

Krumhansl, C.L., & Shepard, R.N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of experimental psychology: Human Perception and Performance, 5*, 579-594. doi: 10.1037/0096-1523.5.4.579

Kruschke, J.K. (2010). Null hypothesis significance testing *Doing bayesian data analysis: A tutorial introduction with r* (pp. 265-287): Academic Press.

Kruschke, J.K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science, 6*, 299-312. doi: 10.1177/1745691611406925

Kruschke, J.K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General, 142*, 573-603. doi: 10.1037/a0029146

Lange, K., & Czernochowski, D. (2013). Does this sound familiar? Effects of timbre change on episodic retrieval of novel melodies. *Acta Psychologica, 143*, 136-145. doi: 10.1016/j.actpsy.2013.03.003

Levitin, D.J. (1994). Absolute memory for musical pitch - evidence from the production of learned melodies. *Perception and Psychophysics, 56*, 414-423. doi: 10.3758/Bf03206733

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*. Thousand Oaks, CA: Sage.

London, J. (2012). *Hearing in time: Psychological aspects of musical meter*. New York: Oxford University Press.

Malmberg, K.J., & Annis, J. (2012). On the relationship between memory and perception: Sequential dependencies in recognition memory testing. *Journal of Experimental Psychology: General, 141*, 233-259. doi: 10.1037/a0025277

Margulis, E.H. (2005). A model of melodic expectation. *Music Perception: An Interdisciplinary Journal, 22*, 663-714.

Margulis, E.H. (2012). Musical repetition detection across multiple exposures. *Music Perception, 29*, 377-385. doi: 10.1525/Mp.2012.29.4.377

McAuley, J.D., Stevens, C., & Humphreys, M.S. (2004). Play it again: Did this melody occur more frequently or was it heard more recently? The role of stimulus familiarity in episodic recognition of music. *Acta Psychologica, 116*, 93-108. doi: 10.1016/J.Actpsy.2004.02.001

McClelland, J.L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105*, 724-760. doi: 10.1037/0033-295x.105.4.734-760

Milne, A.J. (2013). *A computational model of the cognition of tonality*. The Open University.

Milne, A.J., Bulger, D., Herff, S.A., & Sethares, W.A. (2015). Perfect balance: A novel principle for the construction of musical scales and meters *Mathematics and computation in music* (pp. 97-108). London: Springer.

Milne, A.J., & Holland, S. (2016). Empirically testing tonnetz, voice-leading, and spectral models of perceived triadic distance. *Journal of Mathematics and Music*, 1-27.

Milne, A.J., Laney, R., & Sharp, D.B. (2015). A spectral pitch class model of the probe tone data and scalic tonality. *Music Perception, 32*, 364-393. doi: 10.1525/Mp.2015.32.4.364

Milne, A.J., Laney, R., & Sharp, D.B. (2016). Testing a spectral model of tonal affinity with microtonal melodies and inharmonic spectra. *Musicae Scientiae, 20*, 465-494.

Milne, A.J., Sethares, W., & Plamondon, J. (2007). Isomorphic controllers and dynamic tuning: Invariant fingering over a tuning continuum. *Computer Music Journal, 31*, 15-32. doi: 10.1162/comj.2007.31.4.15

Milne, A.J., Sethares, W.A., Laney, R., & Sharp, D.B. (2011). Modelling the similarity of pitch collections with expectation tensors. *Journal of Mathematics and Music, 5*, 1-20.

Milne, A.J., Herff, S.A., Bulger, D., Sethares, W.A., Dean, R.T. (2016). XronoMorph: Algorichmic Generation of Perfectly Balanced and Well-Formed Rhythms. *Proceedings of the 2016 international Conference on New Interfaces for Musical Expression*, Brisbane, Australia

Moelants, D. (2002). *Preferred tempo reconsidered*. Paper presented at the Proceedings of the 7th international conference on music perception and cognition.

Morrison, S.J., Demorest, S.M., Aylward, E.H., Cramer, S.C., & Maravilla, K.R. (2003). Fmri investigation of cross-cultural music comprehension. *Neuroimage, 20*, 378-384.

Müllensiefen, D. (2009). Fantastic: Feature analysis technology accessing statistics (in a corpus): Technical report.

Müllensiefen, D., & Halpern, A.R. (2014). The role of features and context in recognition of novel melodies. *Music Perception*, 418-435. doi: 10.1525/MP.2014.31.5.418.

Nathoo, F.S., & Masson, M.E.J. (2016). Bayesian alternatives to null-hypothesis significance testing for repeated-measures designs. *Journal of Mathematical Psychology*, 144-157. doi: 10.1016/j.jmp.2015.03.003

Nettl, B. (1977). On the question of universals. *The world of music, 19*, 2-7.

Nickerson, R.S. (1965). Short-term-memory for complex meaningful visual configurations - a demonstration of capacity. *Canadian Journal of Psychology, 19*, 155-160. doi: 10.1037/h0082899

Norman, D.A. (2013). *Models of human memory*. New York: Elsevier.

O'Connor, B. (2000). Spss and sas programs for determining the number of components using parallel analysis and velicer's map test. *Behavior Research Methods, Instrumentation, and Computers, 32*, 396-402. doi: 10.3758/Bf03200807

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: A computational implementation of the time-based resource-sharing theory. *Psychonomic Bulletin & Review, 18*, 10-45. doi: 10.3758/S13423-010-0020-6

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review, 19*, 779-819. doi: 10.3758/S13423-012-0272-4

Ollen, J.E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings*. The Ohio State University.

Olson, G.M. (1969). Learning and retention in a continuous recognition task. *Journal of Experimental Psychology, 81*, 381-384. doi: 10.1037/h0027756

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review, 76*, 241-263. doi: 10.1037/h0027272

Palmer, C., & Krumhansl, C.L. (1987). Pitch and temporal contributions to musical phrase perception - effects of harmony, performance timing, and familiarity. *Perception and Psychophysics, 41*, 505-518. doi: 10.3758/Bf03210485

Parncutt, R., & Cohen, A.J. (1995). Identification of microtonal melodies - effects of scale-step size, serial order, and training. *Perception & Psychophysics, 57*, 835-846. doi: 10.3758/Bf03206799

Pearce, M.T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition*: City University London.

Pearce, M.T. (2014). Idyom project. from https://code.soundsoftware.ac.uk/projects/idyom-project

Peretz, I., & Gaudreau, D. (1998). Exposure effects on music preference and recognition. *Memory and Cognition, 26*, 884-902. doi: 10.3758/BF03201171

Peterson, L.R., & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193-198. doi: 10.1037/H0049234

Plantinga, J., & Trainor, L.J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition, 98*, 1-11. doi: 10.1016/J.Cognition.2004.09.008

Poon, L.W., & Fozard, J.L. (1980). Age and word-frequency effects in continuous recognition memory. *Journals of Gerontology, 35*, 77-86. doi: 10.1093/geronj/35.1.77

Posner, M.I. (1966). Components of skilled performance. *Science, 152*, 1712-&. doi: 10.1126/science.152.3730.1712

Prince, J.B. (2014). Contributions of pitch contour, tonality, rhythm, and meter to melodic similarity. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 2319.

Psychology Software Tools, I. (2012). [e-prime 2.0]. from Retrieved from http://www.pstnet.com

R-Core-Team. (2013). R: A language and enviornment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rakover, S.S., & Cahlon, B. (2001). *Face recognition: Cognitive and computational processes* (Vol. 31). Amsterdam: John Benjamins Publishing.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton Jr, C. (2012a). Word perception i: Some basic issues and methods *Psychology of reading* (pp. 49-88): Psychology Press.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton Jr, C. (2012b). Word perception ii: Word identification in text *Psychology of reading* (pp. 135-162): Psychology Press.

Roberts, L.A. (1986). Modality and suffix effects in memory for melodic and harmonic musical materials. *Cognitive Psychology, 18*, 123-157. doi: Doi 10.1016/0010-0285(86)90010-1

Rodrigues, A.C., Loureiro, M., & Caramelli, P. (2014). Visual memory in musicians and non-musicians. *Frontiers in Human Neuroscience, 8*, 424:421-410. doi: 10.3389/Fnhum.2014.00424

Rousseeuw, P.J. (1987). Silhouettes - a graphical aid to the interpretation and validation of cluster-analysis. *Journal of Computational and Applied Mathematics, 20*, 53-65. doi: 10.1016/0377-0427(87)90125-7

Sachs, J.S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2*, 437-442. doi: 10.3758/Bf03208784

Sadeh, T., Ozubko, J.D., Winocur, G., & Moscovitch, M. (2014). How we forget may depend on how we remember. *Trends in Cognitive Sciences, 18*, 26-36. doi: 10.1016/J.Tics.2013.10.008

Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27-52. doi: 10.1016/S0010-0277(98)00075-4

Samson, S., Dellacherie, D., & Platel, H. (2009). Emotional power of music in patients with memory disorders. *Annals of the New York Academy of Sciences, 1169*, 245-255.

Sapp, C.S. (2005). *Online database of scores in the humdrum file format*. Paper presented at the ISMIR.

Schaal, N.K., Banissy, M.J., & Lange, K. (2015). The rhythm span task: Comparing memory capacity for musical rhythms in musicians and non-musicians. *Journal of New Music Research, 44*, 3-10. doi: 10.1080/09298215.2014.937724

Schaal, N.K., Javadi, A., Halpern, A.R., Pollok, B., & Banissy, M.J. (2015). Right parietal cortex mediates recognition memory for melodies. *European Journal of Neuroscience*, 1-7. doi: 10.1111/ejn.12943

Schellenberg, E.G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition, 58*, 75-125. doi: 10.1016/0010-0277(95)00665-6

Schellenberg, E.G., & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Memory and Cognition, 43*, 1021-1031. doi: 10.3758/s13421-015-0519-1

Schellenberg, E.G., Stalinski, S.M., & Marks, B.M. (2014). Memory for surface features of unfamiliar melodies: Independent effects of changes in pitch and tempo. *Psychological research, 78*, 84-95. doi: 10.1007/s00426-013-0483-y

Schellenberg, E.G., & Weiss, M.W. (2013). Music and cognitive abilities. *Psychology of Music, Third Edition*, 499-550. doi: 10.1016/B978-0-12-381460-9.00012-2

Schmuckler, M.A., & Boltz, M.G. (1994). Harmonic and rhythmic influences on musical expectancy. *Perception & Psychophysics, 56*, 313-325. doi: 10.3758/Bf03209765

Schneider, A. (1997). "Verschmelzung", tonal fusion, and consonance: Carl stumpf revisited. In M. Leman (Ed.), *Music, gestalt, and computing: Studies in cognitive and systematic musicology*. New York: Springer.

Schon, D., & Francois, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in Psychology, 2*, 167:161-169. doi: 10.3389/fpsyg.2011.00167

Schulkind, M.D. (2009). Is memory for music special? *Neurosciences and Music III: Disorders and Plasticity, 1169*, 216-224. doi: 10.1111/J.1749-6632.2009.04546.X

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461-464.

Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R., Seither, S., & Lütkenhöner, B. (2007). Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 743-751. doi: 10.1037/0096-1523.33.3.743

Shepard, R.N. (1967). Recognition memory for words, sentences, and pictures. *Journal of verbal Learning and verbal Behavior, 6*, 156-163. doi: 10.1016/S0022-5371(67)80067-7

Shepard, R.N., & Teghtsoonian, M. (1961). Retention of information under conditions approaching a steady-state. *Journal of Experimental Psychology, 62*, 302-309. doi: 10.1037/H0048606

Shiffrin, R.M., & Atkinson, R.C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review, 76*, 179-193. doi: 10.1037/H0027277

Shiffrin, R.M., & Steyvers, M. (1997). A model for recognition memory: Rem-retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*, 145-166. doi: 10.3758/BF03209391

Simons, D.J., & Levin, D.T. (1997). Change blindness. *Trends in cognitive sciences, 1*, 261-267. doi: 10.1016/S1364-6613(97)01080-2

Smith, S.M. (1985). Background music and context-dependent memory. *American Journal of Psychology, 98*, 591-603. doi: 10.2307/1422512

Snodgrass, J.G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology-General, 117*, 34-50. doi: 10.1037//0096-3445.117.1.34

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25*, 207-222. doi: 10.1080/14640747308400340

Stevens, C.J. (2012). Music perception and cognition: A review of recent cross-cultural research. *Top Cogn Sci, 4*, 653-667. doi: 10.1111/j.1756-8765.2012.01215.x

Stevens, C.J. (2015). Is memory for music special? *Memory Studies, 8*, 263-266. doi: 10.1177/1750698015584873

Stevens, C.J., Tardieu, J., Dunbar-Hall, P., Best, C.T., & Tillmann, B. (2013). Expectations in culturally unfamiliar music: Influences of proximal and distal cues and timbral characteristics. *Frontiers in psychology, 4*, 789:781-713. doi: 10.3389/fpsyg.2013.00789

Stroop, J.R. (1992). Studies of interference in serial verbal reactions (reprinted from journal experimental-psychology, vol 18, pg 643-662, 1935). *Journal of Experimental Psychology-General, 121*, 15-23.

Tabachnick, B.G., & Fidell, L.S. (2012). *Using multivariate statistics* (6th Edition ed.). Essex: Pearson Education.

Taruffi, L., Allen, R., Downing, J., & Heaton, P. (2017). Individual differences in music-perceived emotions. *Music Perception: An Interdisciplinary Journal, 34*, 253-266. doi: 10.1525/mp.2017.34.3.253

Tillmann, B., & Dowling, W.J. (2007). Memory decreases for prose, but not for poetry. *Memory and Cognition, 35*, 628-639. doi: 10.3758/BF03193301

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1131-1142. doi: 10.1037/0278-7393.30.5.1131

Tse, P.U., & Cavanagh, P. (2000). Chinese and americans see opposite apparent motions in a chinese character. *Cognition, 74*, B27-B32. doi: 10.1016/S0010-0277(99)00065-7

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & D. W. (Eds.), *Organization memory* (pp. 381-403). New York: Academic Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology-Psychologie Canadienne, 26*, 1-12. doi: 10.1037/H0080017

Vanstone, A.D., Cuddy, L.L., Duffin, J.M., & Alexander, E. (2009). Exceptional preservation of memory for tunes and lyrics: Case studies of amusia, profound deafness, and alzheimer's disease. *Annals of the New York Academy of Sciences, 1169*, 291-294. doi: 10.1111/j.1749-6632.2009.04763.x

Vicary, S.A., Robbins, R.A., Calvo-Merino, B., & Stevens, C.J. (2014). Recognition of dance-like actions: Memory for static posture or dynamic movement? *Memory and Cognition, 42*, 755-767. doi: 10.3758/S13421-014-0395-0

Wagenmakers, E.J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*, 779-804. doi: 10.3758/Bf03194105

Weiss, M.W., Vanzella, P., Schellenberg, E.G., & Trehub, S.E. (2015). Pianists exhibit enhanced memory for vocal melodies but not piano melodies. *The Quarterly Journal of Experimental Psychology, 68*, 866-877. doi: 10.1080/17470218.2015.1020818

White, B.W. (1960). Recognition of distorted melodies. *American Journal of Psychology, 73*, 100-107. doi: 10.2307/1419120

Wiley, R.W., Wilson, C., & Rapp, B. (2016). The effects of alphabet and expertise on letter perception. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 1186-1203. doi: 10.1037/xhp0000213

Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics, 9*, 60-62. doi: 10.2307/2957648

Williams, T.I. (2015). The classification of involuntary musical imagery: The case for earworms. *Psychomusicology: Music, Mind, and Brain, 25*, 5-13. doi: 10.1037/pmu0000082

Winold, A. (1975). Rhythm in twentieth-century music. In G. Wittlich (Ed.), *Aspects of twentieth-century music* (pp. 208-269). Englewood Cliffs, New Jersey: Prentice-Hall.

Wolfe, J.M., Kluender, K.R., Levi, D.M., Bartoshuk, L.M., Herz, R.S., Klatzky, R., . . . Merfeld, D.M. (2012). *Chapter 4: Perceiving and recognizing objects* (Fourth Edition ed.). Sunderland: Sinauer Associates, Inc.

Yonelinas, A.P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517. doi: 10.1006/Jmla.2002.2864

Yonelinas, A.P., Kroll, N.E., Dobbins, I.G., & Soltani, M. (1999). Recognition memory for faces: When familiarity supports associative recognition judgments. *Psychonomic Bulletin and Review, 6*, 654-661. doi: 10.3758/BF03212975

Zatorre, R.J., Chen, J.L., & Penhune, V.B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience, 8*, 547-558. doi: 10.1038/nrn2152

Appendix

Appendix A: Representative Examples of Melodies used in Experiments 1-4

Experiments 1 and 2

Experiments 3 and 4

Figure A.0.1: Representative Examples of the Melodies used in Experiments 1 to 4

Appendix B: Creating and Validating a Corpus of Pre-existing yet Unfamiliar Melodies in Experiments 3 and 4

Experiment 1 and 2 used a corpus of newly composed melodies and demonstrated that the number of intervening melodies do not influence melody recognition. However, this result is potentially limited to the melody corpus used in these experiments. Experiment 3 served as the basis of an extensive stimulus selection procedure to extend these results to a corpus of pre-existing but unfamiliar melodies. A very large set of European folk songs was mathematically analyzed and perceptually tested in an experiment in order to create a heterogeneous corpus of melodies that convincingly represent very common melodies, but are unfamiliar for Australian listeners. The general procedure followed a protocol of placing all melodies in a multidimensional space, identifying clusters of similar melodies, randomly drawing melodies in proportion to cluster size from the clusters, and perceptual testing of the melodies.

Originating Melody Corpus and Musical Feature Analysis

The melody corpora used in this experiment derives from the European folk song portion of the KernScores data base (Sapp, 2005) (through http://kern.ccarh.org/cgi-bin/browse?l=essen/europa). It consists of 8,397 melodies and was chosen to due to easy accessibility, its monophonic midi format, and its relative unfamiliarity to Australian listeners. Melodies longer than 15 seconds were discarded using a custom Max MSP patch (Cycling74, 2014). The MIDI Toolbox (Eerola & Toiviainen, 2004a; 2004b, p. 96) for MatLab, FANTASTIC

(Müllensiefen, 2009, p. 37), for R (R-Core-Team, 2013) were used to analyze the remaining 2194 melodies regarding their musical features. Melconv, a Windows command line program by Frieler, was used to convert the files in the mcsv format required by FANTASTIC. A total of 39 files needed to be deleted, as they were corrupted and not readable by FANTASTIC. A summary of the 2155 remaining melodies' number of notes, pitch range in semitones, and duration can be seen in *Table B.0.1*.

Table B.0.1: Summary of the 2155 melodies analyzed.

	Number of Notes	Pitch range	Duration (s)
Mean	35.69	7.96	10.95
SD	9.58	1.84	2.41

For an overview of the 100+ musical features analyzed, please refer to the documentation of the MIDI Toolbox (Eerola & Toiviainen, 2004a)(p.96) and FANTASTIC (Müllensiefen, 2009)(p 37). Furthermore, several additional musical features analyses were self-implemented in MatLab.

Maximum pitch autocorrelation. A melody is a time series, thus its autocorrelation can be measured (Dean, Bailes, & Dunsmuir, 2014; Dean & Dunsmuir, 2015). Graphically, this feature duplicates a melody and measures the correlation between the two duplicates for every possible lag. The highest pitch autocorrelation is then returned. This feature can be seen as a rough estimate of repetitiveness of a melody.

Dean Tonality. In addition to the measurements of tonality provided by FANTASTIC and the MIDI Toolbox, we implemented an estimate of 'tonalness' as suggested in (Dean, Bailes, & Drummond, 2014). Basically, this form of tonality is defined as the ratio of tonal to non-tonal intervals where intervals of minor seconds, augmented fourth, and major seventh are considered non-tonal.

Pitch class balance and evenness. Rhythms and scales can be described as coordinates around a circle (e.g., in form of a complex vector) (Milne, Bulger, Herff, Sethares, 2015; Milne, Herff, Bulger, Sethares, Dean, 2016). The zeroth and first coefficient of the discrete fourier transformation of this representation can be used to calculate the balance and evenness, respectively. Balance describes the extent to which a scale or rhythm resembles a perfectly balanced rhythm, that is one whose geometrical representation has its center of mass in the middle of the circle. Evenness describes the extent to which a scale or rhythms resembles perfect evenness, that is the geometrical representation has identical step sizes between all events (Milne, Bulger, et al., 2015).

Stimulus Reduction, Dimension Reduction, and Cluster Analysis

An amount of 2155 melodies is too many to present to one participant in a behavioural experiment. Therefore, data pre-processing was conducted to select a representative subsample. To create a representative sample of the 2155 melodies, the melodies were placed in an *n*-dimensional space, where *n* is the number of musical features, in order to identify clusters of similar melodies that can be used as the basis of a random stimulus selection. Due to the large number of musical features computed (over a hundred, some containing co-linearity), a principal component analysis was used in order to reduce the dimensions. A 2500 data sets Monte Carlo Simulation with a 95% confidence level was calculated to estimate the number of underlying components. A permutation approach was chosen, as there is no reason to believe that all musical features of the melodies follow a normal distribution. This approach generates numerous permutations of the given data, runs a principal component analysis for each of them, and compares them to the principal components of the real data. This allows identification of the number of statistically significant principal components. The simulation was conducted using a syntax provided by O'Connor (2000).

The first 21 components provided a significant result. This 21-component solution showed a Kaiser-Meyer-Olkin measure of .796 (*Bartlett's sphericity* < .001 and a maximal correlation between two components of .244. Literature suggests that orthogonal rotations can be used when no correlation is bigger than .32 (Tabachnick & Fidell, 2012) p. 646) but we chose the conservative approach and used an oblique rotation. A solution converged after 55 iterations using direct Oblimin Rotation (Clarkson & Jennrich, 1988). Single factor interpretability was not required for the purpose of this stimulus selection procedure. The score on every principal component was calculated for every melody in order to cluster the melodies in this new reduced dimension space.

Using any kind of Euclidean distance resulted in one big cluster, regardless of the linkage method or whether Akaike's Information criterion or Schwarz's Bayesian criterion was used. When reduction in Log-likelihood was implemented as distance measurement, a nine-cluster solution with fairly low separation (average silhouette measure of cohesion and separation = .01) and a smallest to biggest cluster ratio of 2.587 emerged. This result is not surprising as the corpus consists of purely European Folk songs, predominantly of German derivation. The statistics would allow us to treat the corpus as one big cluster. However, in order to produce a representative subsample, melodies were randomly drawn out of each cluster.

The number of melodies drawn was determined by the aim of the study. In this case, 110 melodies were drawn, later to be used in the behavioural Experiments 3 and 4. This number was identified to satisfy the following requirements. Firstly, being twice the amount of melodies used in Experiment 1 and 2, the remaining corpus would still be big enough to be useful for further experiments in case some melodies were identified to be perceptually familiar. Secondly, 110 melodies can easily be tested in an hour or less, thus making the experiment more tolerable for participants. Thirdly, relative cluster size determined the amount of melodies drawn from each cluster and 110 was a number that allows this procedure to resolve. A summary of the 110 melodies' properties can be seen in *Table B.0.2*. When one compares *Table B.0.1* and *Table B.0.2*, it is clear that the subsample reasonably represents the underlying corpus.

Table B.0.2: Summary of the 110 melodies drawn from the 9 underlying clusters in the corpus.

	Number of Notes	Pitch range	Duration (s)
Mean	35.26	7.93	10.86
SD	10.81	2.07	2.27

As seen in *Table B.0.3*, the 21 underlying components that were shown to be significant by the principal component analysis conducted earlier, explained 70% of the variance in all melodies.

Table B.0.3: Percentage variance explained by 21 components identified to be significant by the Monte Carlo simulation.

Component	% of variance explained	Cumulative %
1	11.921	11.921
2	9.250	21.171
3	6.684	27.598
4	5.826	33.424
5	4.163	37.587
6	3.523	41.109
7	3.157	44.266
8	2.899	47.165
9	2.530	49.696
10	2.264	51.960
11	2.138	54.098
12	2.087	56.185
13	2.085	58.270
14	1.829	60.099
15	1.742	61.841
16	1.664	63.505
17	1.513	65.018
18	1.482	66.500
19	1.393	67.893
20	1.339	69.232
21	1.207	70.439

Perceptual Validation of the New Corpus

A behavioural experiment was designed to identify and exclude melodies in the corpus that evoke atypical perceived familiarity distributions or response tendencies in the target population of Australian listeners.

Participants. Twelve participants (M_{age} = 33, SD_{age} = 12.5) from a heterogeneous musical background at the University of Western Sydney were invited this perceptual experiment.

Procedure. After signing a consent form, participants were instructed that they would hear one melody after another through a pair of headphones. After every melody they were required to indicate 'How familiar is this melody?' on a 100-point click- and slideable vertical visual analog scale (Freyd, 1923). The scale had an 'unfamiliar' label at the bottom and a 'familiar' label at the top and was presented at a length of 10 cm on a MacBook Pro Retina display. The response could be changed during and after stimulus presentation but was logged in as soon as the participant clicked 'Next Melody'. The 'Next Melody' button was only functional after the melody was presented to ensure that participants listened to the entire stimulus. The 110 melodies were presented in 220 trials, as every melody occurred a second time throughout the experiment after a random number of intervening melodies. Note that participants were not informed about the melodies' reoccurrence throughout the experiment. This experiment used the same continuous recognition paradigm with two presentations of each melody as all the other experiments, even though only responses towards the first presentations were analyzed to test initial perceived familiarity of the melodies in the subset. The reason why a paradigm with second occurrences was used is twofold. Firstly, a close replication of the paradigm reduces potentially unpredictable task effects. Secondly, the data of this validation study can later be compared to a planned and conducted experiment that investigates the effect of intervening melodies on perceived familiarity (Experiment 4).

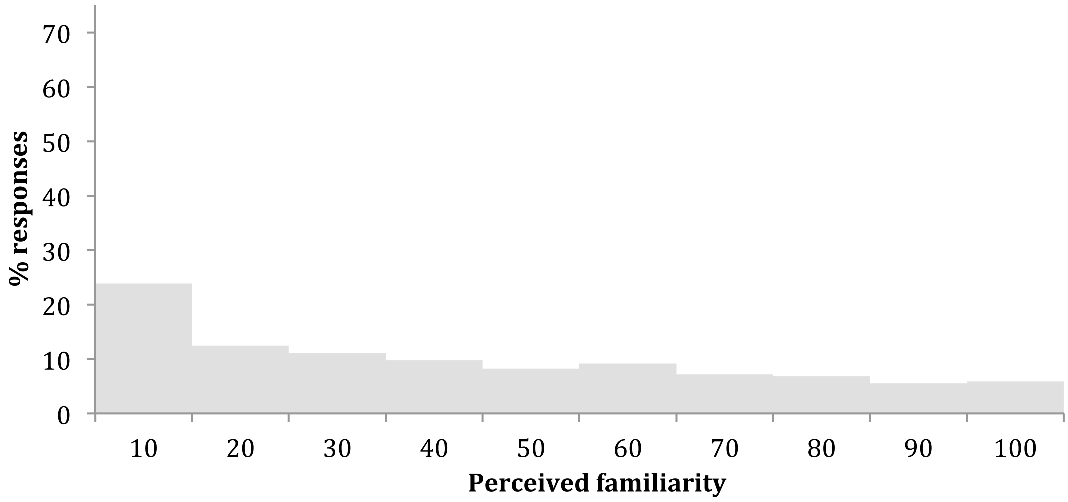
Results

In a first step, melodies that showed a different distribution of perceived familiarity responses to their first occurrence across participants compared to all other melodies were removed from the sample. This step was performed by using a non-parametric Kolmogorov-Smirnov test in order to compare the distribution of every melody with the distribution of all melodies minus that melody. One could argue that the alpha-error level should be adjusted as a result of the huge amount of direct comparisons; however, a beta error would be more severe than an alpha error during this stimuli selection process. This is because the purpose of this analysis was to remove melodies that might be perceived as familiar and the null hypothesis assumed the response distributions to be identical. A conservative approach was chosen by removing all melodies on a .05 alpha-error level.

Using this method, 12 melodies were removed based on the data given by the participants. *Figure B.0.2*a shows the distribution of responses to the first occurrence of all melodies. Compare *Figure B.0.2*b to see the distributions of the 12 removed melodies. A summary of the remaining 98 can be seen in

Table B.0.4.

a) Average response distribution towards unfamiliar melodies



b) Average response distribution towards the 12 significantly different melodies

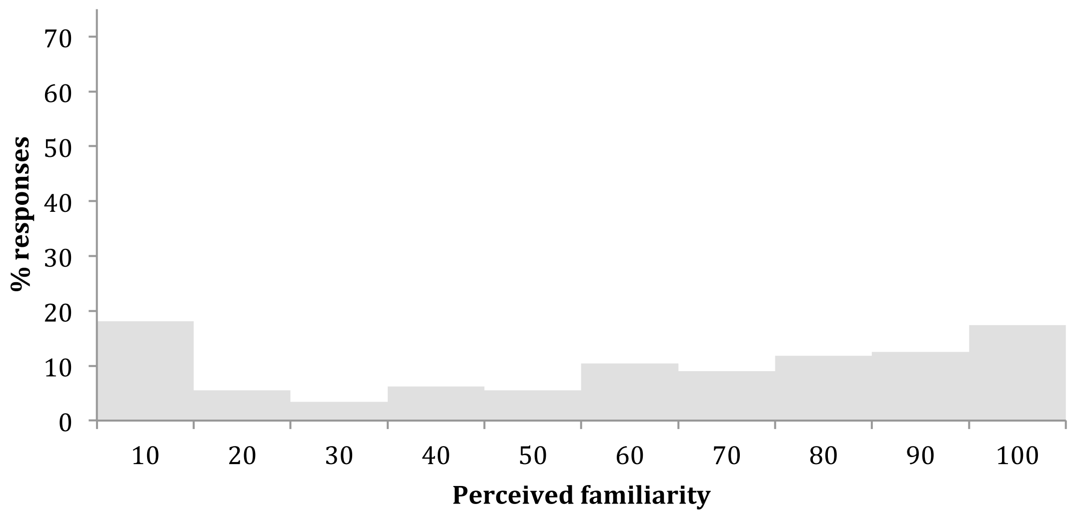


Figure B.0.2: Average response distribution of perceived familiarity towards unfamiliar melodies.

Table B.0.4: A summary of the 98 melodies remaining after the mathematical and perceptual analysis.

	Number of Notes	Pitch range	Duration (s)
Mean	35.26	7.93	10.86
SD	10.81	2.07	2.27

Discussion

An exhaustive procedure for stimulus selection is presented in Experiment 3. A large pre-existing melody corpus that was unfamiliar to the tested cohort was mathematically analyzed. The melodies were clustered on the basis of more than a hundred musical features. Melodies were randomly drawn from the clusters to adequately represent the initial corpus. These melodies were then subjected to a perceptual test. Melodies that showed response distributions of perceived familiarity that were significantly different to the mean distribution of novel melodies were excluded from the corpus. The result is a mathematically and perceptually tested corpus of melodies that are unfamiliar to Australian listeners, while still reflecting common, pre-existing melodies of the Western tonal tradition. This set of melodies will be used in the following experiments to further investigate the number of intervening melodies in melody recognition. All melodies can be found in midi form in the online supplement S1-Stimuli.zip.

Appendix C: Results from the Dynamic Response Tendency Analysis in Chapter 2

Participants' response tendencies can change over the course of an experiment. In the main statistical analyses presented throughout the paper, any shifts in response tendencies were taken into account by building *Dynamic Response Tendency* models that predict the baseline tendency (which is a bias) for each participant to press 'old', and how this tendency changes throughout the experiment. To do this, a generalized mixed effects model was trained to predict 'old' responses on first melody presentations (False Alarms) based on trial number. The resultant *Dynamic Response Tendency* models were then used to predict the tendency (probability) of pressing 'old' on second melody presentations, for each participant and trial. The predictions of the models were then included as a fixed factor in the main statistical analyses. Effectively, the procedure corrects participants' responses with the to-be-expected false alarm rate for that particular trial in a given participant. This is analogous to common methods of adjusting participants' responses depending on false alarm rates, however, here we extend this correction dynamically over the number of trials (Snodgrass & Corwin, 1988). Results of *Dynamic Response Tendency* models are summarized here: Experiments 1, 2, and 4 did not exhibit significant response tendency shifts (all coefficients p-values > .20). However, Experiment 3 did show significant dynamic response tendency shifts over the course of the experiment ($p < .001$, coefficient = 2.59). As can be seen in *Figure C.0.3*, the average probability of producing false alarms increased in Experiment 3 as trial numbers increased. This is likely due to the design of Experiment 3's continuous recognition task, which presented nearly twice as many trials as Experiments 1 and 2.

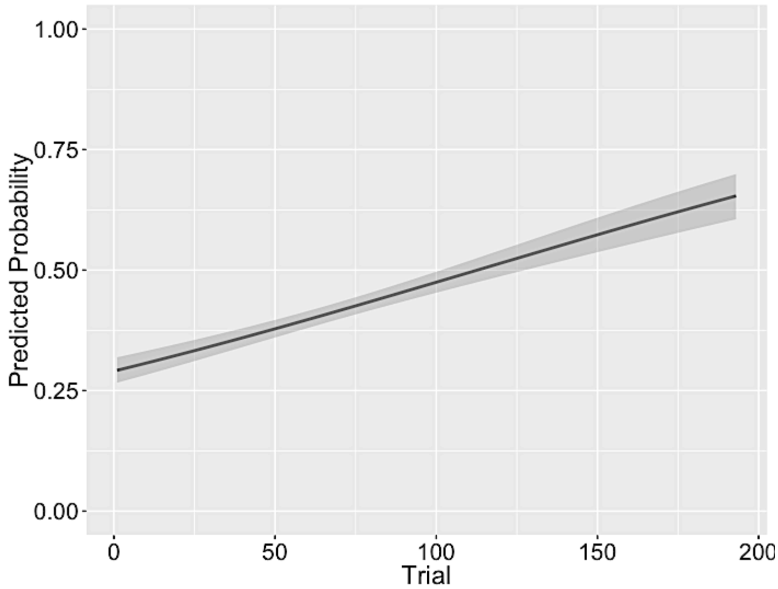


Figure C.0.3: Average dynamic response tendency shift in Experiment 3. The graph shows how the probability of producing a false alarm increased with trial number. The dynamic response tendency shift was controlled in the statistical analyses and implemented as a predictor in the mixed effects models reported in the main Results sections. The grey area around the prediction line represents a 95% confidence interval.

For comparison, *Figure C.0.4* shows uncorrected hit rates in Experiment 1 and *Figure C.0.5* shows uncorrected hit rates of Experiment 3.

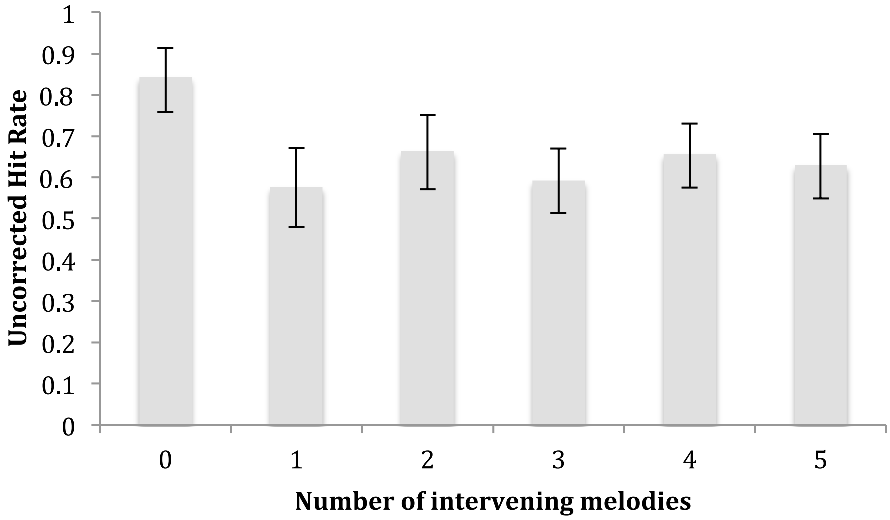


Figure C.0.4: Uncorrected hit rates in Experiment 1. In our analysis, we corrected the hit rates to capture and account for participants' response biases, as well as potential shifts of these response biases over the course of the experiment.

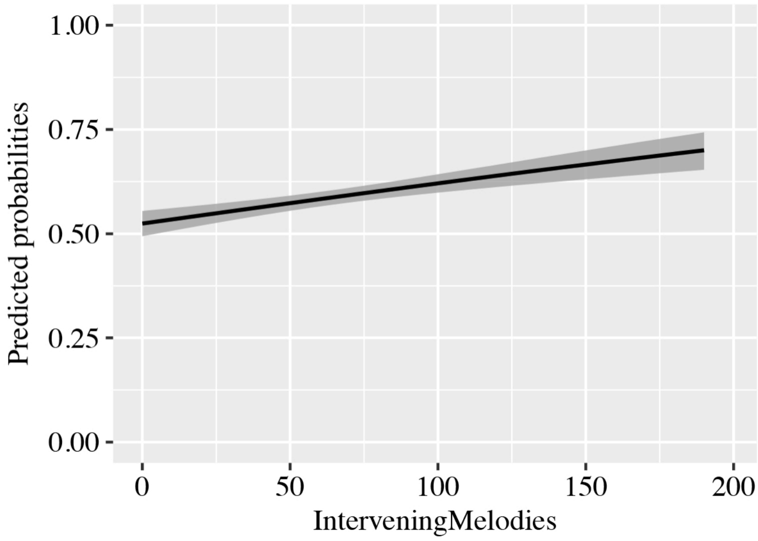


Figure C.0.5: Uncorrected hit rates in Experiment 3. In our analysis, we corrected the hit rates to capture and account for participants' response biases, as well as potential shifts of these response biases over the course of the experiment. When comparing *Figure C.0.5* with *Figure 2.8*, performance appears better in the uncorrected hit rates and performance seems to significantly increase over the course of the experiment. However, this is due to systematic response-tendency shifts (see further detail in Appendix C). When the analysis takes these response-tendency shifts into account, then performance is stable across all intervening item conditions (cf. *Figure 2.8*).

Appendix D: Cosine Dissimilarity

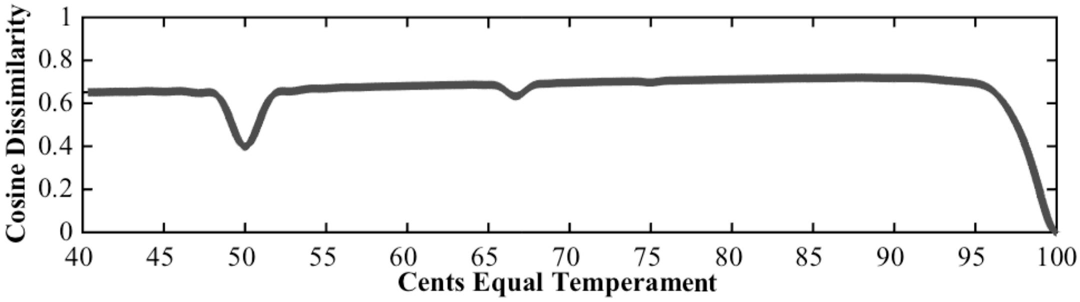


Figure D.0.6: Cosine Dissimilarity between 12-Tone-Equal-Termperament (or 100 Cents-Equal-Temperament) tuning system familiar to most Western listeners and different Cents-Equal-Temperament tuning systems. 88.08-CET shows the maximum dissimilarity with 12-TET.

Appendix E: Pilot Data of a High Musical Expertise Group

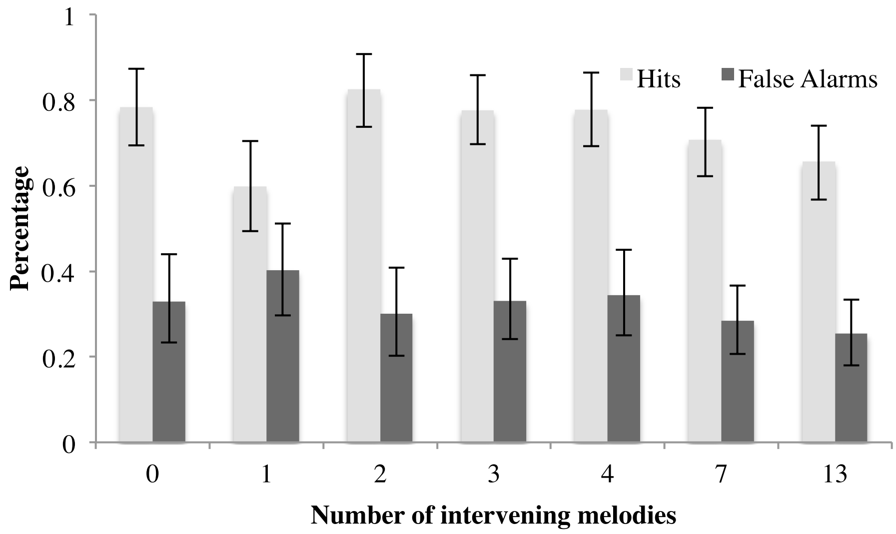


Figure E.0.7: Average hit rates and false alarm rates for a high musical expertise group (N = 14, more than 8 years of formal training) in Experiment 5 (0, 1, 2, 3, 4, 7, and 13 intervening melody conditions). In this experiment, a high musical expertise group performed a continuous recognition task with melodies in an unfamiliar tuning system. Error bars represent a 95% confidence interval.

Appendix F: General Participant Information Sheet

MARCS Institute
University of Western Sydney
Locked Bag 1797
Penrith NSW 2751
Australia
Telephone : 02 9772 6107
e-mail : y.leung@uws.edu.au

University of Western Sydney
Bringing knowledge to life

Participant Information Sheet (General)

Project Title: A systematic investigation of memory for melodies.

Project Summary: This study is interested in how non-musicians store music in their memory. It will examine perceived familiarity and memory of musical stimuli in dependency to musical features.

You are invited to participate in a study conducted by Steffen Herff, PhD Candidate, MARCS Institute, under the Supervision of Prof Roger Dean, Research Professor, MARCS Institute.

How is this study being paid for?
The study is being sponsored by the MARCS Institute.

What will I be asked to do?
You will be asked to perform a computer task which involves listening to some melodies through a pair of headphones, and responding to them by pressing keys on a computer keyboard. You will also be asked to complete some questionnaires about your demographic information and music background.

How much of my time will I need to give?
This experiment will take approximately 60 minutes, with a short break in the middle.

What specific benefits will I receive for participating?
The study will possibly provide you some knowledge about psychological experiments and some insights to your ability to perceive different types of musical sounds. You will also gain partial course credit points from your participation in the study.

Will the study involve any discomfort for me? If so, what will you do to rectify it.
There will be no discomfort for you. You will be listening to musical sounds at a comfortable loudness level.

How do you intend on publishing the results.
All aspects of the study, including results, will be confidential and only the researchers will have access to information on participants. Results will be disseminated in peer-reviewed journal articles such as 'Music Perception' and national and international conference presentations.

There are a number of government initiatives in place to centrally store research data and to make it available for further research. For more information, see http://www.ands.org.au/ and http://www.rdsi.uq.edu.au/about. Regardless of whether the information you supply or about you is stored centrally or not, it will be stored securely and it will be de-identified before it is made to available to any other researcher.

Can I withdraw from the study?

Participation is entirely voluntary: and you are not obliged to be involved. If you do participate, you can withdraw at any time without giving any reason.

If you do choose to withdraw, any information that you have supplied will be discarded.

Can I tell other people about the study?

Yes, you can tell other people about the study by providing them with the chief investigator's contact details. They can contact the chief investigator to discuss their participation in the research project and obtain an information sheet.

What if I require further information?

Please contact Steffen Herff should you wish to discuss the research further before deciding whether or not to participate.

Steffen A. Herff
PhD Candidate, MARCS Institute
Ph. +61 2 9772 6801

What if I have a complaint?

This study has been approved by the University of Western Sydney Human Research Ethics Committee. The Approval number is _____H10847___.

If you have any complaints or reservations about the ethical conduct of this research, you may contact the Ethics Committee through the Office of Research Services on Tel +61 2 4736 0229 Fax +61 2 4736 0013 or email humanethics@uws.edu.au.

Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.

If you agree to participate in this study, you may be asked to sign the Participant Consent Form.

Online Supplement

The Online Supplement S1-Stimuli.zip contains the stimuli of all 10 experiments and can be accessed through:

http://www.dynamictonality.com/memory_for_melody_data/S1_Stimuli.zip