THE UNIVERSITY OF HONG KONG
FACULTY OF EDUCATION

Master of Science in Library and Information Management (MSc[LIM])

A Field Experiment on Music Preference during Learning

Submitted by: LI, Fanjie

Dissertation presented in part-fulfilment of
the requirements of the degree of
Master of Science in Library and Information Management,
The University of Hong Kong

December 2019

Declaration

I hereby declare that this dissertation represents my own work and that it has not been previously submitted to this University or any other institution in application for admission to a degree, diploma or other qualifications.

Li Fanjie

LI Fanjie, December 2019

Abstract

Empirical evidence of how background music benefits or hinders learning becomes the crux of optimizing music recommendation in educational settings. This study aims to explore how background music befits learning through an experiment in naturalistic setting. A one-week field experiment was conducted in participants' own study places. During the experiment, participants were asked to conduct learning sessions with music in the background and collect the tracks they deemed suitable for learning using a novel mobile music app (i.e., Moody App). A set of participant-related, context-related, and music-related data were collected via the pre-experiment survey as well as the logging system and survey system of the music app.

Our findings revealed some general tendencies of learners' music preference in terms of music style, music emotion (i.e., happiness, energy), and the dynamical and timbral characteristics of music. Group-wise difference in music preference was also observed when grouping participants by certain personal factors (e.g., personality, working memory capacity, prior habit of studying with background music). In regard to the association between music characteristics and listeners' learning experience, both affective dimensions of music emotion (i.e., happiness, energy) were found to significantly correlate with participants' learning engagement. Though no overall effect of dynamical, rhythmic, and timbral features was observed, the timbre quality of music showed significant effect in certain condition when the potential moderating effect of task load and learners' traits was considered.

This study is expected to provide evidence for understanding the effects of background music on learning, as well as implications for designing music recommendation systems that are capable of intelligently selecting background music for facilitating learning.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Statement of the Problem

Music, often used as a background accompaniment for learning, has been deemed an effective tool for concentration and mood modulation by a considerable number of learners. In fact, the "studying music" has already become one of the most popular tags for music discovery, and thousands of playlists hosted on online music streaming services (e.g., Spotify) are specifically curated for daily learning activities. In spite of the prevalence of studying with music in the background, there is also evidence that the positive effect of music might not be universally true among learners (Furnham & Strbac, 2002; Perham & Currie, 2014).

How background music plays a role in students' learning process has been studied by educators, cognitive psychologists, and neuropsychologists for over eight decades (Hu et al., 2019; Whitely, 1934). Although the positive effect of music has been revealed by a number of studies in the research literature (Angel et al., 2010; Ferreri & Verga, 2016; Hallam et al., 2002), inconsistent and conflicting findings (i.e., detrimental or no effect) have been reported as well (Furnham & Strbac, 2002; Jäncke & Sandmann, 2010; Perham & Currie, 2014).

With respect to these inconsistent and inconclusive findings, one possible explanation suggested by relevant research is that the effect of music on learning might vary across the types of music in the background (Jäncke et al., 2014), the complexity of learning tasks (Kämpfe et al., 2010), and the personal characteristics of learners (Küssner, 2017; Lehmann & Seufert, 2017). To this end, identifying the right piece of music for facilitating learning based on the characteristics of learning task and traits of learners is thus worthy of our attention (Ferreri & Verga, 2016; Kämpfe et al., 2010).

Furthermore, thanks to the recent development of music processing techniques, more fine-grained music features on various music trait dimensions (e.g., dynamics, rhythm, timbre) could be measured in a standardized and objective way through the automatic processing of audio samples. This further provides opportunity for disentangling and anatomizing how each specific music trait dimension plays a role in the interaction between background music and listeners' learning process.

To sum up, though studying with background music is popular among learners, the empirical evidence of how background music benefits or hinders learning is still inadequate, which becomes the crux of optimizing music recommendation in educational settings. Towards the ultimate goal of intelligently selecting background music for learners, further research is needed to provide empirical evidence for predicting preferable study music for individual learners, in view of the particular learning task being performed and the undergoing learning context.

1.2 Purpose Statement

Based on the above discussions, this study, therefore, aims to further probe how background music befits learning in view of the potential moderating effect of the music characteristics, learners' traits, and learners' task load. Furthermore, instead of following the laboratory experimental approach adopted in previous research, a field experiment was conducted so as to achieve a higher level of ecological validity.

1.2.1 Objectives

Overall, as a user study which primarily focuses on music recommendation in educational settings, the main objective of this research is to profile learners' needs for music selection, and, therefore, provide empirical evidence for optimizing the recommendation of music for learners from a user-centric perspective. Specifically, we aim to achieve the following research objectives:

1) To profile the subjective music preference of learners in view of the potential individual difference;
2) To investigate the association between music characteristics and listeners' learning experience (i.e., the perceived enhancing versus distracting effect of music, learning engagement) in view of the potential moderating effect of learners' traits and task load.

1.2.2 Research Questions

Specifically, the research questions for the present study could be summarized as follows:

RQ1: What kind of music would be deemed suitable for learning? Would the music preference of learners vary across different user groups?

As a user study aiming to provide empirical evidence for background music selection for learners, we first profiled the characteristics of preferable study music against those failed to match learners' preference. Moreover, apart from the general tendency of learners' music preference, we also aim to depict the characteristics of music preferred by different user groups.

RQ2: What kind of music would be deemed as benefiting learning versus impairing learning? Would task load and learners' traits moderate this effect?

Taking one step further, this study also investigated the association between music characteristics and listeners' learning experience to anatomize the effect of each music trait dimension on learning. Particularly, in light of the research literature, the potential moderating effect of learners' traits and task load was considered as well.

1.3 Significance of the Study

This study is expected to deepen our understandings of the interplay between music and learning, enable us to reasonably predict preferable music for individual learners in different learning contexts, and ultimately provide implications for designing music information retrieval (MIR) systems which are capable of intelligently selecting background music and generating playlists for facilitating learning.

Particularly, as specified in chapter 3, this study was expected to further improve the ecological validity of research findings on the current topic by collecting longitudinal data in naturalistic setting and providing a large-scale music pool for participants to explore.

Besides, as one of the early studies which incorporated fine-grained music characteristics into analysis, this study was expected to further anatomize how each specific music trait dimension plays a role in the interaction between background music and listeners' learning process and explore the music aspects (e.g., timbre) under-represented in the current research literature.

Furthermore, by incorporating task load and learners' traits in our research framework, this study was also expected to further our understanding on the association between music characteristics and listeners' learning experience by disentangling the confounding effect of personal and contextual factors.

Practically, our findings could also provide implications for designing personalized and context-aware music recommendation system. With consideration of the potential effect of individual and contextual difference, this study empirically depicted the preferable music for learning in terms of music type, music style, music emotion (i.e., happiness, arousal), as well as dynamical, rhythmic, and timbral characteristics of music.

Chapter 2: Literature Review

2.1 Theoretical Background

Although how background music plays a role in students' learning process has been studied by educators, cognitive psychologists, and neuropsychologists for over eight decades (Hu et al., 2019; Whitely, 1934), the relationships thereof are still inconclusive. Some studies have found that background music is beneficial to various learning tasks, including verbal learning (Ferreri & Verga, 2016), arithmetic problem solving (Hallam et al., 2002), and spatial processing (Angel et al., 2010), etc. Other studies, on the other hand, reported detrimental (Furnham & Strbac, 2002; Perham & Currie, 2014), or no effects (Jäncke & Sandmann, 2010; Kämpfe et al., 2010) of background music on learning.

To explain these heterogeneous findings, two theoretically opposite hypotheses have thus been developed. From the perspective of academic emotion, Husain et al. (2002) suggest a possible explanation (i.e., arousal-mood-hypothesis, AMH) for the beneficial effect of music on learning. They point out that music is powerful in mood modulation and thus can keep learners in a positive mood and help them reach the optimal level of arousal, which, in turn, exerts a positive influence on their learning performance. Relevant studies on the relationships between music, emotion, and learning are found consistent with this hypothesis: For one thing, the effects of music on emotion (i.e., arousal and valence) has already been proved by the previous studies in musical psychology. Generally, upbeat music often increases listeners' level of arousal, whereas slow music often decreases their arousal level (Balch & Lewis, 1996); major-mode music pieces often induce positive mood status (e.g., pleased), while its counterpart (i.e., music in minor mode) is often associated with negative mood status (e.g., depressed) (Webster & Weir, 2005). For another, the effect of emotion on learning performance has also been demonstrated by the previous psychological and educational research. As illustrated by the influential Yerkes-Dodson law (Yerkes & Dodson, 1908), the influence of arousal on learning performance is in conformity with the pattern of "inverted-U", which means the increasing level of arousal will first lead to improved learning performance but exert detrimental effect afterwards once the turning point has been reached. Moreover, according to Schellenberg (2012), the mood valence (i.e., level of happiness) was also found to be correlated with students' performance; i.e., negative and positive emotional status are deemed unfavorable and favorable to cognitive processing respectively. Also closely related to affection, Schnotz and Kürschner (2007)'s research further developed such idea from the motivational aspects. Specifically, based on Schnotz and Kürschner (2007)'s hypothesis, the enjoyableness and energizing effect of music might increase learners' activation and persistence in learning, which, ultimately, exerts a positive effect on their learning performance.

Despite the theoretical contribution of the above affection-related hypothesis, it was still criticized for its inadequacy in explaining the negative influence of studying with music in the background, which gives rise to a disparate theoretical assumption. From the perspective of the cognitive functioning in the learning process, the irrelevant-sound-effect (ISE) stresses that the information-load characteristics of the background music would, nonetheless, increase the cognitive loads of learners, and thus impairs learning (Beaman, 2005; Boyle & Coltheart, 1996). The ISE points out that, the background music could generally be considered as a seductive detail, which would overburden learners' working memory and distract their attention (Kantner, 2009; Rey, 2012). Specifically, as the auditory reception function is intrinsic to our brain, listening to music while learning would inevitably consume the finite cognitive resources of our brain (e.g., working memory, attention) and thus brings extra cognitive burden to learners (Lehmann & Seufert, 2017). In this regard, Sanchez and Wiley (2006) further proposed the assumption that learners with lower working memory capacity are more likely to suffer from the detrimental effect of such seductive details, as one's working memory capacity is closely related to the ability of maintaining attention on the major task and suppressing extraneous task-irrelevant information.

2.2 Related Work

2.2.1 Interactions among Learner, Music, and Learning Task

Based on the above hypotheses, several possible moderators in relation to the effect of background music on learning have been subsequently proposed, including information-load characteristics of music (e.g., tempo, loudness, vocal element) (Jäncke et al., 2014; Kantner, 2009; Kiger, 1989), task complexity (Jäncke & Sandmann, 2010; Kämpfe et al., 2010; Lehmann & Seufert, 2017), and learners' working memory capacity (Christopher & Shelton, 2017; Lehmann & Seufert, 2017), personality traits (Furnham & Strbac, 2002; Küssner, 2017), etc.

One of the early attempts of investigating the effect of information-load characteristics of music on learning could be traced back to Kiger (1989)'s research. Motivated by their music information load hypothesis, Kiger (1989) conducted an experimental study which compared learners' reading comprehension performance in three conditions (i.e., silence and music with high versus low information load). Specifically, the music materials in the latter two conditions were both selected by the researcher based on several music traits criteria (e.g., loudness, tempo, tonal quality): according to their published report, the low-information load piece was characterized by high repetitiveness and small tonal variation, while the selected high information load music was reported to be dissonant and possess high rhythmic and dynamical variation. The results of their study showed that participants who performed the reading task with low information load piece achieved significantly higher reading comprehension score compared to those who read in silence or with high information load music. And the reading comprehension performance was worst in the high information load condition.

Apart from the aforementioned music traits, the effect of vocal element has also been investigated in recent studies. For instance, in view of the irrelevant speech effect (ISE), Kantner (2009) conducted two experiments in different learning scenarios to probe whether learners' serial recall performance vary across vocal and instrumental background music stimuli. The results of the two experiments were both found to be consistent with the ISE hypothesis (i.e., participants listened to vocal music performed significantly worse than those studied with instrumental music). However, another study conducted by Jäncke et al. (2014) on verbal learning revealed a slightly opposite result. Specifically, Jäncke et al. (2014) also adopted a between-subjects design and assigned participants to five experimental conditions (i.e., silence control group and four music conditions: vocal/instrumental × high/low music intensity). Nonetheless, as a non-significant difference in the perceived intensity was observed, they finally decided to merge the conditions on the intensity dimension, resulting in three groups (i.e., control, vocal, instrumental) for the final comparison. The result of Jäncke et al. (2014)'s study showed that there is no significant group difference, and hence they concluded that there might be no influence of the specific type of music (i.e., instrumental or vocal).

Besides, Jäncke and Sandmann (2010) also conducted another research on the effect of tempo and consonance prior to the aforementioned study. Particularly, they used computer-manipulated music excerpts (i.e., the final four music conditions: slow/fast × in-tune/out-of-tune were generated from the same original piece) to control for the potential confounding effect introduced by difference in other music trait dimensions. The results of this study showed no significant group difference in the verbal memory test score. However, event-related desynchronization was found to be associated with both high tempo conditions in their EEG signal analysis. Finally, as consistent with Jäncke and Sandmann (2010)'s finding on the non-significant effect of tempo, Kämpfe et al. (2010)'s meta-analysis also suggest that there might be no overall effect of the speed of music, though they concluded from their reviewed article that the tempo of music seemed be strongly correlated with the quickness of listeners' behaviour. Besides, the meta-analysis also revealed a non-significant overall effect of loudness.

Notwithstanding the impact of the aforementioned research, there are still several limitations and potential gaps to be noted:

1) *Representativeness of music materials (stimuli).* Most of the existing research were based on the researcher-selected music materials and each music condition was mostly represented by a single music piece. However, as mentioned in Schnotz and Kürschner (2007)'s hypothesis, learners' perceived enjoyableness of music was also assumed to play an important role in the interaction between music and learning. To reduce the confounding effect of learners' personal music preference and music perception, a larger music pool might be necessary to further underpin the generalizability of our research findings.
2) *Granularity of Music Traits.* Kämpfe et al. (2010) has pointed out that one important challenge faced by the earlier research is the lack of standard for the comparison and characterization of various traits of music (e.g., tonal quality, timbre, harmony) except for a few relatively standardized measure (i.e., tempo, loudness). Hence, it is not surprising that little research was focused on the effect of the timbre quality of music, except for some preliminary feasibility study such as Matney (2017)'s research on the effect of instrumentation on university students' anxiety reduction. However, thanks to the recent development of music processing techniques, more fine-grained music features on various music trait dimensions (e.g., dynamics, rhythm, timbre) could be measured in a standardized and objective way through the automatic processing of audio samples. It is promising that incorporating the insights from music processing field might generate more fruitful findings.
3) *Integration of personal and contextual factors.* Recalling our previous discussions, one may find that there is still limited body of research investigating the effect of music characteristics on learning in view of the potential moderating effect of various personal and contextual factors. For instance, though the studies concluded that there might be no overall effect of tempo, it is still worth probing if the effect of tempo could be confounded by learners' traits (e.g., working memory capacity) or the learning context (e.g., task complexity). Related studies, which investigated the general effect of learners' traits, etc., were further discussed in the following sections.

As for learners' traits, the experiment results of Lehmann and Seufert (2017)'s research revealed a positive correlation between reading comprehension performance (i.e., recall, comprehension) and subjects' working memory capacity. Consistently, such relationship was also observed in Christopher and Shelton (2017)'s research of which the results suggest that the moderating effect of working memory capacity could help predict the performance of two types of learning tasks (i.e., math-related task and reading comprehension) in the presence of background music. Based on a mini-review, Küssner (2017)'s study suggested that, in contrast to extraverts, introverted learners were more likely to suffer from the detrimental effect of music because of their relatively higher cortical arousal. And another experimental study conducted by Etaugh and Ptasnik (1982) reported that, compared to participants who were unaccustomed to background music, subjects who habitually listened to music while studying achieved better performance in their verbal learning task.

However, unlike the increasing research on the moderating effect of learners' traits, a limit body of research was found to systematically manipulate the characteristics of learning task, even though variables such as task difficulty and task complexity have been frequently proposed as potential moderators in relevant studies (Jäncke & Sandmann, 2010; Kämpfe et al., 2010; Lehmann & Seufert, 2017). One possible reason of this phenomenon might be the difficulty in designing a wide range of naturalistic learning tasks in the laboratory settings. Actually, the limited ecological validity of previous laboratory-based studies has already been deemed to be responsible for the inconclusive findings in the research literature (Kämpfe et al., 2010). With tight schedules of the laboratory experiments, constrained music stimulus, and artificial learning tasks, it could be hard for laboratory experiments to simulate real-life learning scenarios.

2.2.2 Measuring Learning Experience: The State of Flow

Apart from the aforementioned limitations, another issue that worth our attention is the selection of learning outcome measures, especially when designing experiment in naturalistic settings. As shown in the previous discussions, the outcome variables identified by the previous studies were mostly performance-related measures (e.g., accuracy of calculation, vocabulary recall, comprehension test score). However, there is limited feasibility of incorporating such outcome measures in studies primarily based on naturalistic observations, given the potential range of learning tasks that could be covered during the experiment execution.

Furthermore, as hypothesized by Schnotz and Kürschner (2007), it is noteworthy that the influence of music could also be motivation-and-persistence-related, instead of a direct influence on the learning performance per se. Actually, the core idea of the affection-related hypothesis discussed in section 2.1 is closely related to the power of music in enhancing learners' learning experience as well as the underlying relationship between emotion modulation and learners' motivation (reward) system (Schnotz & Kürschner, 2007).

Hence, given the concerns about measuring performance in naturalistic settings and the theoretical motivation discussed above, this study further reviewed the related work on measuring learning experience and, especially, learning engagement.

Csikszentmihalyi (1997)'s flow theory is a well-established theoretical framework for the conceptualization of optimal psychological experience. As a positive psychological state, the state of flow often co-occurs with high performance and high intrinsic motivation (Jackson et al., 2012), and has been adopted in a wide range of areas for investigating engagement and positive experiences, including educational and learning-related studies (Admiraal et al., 2011; Hamari et al., 2016; David J Shernoff & Csikszentmihalyi, 2009; David J. Shernoff et al., 2014).

Phenomenologically speaking, the state of flow depicts a special absorbing experience and is closely associated with high level of task involvement (Csikszentmihalyi, 2014, p. 240; Jackson et al., 2012). In flow theory, such absorbing experience was further broken down to several sub-constructs, including *altered sense of time* and *total concentration on the task at hand*. Particularly, the *total concentration on the task at hand* referred to ones' intense focus on their task being performed at the moment, disregarding any extraneous thoughts and distractions, while the *altered sense of time* generally refers to the distortion of temporal awareness, which is typically experienced as a sense that time seems to pass faster than normal (Csikszentmihalyi, 2014, p. 240).

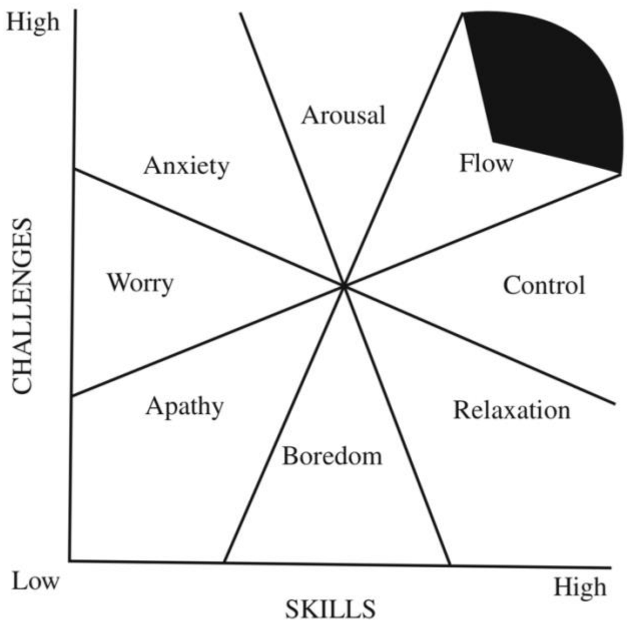


Figure 1. Model the flow state in regard to challenge-skill balance (Csikszentmihalyi, 2014, p. 201).

Moreover, in flow theory, Csikszentmihalyi (1997) also proposed an insightful theorization of the condition of entering the flow state. Specifically, the *challenge-skill balance* was identified as a crucial pre-requisite to flow (Jackson et al., 2012), and has been modeled graphically as shown in Figure 1 (Csikszentmihalyi, 2014, p. 201). As indicated by the graphical model, when learners found the challenge of the task outweigh their skill, they would be aroused. And, if little skill was equipped, anxiety could thus be predicted. To the contrary, when learners found their skill could cope with the task, boredom or relaxation would be predicted.

The challenge-skill hypothesis would be adopted as the theoretical background for investigating the moderating effect of learners' task load (i.e., the mental and temporal demand of the learning task) in section 4.2.2.

2.3 Conceptual Framework

To sum up, in light of the affection-related and cognition-related hypothesis discussed in section 2.1, this study aims to investigate how background music could possibly play a role in both the cognitive and emotional aspects of learning process. Meanwhile, given the limitations of the laboratory studies discussed in section 2.2.1, a naturalistic experiment design was adopted so as to achieve a higher level of ecological validity. Based on the synthesis of the potential moderators proposed in the research literature, the conceptual framework of this study was developed and presented below (Figure 2).

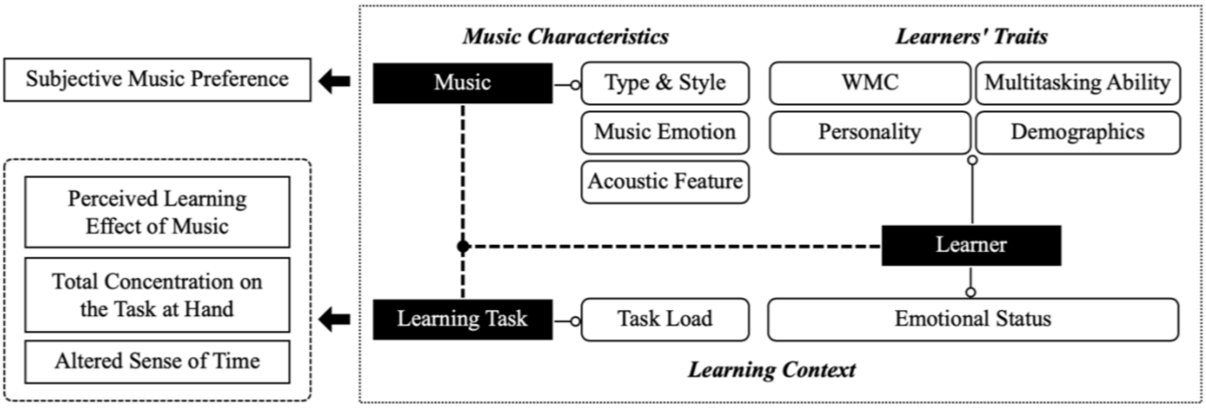


Figure 2. Conceptual framework. WMC = Working memory capacity.

Details on the operationalization and measurement of the variables could be found in section 3.4.

Chapter 3: Methodology

3.1 Overview of Research Design

Overall, as an exploratory study, this research mainly adopted the mixed method approach and an observational and longitudinal research design.

Specifically, to simulate the real-life music discovery experience of learners, a field experiment was conducted in participants' own study places for one week. During the experiment, participants were asked to conduct learning sessions with music in the background and add the tracks they deemed suitable for learning to their personal study music playlists. To facilitate longitudinal data collection in naturalistic settings, a novel mobile music app (i.e., Moody App) was provided to participants for music searching and listening. Participants' searching behavior, learning engagement, and their perceived learning effect of music (i.e., enhance versus distract) were collected by the Moody App via system log files and pop-up surveys.

As specified in the conceptual framework, a set of participant-related, music-related, and context-related data were collected for further analysis. The data analysis strategies were primarily quantitative. Some qualitative insights (e.g., visualization of music metadata, participants' comments in the post-experiment interview) were incorporated as triangulation as well.

Details on the user experiment design, data sources and measures, and data analysis strategies were described in the sections below.

3.2 The Music App

The Moody App is a novel mobile-based music information retrieval system, which supports mood-based music discovery (i.e., music selection based on its happiness and energy), automatic and customizable playlist generation, user activity monitoring and logging, and interactive online survey.

3.2.1 The Music Information Retrieval Subsystem

Moody Music Library

A total of 10K music pieces in the Moody database are freely accessible to the system users, where all tracks are originally obtained from the Jamendo music database under Creative Common licenses (Hu et al., 2017). Specifically, the MP3 audio files were all stored on our web server (over 60 GBs in total), along with a rich set of music metadata (e.g., title, artist, genre, instrumental or vocal) supplied by the Jamendo dataset. On the whole, the Moody music library covers 137 unique genres, including easy listening (694), electronic (3772), country (116), pop (1267), folk (464), jazz (568), classical (509), rock (2079), R&B (180), etc.[1] Particularly, all tracks in the music library were associated with at least one genre tag, and 4994 of them were labeled for being vocal or instrumental (i.e., with or without lyrics).

To facilitate mood-based music discovery, all songs in the Moody database also had a set of pre-computed music emotion metadata on the continuums (-1~1) of two affective dimensions (i.e., happiness, energy). Details on the computation of music emotion could be found in section 3.4.3.

Music Searching and Filtering

In light of the arousal-mood-hypothesis discussed in section 2.1, the Moody App implemented a novel mood-based music discovery feature so that users could find music in accordance with their preferred music mood at the moment. Specifically, users could specify the preferred happiness and energy level of music using the sliders (as shown in Figure 3).

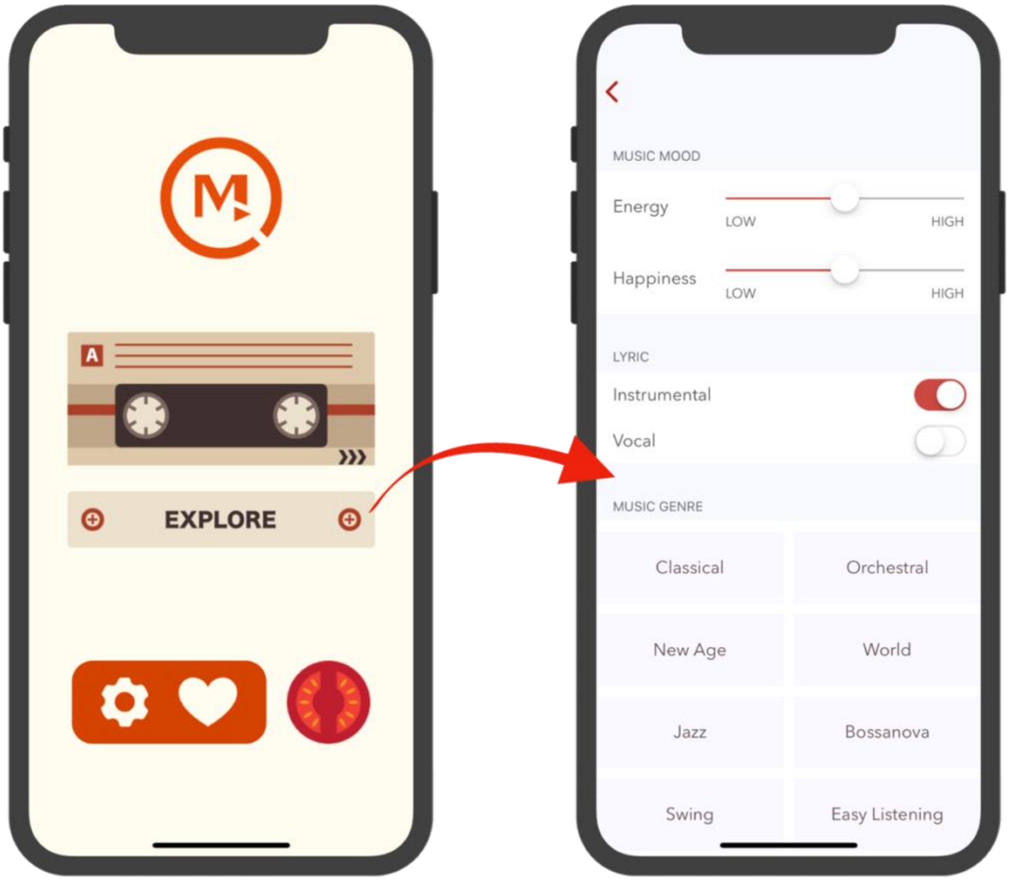


Figure 3. Search interface of the music information retrieval system (Moody App)

[1] The numbers in parentheses denote the number of tracks associated with each genre tag.

Moreover, for the sake of more personalized listening experience, users could further customize the music filters in terms of music type (i.e., instrumental versus vocal) and music genre (see Appendix 5 for all genre options), and the MIR system would generate playlist based on the aforementioned user-specified music filtering criteria.

Particularly, as mentioned above, only 4994 tracks were accompanied with the music type metadata. Hence, if a user specifies one particular music type (i.e., instrumental or vocal) for automatic playlist generation, the MIR system would select music from the 4994 tracks with known music type. Otherwise, if the music type is not of user's concern, the Moody App would select music from the 10k music pool. Besides, to ensure the diversity of participants' listening experience, a random offset was incorporated in the playlist generation procedure so that the MIR system would be able to return diversified music recommendation even under the same set of music filtering criteria.

Personal Playlist Management

Apart from exploring new music via the aforementioned system-generated playlist, the system users could study with their own favourite music by either looping through their personal playlist or looping a single favourite piece as well (Figure 4).

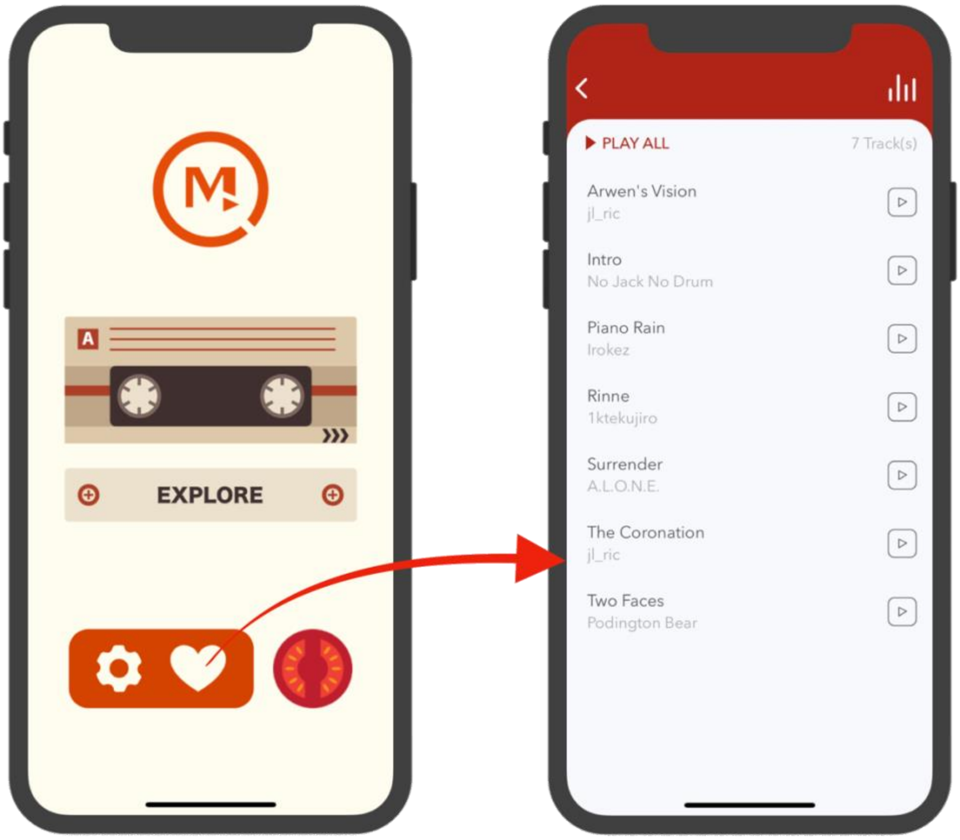


Figure 4. Personal playlist interface of the MIR system (Moody App)

Particularly, users could add a track to his/her personal study music playlist via either the *like/unlike* button on the player panel or the button on the music rating widget which would always appear upon the completion of each session (Figure 5). Hence, given this flexibility, participants could feel free to immerse themselves with music and learning task during the course of the learning session and carry out their music collection task at the end of each session.

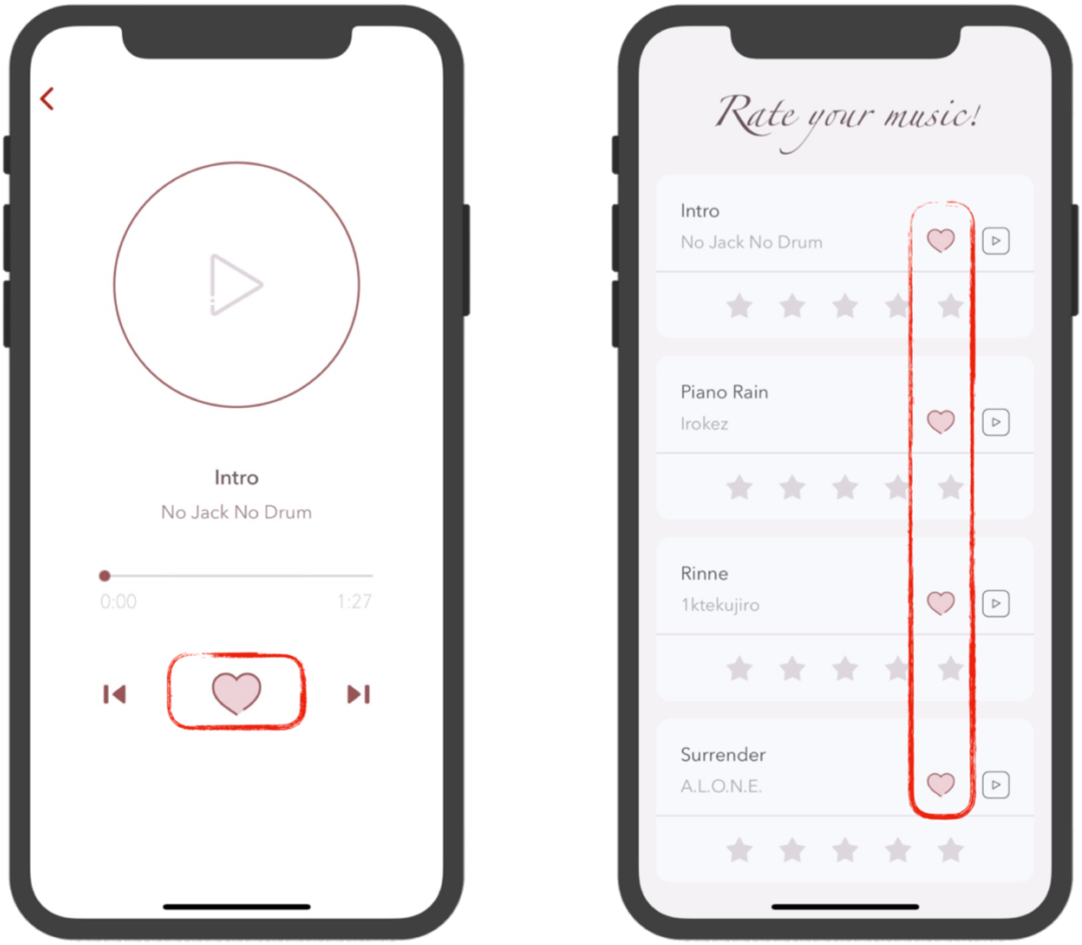


Figure 5. Like/unlike button on the player panel (left) and music rating widget (right)

User Activity Monitoring and Logging

As an observational research, the Moody App also incorporated the user activity monitoring feature to unobtrusively record users' searching and listening behaviour. Specifically, participants' music searching history, session listening history and other timestamped user-related event (e.g., play/pause, skip, rewind, like/unlike a song) were logged to the back-end database in real time so as to facilitate future fine-grained user-centric analysis (e.g., identifying skipped songs, tracing back listening context of liked songs, etc.)

3.2.2 The Online Survey Subsystem

To facilitate tracking and snapshotting the contextual information associated with each learning session, the proposed music app implemented a popup survey subsystem to periodically collect participants' emotional status (i.e., valence and arousal), task description and task load, as well as their self-reported learning engagement and perceived learning effect of music (i.e., enhance versus distract), etc.

During the field experiment, the music app was designed to conditionally and periodically pop-up pre-survey or post-survey so that the longitudinal data could be analyzed per session (i.e., a studying/music listening period between pre-survey and post-survey).

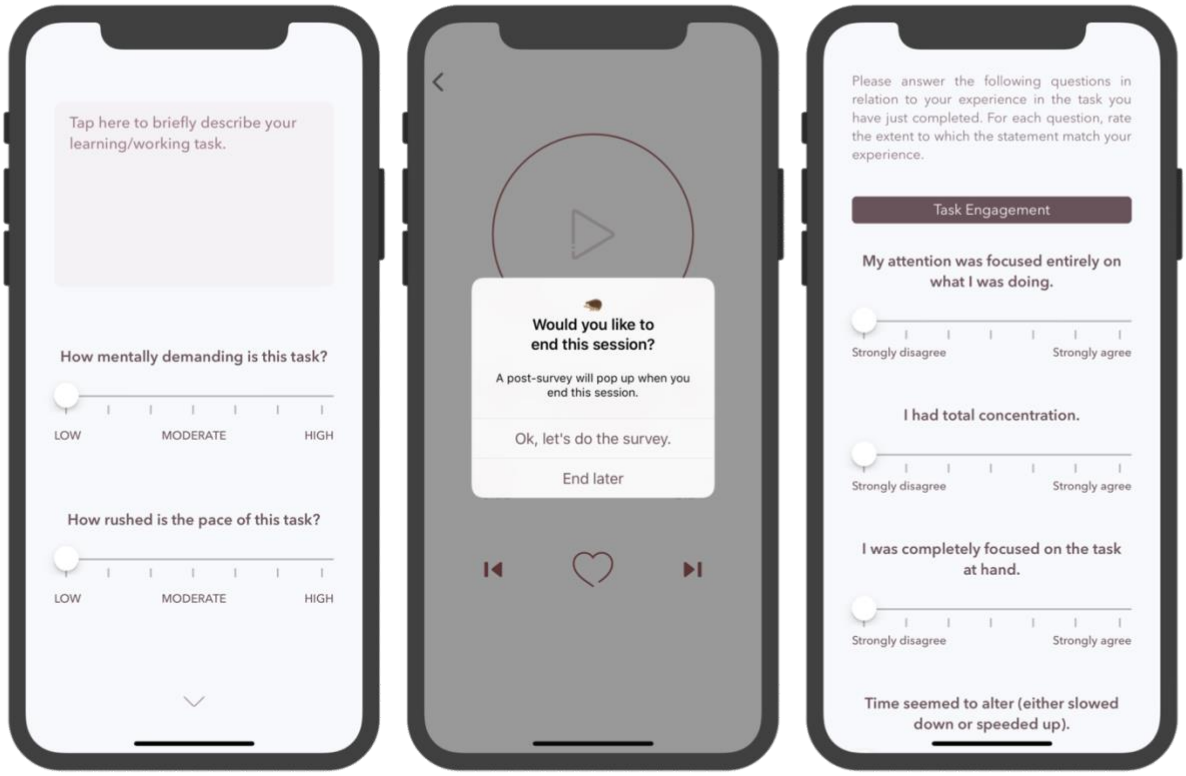


Figure 6. A sample of interfaces of the survey subsystem. Note that not all questions are included.

Particularly, the pre-survey, which contains questions on participants' emotional status and learning task, would be automatically popped up after music searching. And a 25-minutes tomato timer would be invoked upon the completion of the pre-survey. A prompt for ending the session and completing the post-survey would be subsequently popped up in 25 minutes or when users pause the music, though the user could opt for postponing the post-survey and continuing learning as well. Generally, the post-survey would include questions on an optional update of learning task, participants' emotional status at the moment, their self-reported learning engagement and their perceived learning effect of music (i.e., enhance versus distract).

Moreover, as mentioned in the previous section, ratings for music listened during the learning session would be included at the end of the survey as well.[2]

3.3 Procedures

The whole user experiment could be divided into three major phases, including pre-experiment survey and instructions, the one-week field experiment, and the post-experiment interview.

3.3.1 Phase 1: Pre-Experiment Survey and Instructions

During registration, participants needed to complete a pre-experiment questionnaire for demographical information collection and personality trait assessment (Ten Item Personality Inventory, TIPI) (Samuel D Gosling et al., 2003).

Subsequently, a set of computer-based cognitive ability assessment were arranged during the face-to-face instruction session so as to measure participants' working memory capacity and multitasking ability.

The consent form was presented to and signed by the participant at the beginning of each instruction session, and an introduction to the experimental task and the Moody App was covered in the user instruction in detail. During App installation, participants would have several minutes of practice of the Moody App using a simulator device provided by the facilitator, being told that the facilitator is ready to answer any questions they may have and they can feel free to contact the facilitator via WhatsApp or email if they have any enquiries after the instruction session.

3.3.2 Phase 2: One-Week Field Experiment

During phase 2, participants were encouraged to perform their learning tasks with music playing in the background. Specifically, their tasks in the one-week field experiment was specified as follows:

1) Use Moody to find as many pieces of music as possible (but at least 20 pieces) which you deem *suitable for listening while studying* and add them to your personal playlist;
2) Use Moody for as many learning sessions as possible (at least 1 session each day) and complete the pop-up surveys before and after each session.

[2] Participants could re-play the song via the music rating widget to remind themselves their feeling about a particular music piece.

During the course of each learning session, to minimize interruption, the Moody App would, by default, automatically select the next music piece to play based on participants' pre-specified music filtering criteria. Participants could build their personal music library effortlessly by simply skipping the disliked tracks and collecting the preferable ones. Meanwhile, the interactive online survey would also periodically pop up on the Moody App so as to track and record a set of context-related data as well as participants' self-reported learning engagement and perceived learning effect of music.

3.3.3 Phase 3: Post-Experiment Interview

Upon the completion of the field experiment, a face-to-face interview was arranged for each participant. Specifically, the interview questions were mainly concerning participants' insights on studying with background music, comments on the characteristics of preferable study music and general music preference during learning as well as their feedbacks on and suggestions for the experiment procedure and the MIR system (i.e., Moody App).

Each participant was paid a nominal renumeration at the end of the experiment.

3.4 Measures

Recalling the variables identified in the conceptual framework, the data collected in this study could mainly be categorized into three types, i.e., participant-related, context-related, and music-related. Sources and measures of the three categories of data were described in the sections below.

3.4.1 Participant-Related Data

Overall, all participant-related data (i.e., cognitive ability, personality traits, demographics) were collected during the first phase of the experiment.

Cognitive Ability

In light of the research literature, two cognitive ability of interest, i.e., working memory capacity and multitasking ability, were assessed using a set of computer-based cognitive tests during the pre-experiment instruction session.

Working Memory Capacity (WMC). From the perspective of executive attention, the WMC has been conceptualized as one's ability to sustain attention or suppress irrelevant information in the presence of distraction (Engle et al., 1999, p. 104; Wilhelm et al., 2013), and thus has received increasing attention in recent studies on relevant topics (Burunat et al., 2014; Christopher & Shelton, 2017), in view of the irrelevant sound effect hypothesis. In this study, participants' working memory capacity was assessed using the N-back Test, one of the most popular assessment techniques of WMC in the field of neuropsychology (Lezak et al., 2004, p. 441). Generally, in the N-back task, participants see a sequence of stimuli, and are asked to indicate whether the current stimulus is the same as the stimulus item "n" steps back. Particularly, this study adopted the 2-back condition. Participants' averaged reaction time of correct response were used as indicators of participants' working memory capacity.

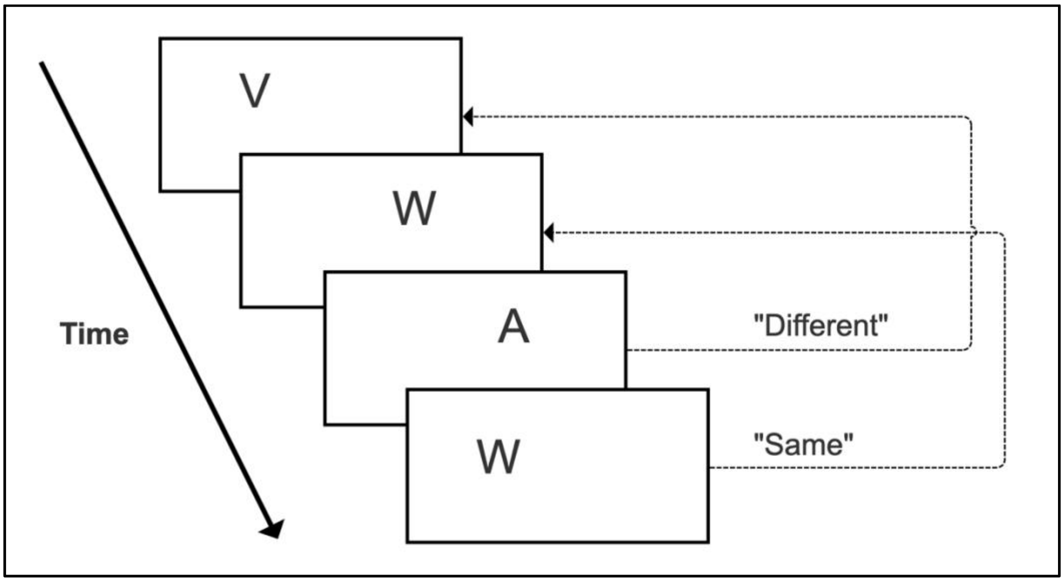


Figure 7. Illustration of the n-back task. The expected correct answers were noted in quotation.

Multitasking Ability. The multitasking ability, which requires simultaneously processing more than one stream of information, is also known to be closely related to one's executive functions such as suppressing task-irrelevant information. Particularly, this study followed Stoet et al. (2013)'s Multitasking Test design for assessing participants' multitasking ability. During the test, the participants would see four types of figures (Figure 8) presented at either top or bottom half of a box. Participants were asked to respond to the figure stimulus according to its shape, filling, and position. In the shape-only task, the figure stimulus was presented in the top half of the box. Participants only needed to judge the shape of the stimulus (i.e., square or diamond) and press the corresponding keyboard button as response. In the filling-only task, the figure stimulus was presented in the bottom half of the box. Participants were asked to judge the filling of the figure (i.e., two dots or three dots). In the final dual task block, the stimulus was randomly presented in either half of the box. Participants then needed to judge either its shape or filling according to the position of the stimulus (i.e., top or bottom half of the box). The response time of correct answers (RTCA) in the first two blocks are deemed as the base line, and the multitasking ability is generally measured as the RTCA in the mixed-task block minus the averaged RTCA during the first two single-task blocks.

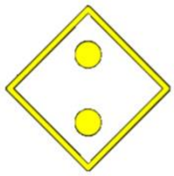
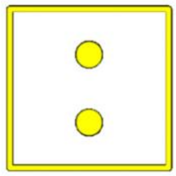
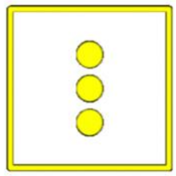
Figures				
Shape Task	Left Key	Left Key	Right Key	Right Key
Filling Task	Left Key	Right Key	Left Key	Right Key

Figure 8. Four types of figure stimuli and the expected response for each task. Rules: in the shape task, press the left key for figures with diamond and the right key for those with square; in the filling task, press the left key for figures with two dots and the right key for those with three dots.

On a side note, each cognitive test was accompanied with a pre-test practice session and the cognitive tests were conducted in counterbalanced order for all participants.

Personality Traits

Participants' personality traits of the Big Five dimensions were measured using a brief but validated inventory (Ten Item Personality Inventory, TIPI) developed by Samuel D Gosling et al. (2003) (Appendix 2) during registration. Particularly, to investigate the group-wise difference of learners' music preference, this study followed the TIPI norms provided by S. D. Gosling et al. (2014) for the personality group assignment (i.e., high versus low for each personality dimension). The norms for the five personality trait dimensions are 4.44 (extraversion), 5.23 (agreeableness), 5.4 (conscientiousness), 4.83 (emotional stability), 5.38 (openness to experience), respectively (S. D. Gosling et al., 2014).

Demographics

As mentioned in section 3.3.1, this study also collected a set of demographical information of each participant through the pre-experiment survey. Specifically, the survey mainly includes questions on participants' gender, age, major, prior music listening frequency during learning (i.e., habitually versus rarely/never study with background music), music listening behavior in everyday life, musical training background, genre preference, etc. Details on the questions and scales could be found in Appendix 1.

3.4.2 Context-Related Data

The context-related data, i.e., participants' emotional status, task description and task load, learning engagement and perceived learning effect of music, were mainly collected via the survey periodically popped up in the Moody App. Table 1 summarized the operationalization and measures of the aforementioned variables.

Table 1

Measures of the Context-Related Data

Variable	Related survey	Measures
Task load	Pre-survey	
Task description		Open-ended textual task description
Mental demand		NASA Task Load Index (adapted) (Hart & Staveland, 1988)
Temporal demand		
Emotional status	Pre/post-survey	
Mood type		Eight discrete mood categories (Scherer, 2005)
Valence		Continuums of -1~1 (Russell et al., 1989)
Arousal		
Learning Engagement	Post-survey	
AST		Flow State Scale (adapted) (Jackson et al., 2012)
CONC		
Perceived learning effect	Post-survey	
Enhance vs. distract		7-point Likert scale question adapted from Mayfield and Moss (1989)

Note. AST = altered sense of time, CONC = total concentration on the task at hand

Task Load

In view of the potential moderating effect of task load on the interaction between music and learning, participants' workload (i.e., the mental demand and temporal demand of learning task) were measured using the instrument adapted from the NASA task load index (Hart & Staveland, 1988). Particularly, the mental demand refers to the amount of mental or perceptual activity required such as calculating, remembering, thinking, etc. (i.e., easy and simple versus demanding and complex), while the temporal demand refers to the time pressure felt due to the task (i.e., slow and leisurely versus rapid and hurried). In addition, the textual description of the learning task was also obtained from participants for further analysis.

Emotional Status

Both discrete and dimensional measures were used to indicate participants' emotional status. Specifically, Scherer (2005)'s semantic space for emotions was adapted to select the discrete emotion adjectives (e.g., excited, happy, pleased, sleepy, bored, depressed, frustrated, annoyed). Besides, based on Russell et al. (1989)'s circumplex model of emotion, two affective dimensions, i.e., valence (unpleasant vs. pleasant) and arousal (calm vs. energetic) were measured using a set of continuums (-1~1) as well.

Learning Engagement and Learning Effect

Finally, given our primary interest in how music benefits learning, two constructs, i.e., total concentration on the task at hand and altered sense of time, have been borrowed from the flow theory to indicate participants' learning engagement and were measured using the adapted flow state scale (Jackson et al., 2012). Specifically, both constructs depict a special absorbing experience from two different perspectives. *Total concentration on the task at hand* was measured as participants' averaged rating of three statements (i.e., *"My attention was focused entirely on what I was doing", "I had total concentration", "I was completely focused on the task at hand"*) on 7-point Likert Scales, while the *altered sense of time* was measured as participants' averaged rating of *"Time seemed to alter (either slowed down or speeded up)"*, *"It felt like time went by quickly"*, and *"I lost my normal awareness of time"* also on 7-point scales.

Additionally, given the limited feasibility of measuring learning performance in naturalistic experiment design, this study only measured participants' perceived effect of music on task performance using the question (i.e., *"To what extent did the music affect your performance on this task?"*) adapted from Mayfield and Moss (1989)'s research design. And participants responded to this question on a 7-point Likert scale (1=Very much distracted me, 4=Had no effect, 7=Very much enhanced my work) as well.

3.4.3 Music-Related Data

Towards the goal of anatomizing the effect of music on learning, both high-level and low-level music characteristics were incorporated for music feature representation. Particularly, the high-level music characteristics mainly included the music emotion predicted from the fine-grained music features and the music metadata provided by the Jamendo dataset, while the low-level ones were acoustic features obtained from audio signal processing, which are known to be closely related to the dynamical, rhythmic, and timbral quality of music.

Music Types and Music Styles

As mentioned in section 3.2.1, all tracks in the Moody music library were associated with a rich set of music metadata provided by Jamendo, including music type (i.e., instrumental or vocal) and music genres. Particularly, in the Moody database, a track could be associated with multiple genres and fine-grained sub-genre tags (e.g., 70s, Asian, jazz fusion). Hence, a simple one-hot encoding of the genre tags would produce an extremely sparse matrix.

To ensure the further analysis could be conducted at a reasonable degree of granularity, the genre tags were further grouped into five music styles (i.e., mellow, unpretentious, sophisticated, intense, contemporary) based on a five-dimensional musical preference dimensions model (MUSIC) (Bonneville-Roussy et al., 2013). In Bonneville-Roussy et al. (2013)'s original research, the five dimensions were identified based on the clustered preference of genres and the resulting five music styles were reported to share relatively consistent psychological and musical characteristics (e.g., *psychological:* affect, complexity; *musical:* timbre, loudness) (Hallam et al., 2016, p. 264). The mapping between genres and styles was specified in Table 2.

Table 2

Mapping between Genres and Styles

Styles	Genres
Mellow	New Age, Dance, Electronic(a), World
Unpretentious	Pop, Religious, Country
Sophisticated	Classical, Opera, Orchestral, Jazz, Blues, Folk, Gospel
Intense	Rock, Heavy Metal, Alternative, Punk
Contemporary	Funk, Soul, R&B, Hip-hop, Rap, Reggae

Note. Adapted from Bonneville-Roussy et al. (2013)

On a side note, as a music piece could still be associated with genres of more than one styles (e.g., simultaneously being pop and electronic), it's possible that a single track would be associated with more than one style tags, and the style tags hence were further one-hot encoded.

Low-Level Music Features

To measure the dynamical, rhythmic, and timbral characteristics of music, a set of low-level acoustic features were extracted through the audio signal processing as well. Particularly, the acoustic features were extracted using a specialized Python library for music and audio signal analysis (i.e., LibROSA) (McFee et al., 2015). Table 3 summarized the low-level music features involved in this study.

Table 3

Specification of Low-Level Music Features

Category	Code	Statistics	Description	Dim.
Loudness (Dynamics)	rmse	mean, std	The root-mean-square energy (RMSE) of each frame computed from the audio spectrogram. (Herrera-Boyer et al., 2006, p. 172)	2
Rhythm	tempo	mean	The averaged tempo of the audio sample (beats per minute). (Hainsworth, 2006, p. 101)	1
	onset_frq	mean	The average onset frequency, i.e., the number of notes per second. (Xiao & Yi-Hsuan, 2017)	1
Timbre	rolloff	mean, std	The roll-off frequency of each frame below which a certain fraction of the total energy of that frame is contained (normally set to 0.85). (Tzanetakis & Cook, 2002)	2
	centroid	mean, std	The centroid of the spectral distribution at each frame. (Fitzgerald & Paulus, 2006, p. 136)	2
	flatness	mean, std	The spectral flatness at each frame which indicates whether the spectral distribution is spiky or smooth. (Herrera-Boyer et al., 2006, p. 173)	2
	zcr	mean, std	The zero-crossing rate at each frame which indicates the noisiness of the audio signal. (Fitzgerald & Paulus, 2006, p. 136)	2
Total				12

Note. Dim. = the number of feature dimensions.

The loudness features (also known as music dynamics) generally represent the volume of a sound (rmse_mean) and the variation thereof (rmse_std). Particularly, for the sake of more accurate acoustic quality representation, the RMSE was computed from the STFT spectrogram instead of the raw audio samples.

Besides, the rhythmic features were also computed to represent the systematic pattern and arrangement of the musical notes. As for the rhythm-related music characteristics representation, the tempo was estimated from the note onsets correlation and the average onset frequency was computed based on the detected onset frames.

Finally, the timbre quality and the nosiness of music were characterized by the spectral distribution and spectral feature of the audio sample. Specifically, the roll-off is generally used as an estimation of the amount of high frequency consists in the audio signal, and a roll percentage of 0.85 was adopted following the well-accepted convention (Tzanetakis & Cook, 2002). Moreover, the spectral centroid and flatness were also computed so as to represent the central tendency and the smoothness of the spectral distribution. The zero-crossing rate was computed from the audio sample time series for nosiness indication as well.

On a side note, the aforementioned features were all extracted using the default parameter settings (e.g., sample rate = 22050, hop length = 512) of the music processing tool (i.e., LibROSA) and the aggregated statistics (i.e., mean, standard deviation) of the frame-wise feature vectors were mainly adopted as the clip-level feature representation.

Music Emotion

In light of the arousal-mood-hypothesis, to further probe the role of the affective traits of music, this study also computed the music emotion on the continuums (-1~1) of two affective dimensions (i.e., happiness and energy).

Specifically, the happiness and energy level of music were predicted from the low-level acoustic features using the support vector machine (SVM). And the prediction model was built on the training data obtained from Xiao and Yi-Hsuan (2017)'s previous study.[3]

On a side note, the music processing toolkit adopted by this study (i.e., LibROSA) also supports the extraction of a set of matrix-like audio features (e.g., MFCC, tonnetz, chromagram, etc.). Though these matrix features were not adopted for further analysis on low-level features because of their limited interpretability, we did originally include such features in the feature set for music emotion prediction. Besides, with respect to the feature set for model building, it's also noteworthy that the regression model for happiness and energy prediction were built on two separate reduced feature set obtained from univariate feature selection. Particularly, the selected features for happiness prediction included *centroid_mean*, *flatness_std*, *rolloff_mean*, etc., while the selected features for energy prediction included *rmse_mean*, *centroid_mean*, *flatness_mean*, *flatness_std*, *rolloff_mean*, *rolloff_std*, *zcr_mean*, etc.[4]

The 10-fold cross-validated root-mean-squared error (RMSE) for happiness and energy prediction were 0.311 and 0.229 respectively.

[3] Statistics of the training data. Happiness: M = .061, SD = .309; Energy: M = .173, SD = .453

[4] To save space, only the features finally adopted as low-level features of interest (i.e., features listed in the previous section) were reported.

3.5 Participants

30 postgraduate students (*Gender:* 13 males, 17 females; *Age:* M = 24; SD = 1.99) in a comprehensive university in Hong Kong were recruited as the participants of this field experiment, whose majors of study covered information science, education, computer science, engineering, architecture, and linguistics.

Additionally, in regard to their musical training background and music listening frequency in everyday life, 12 of them had received formal music training and 17 of them reported listening to music on a daily basis (i.e., from approximately once a day to almost always). As for those who are less involved in music, 8 of the recruited participants reported listening to music less than several times a week (*several times a month:* 7, *seldom:* 1). Besides, concerning their prior habit of studying with background music, they responded with "almost always" (2), "usually" (5), "sometimes" (11), "rarely" (8), and "never" (4) when asked how frequently they listened to music while learning. The most popular genres among our recruited participants were pop (26), folk (13), classical (13), easy listening (13), and country (10).[5]

Table 4

Specification of Participant Grouping

Learners' Traits	Number of participants		Grouping criterion
	Low	High	
Personality			
Extraversion	15	15	TIPI norms
Agreeableness	21	9	
Conscientiousness	17	13	
Emotional stability	16	14	
Openness to experience	19	11	
Cognitive ability			
Working memory capacity	15	15	Median
Multitasking ability	15	15	
Prior music listening frequency			
During learning	12	18	Center of the scale

Note. Participants who reported sometimes/usually/always studying with background music were assigned to the high group while the rest of them were assigned to the low group.

[5] The numbers in parentheses denote the number of participants who have stated their preference for a particular genre.

Finally, to investigate the group-wise difference of learners' music preference, we divided participants into two groups (i.e., high versus low) on various traits dimensions, including personality, cognitive ability, and prior music listening frequency during learning. Details on group distribution and discretization method could be found in Table 4. Particularly, as shown in Table 4, the group distribution on the agreeableness dimension was extremely unbalanced. Hence, to avoid the potential bias, this personality dimension was not considered for further group-wise analysis.

3.6 Data Analysis

3.6.1 Data Preprocessing

As mentioned in section 3.2.1, not all tracks in the Moody music library were associated with music type metadata (i.e., instrumental or vocal), while the Moody App could recommend music without such metadata in certain cases (e.g., the participant opted for both instrumental and vocal music). Hence, there is chance that certain tracks in participants' listening history were not accompanied with music type information. As for such missing data, we manually labeled the tracks at the preprocessing stage. Particularly, 379 music pieces were manually labeled during preprocessing.

Additionally, to facilitate the comparison of playlist tracks (i.e., music selected by participant) and non-playlist tracks (i.e., music listened but not selected) as well as the comparison of music selected by different user groups, all tracks traced from participants' listening history and those selected into their personal playlists were labeled accordingly. Particularly, for each learners' traits dimensions, the music pieces selected by both user groups (i.e., high and low) were labeled as *Mix*, which were temporarily excluded in the group-wise analysis on that particular dimension.

Finally, as for the learning session-related data, we excluded the sessions with duration less than five minutes and sessions containing missing data caused by technological failures (e.g., App crashes), which resulted in a total of 195 valid learning sessions. In order to investigate the effect of music on learning, we also computed the general music characteristics associated with each session, including dominant genre (i.e., genre tags that appeared most in each session), averaged happiness and energy level, and averaged low-level music features.

3.6.2 Overview of Data Analysis

Table 5 summarized each of our research questions and the corresponding analysis method. Particularly, given that the data collected was found to violate the normality assumption, most of the statistical analysis adopted were based on the ranked data (i.e., non-parametric test or test in its robust version).

Table 5

Mapping between Research Questions and Data Analysis Strategies

Research questions	Data analysis strategies
RQ1.1: What kind of music would be deemed suitable for learning? RQ1.2: Would the music preference of learners vary across different user groups?	1. Pearson's Chi-Square test 2. MANOVA 3. Discriminant function analysis 4. Wilcoxon rank-sum test 5. Qualitative triangulation
RQ2.1: What kind of music would be deemed as benefiting learning versus impairing learning? RQ2.2: Would task load and learners' traits moderate this effect?	1. Repeated measures correlation 2. Qualitative triangulation

To investigate learners' subjective music preference (RQ1.1) and the underlying individual difference thereof (RQ1.2), the comparison of playlist and non-playlist tracks as well as the comparison of music selected by different user groups were conducted based on a set of inferential statistical analyses. Specifically, the Pearson's Chi-Square test was adopted for comparing music types and music styles, and the word cloud of genre tags was included as triangulation as well. As for another high-level music characteristics, the music emotion was first compared using MANOVA on the whole, and the discriminant function analysis as well as the univariate tests were subsequently applied to further anatomizing the effect. Finally, a set of Wilcoxon rank-sum tests were adopted for comparing low-level music features. Benjamini-Hochberg procedure at a false discovery rate of 0.1 was applied for controlling potential Type I error introduced by multiple comparisons.

To further explore the effect of music on learning (RQ2.1), this study also investigated the correlation between music features and the proposed learning outcome measures (i.e., perceived learning effect, altered sense of time, total concentration on the task at hand). Particularly, the repeated measures correlation (RMCORR) was applied using Bakdash and Marusich (2017)'s R implementation to model the intra-individual association. The RMCORR analyses were also conducted on dataset partitioned by participants' traits and task load so as to control for the potential confounding effect of these variables.

Chapter 4: Results and Discussion

As mentioned in the previous section, overall, a total of 195 valid learning sessions were recorded by the Moody App during the one-week field experiment. Among the 1525 music pieces traced from participants' session listening histories, 529 pieces were selected into their personal study music playlists.[6] Based on a set of playlist-wise and session-wise analyses, findings and discussions on learners' music preference (RQ1) and the effect of music on learning (RQ2) were presented in section 4.1 and 4.2 respectively.

4.1 Music Preference of Learners

4.1.1 General Tendencies of Learners' Music Preference

The comparison between playlist tracks and non-playlist tracks from users' listening history revealed some general tendencies of learners' preference on music styles, music emotion, and other fine-grained music features.

Music Types and Music Styles

Each music piece involved in this study had a set of metadata provided by Jamendo, including types (i.e., instrumental or vocal) and genres. A set of Pearson's Chi-Square tests were thus applied to probe the association between learners' preference and the high-level music characteristics described by the metadata.

Particularly, given that a single piece could be associated with multiple genres and sub-genres, we further grouped the genre tags as five music styles (i.e., mellow, unpretentious, sophisticated, intense, contemporary) according to Bonneville-Roussy et al. (2013)'s MUSIC musical preference dimensions model (see details in section 3.4.3)[7] so that the statistical tests could be conducted at a reasonable degree of granularity. Benjamini-Hochberg procedure at a false discovery rate (FDR) of 0.1 was applied for controlling Type I error introduced by multiple comparisons. The mosaic plots presented in Figure 9 visualize the contingency tables with each cell shaded in accordance with the direction and the significance of the effect.

---

[6] The above summary statistics have excluded the duplicate records (note that a single piece could be listened or selected by multiple participants). In terms of double counting, the number of pieces listened and selected were 2618 and 625 respectively (approximately 13 songs listened per session and 21 songs collected per participant).

[7] Note that a single track could still be associated with more than one style tag, and the style tags were hence transformed into one hot encoding in the following analysis.

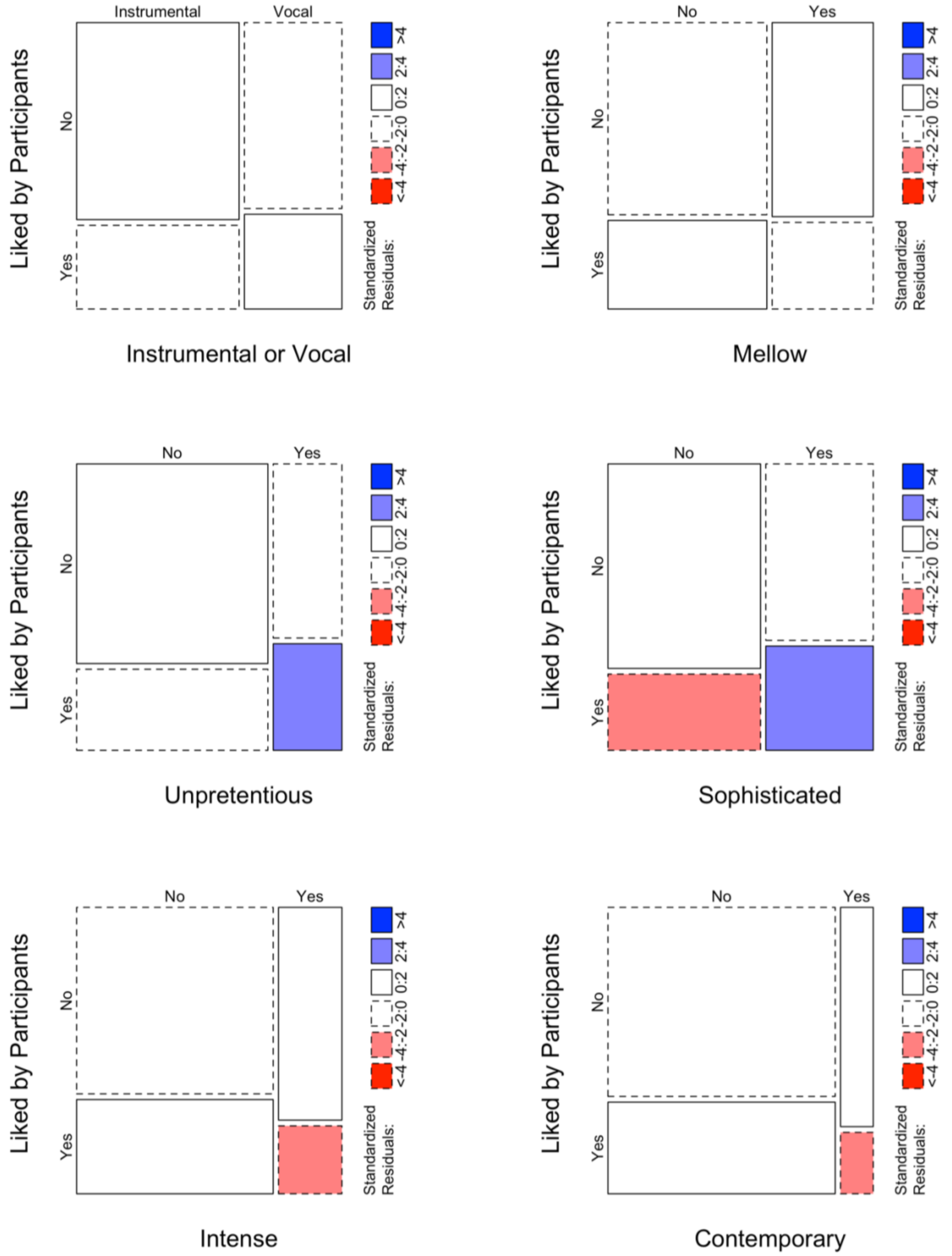


Figure 9. Mosaic plots of music metadata and participants' music preference. a) N=1525; b) A track would be labelled “liked by participants” as long as at least one participant selected it into his/her playlist; c) The sign and magnitude of standardized residuals indicate the direction and the significance of the effect.

Even though the existence of vocal element has been reported as an important music selection criterion by some participants in their post-interviews, the Pearson's Chi-Square test did not reveal a significant association between learners' general music preference and music type (i.e., instrumental or vocal; $\chi^2(1)=2.582$, $p=.108$). Nonetheless, some interesting patterns did emerge when participants' personal factors were considered. Relevant discussions could be found in section 4.1.2.

As a prior step of breaking down and interpreting the association between learners' music preference and music styles, a genre tags word cloud of participants' "liked" music was plotted to provide some complementary qualitative insights (Figure 10).

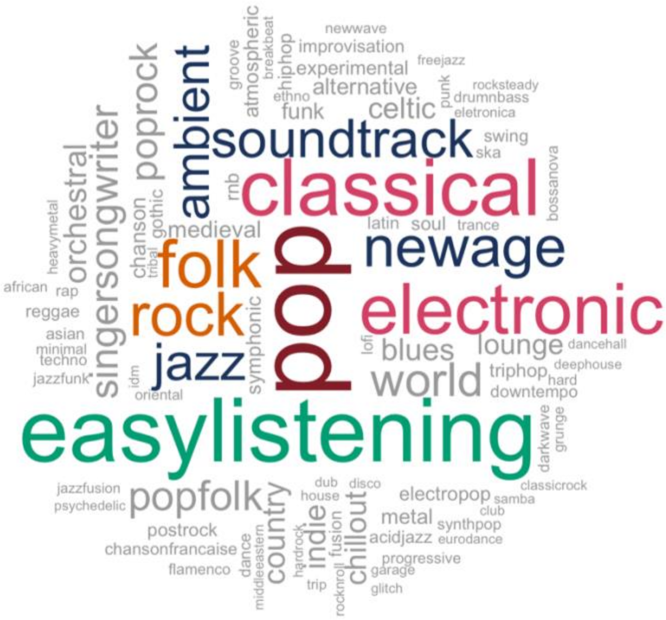


Figure 10. Word cloud produced from genre tags of participants' "liked" music. a) The size, centeredness, and colour of genre tags indicate its frequency; b) The top 5 genres were pop (n=170), easy listening (n=141), classical (n=123), electronic (n=107), and folk (n=100).

As shown in Figure 9 and Figure 10, unpretentious (e.g., pop) and sophisticated (e.g., classical, folk, jazz) music stand out from the music selected by participants, whose positive relationship with learners' music preference was further confirmed by the Pearson's Chi-Square test (*unpretentious:* $\chi^2(1)=11.393$, $p<.001$, z=2.400; *sophisticated:* $\chi^2(1)=17.174$, $p<.001$, z=2.632). On the contrary, the intense (e.g., rock, punk, alternative, heavy metal) and contemporary (e.g., hip-hop/rap, soul/R&B, funk, reggae) music seems to be less common in participants' personal study music playlists. As shown in the Chi-square analysis, the number of intense and contemporary music selected into participants' playlist was far less than the expected frequency (*intense:* $\chi^2(1)=11.491$, $p<.001$, z=-2.443; *contemporary:* $\chi^2(1)=9.036$, $p=.002$, z=-2.330). The above observations might lead us to the conclusion that, generally speaking, learners might prefer unpretentious and sophisticated music (e.g., pop, classical, folk) to intense and contemporary music (e.g., rock, punk, hip-hop/rap).

With regard to learners' general preference on music styles, another interesting observation is that, although mellow music (e.g., electronic, new age) also frequently appeared in participants' playlist, the Pearson's Chi-square test did not detect a significant association between this music style and participants' add-playlist behaviour ($\chi^2(1)= 0.079$, $p=.778$). Some possible explanations could be: 1) only certain group of users share a preference for mellow music; 2) only a subgroup of mellow music was deemed suitable for listening while learning and participants were thus more selective in this music style. Particularly, the first assumption was also corroborated by one of the findings reported in section 4.1.2.

Music Emotion

In light of the arousal-mood-hypothesis (Husain et al., 2002) discussed in section 2.1, a MANOVA was applied to investigate learners' general preference on two affective dimensions of music emotion (i.e., happiness, energy). Particularly, as specified in section 3.4.3, the music emotions were obtained via the predictive modeling of acoustic features using support vector machine (SVM). And the MANOVA was conducted using Munzel and Brunner (2000)'s robust method which revealed a statistically significant difference between playlist tracks and non-playlist tracks in terms of music emotion (F=12.306, $p<.001$).

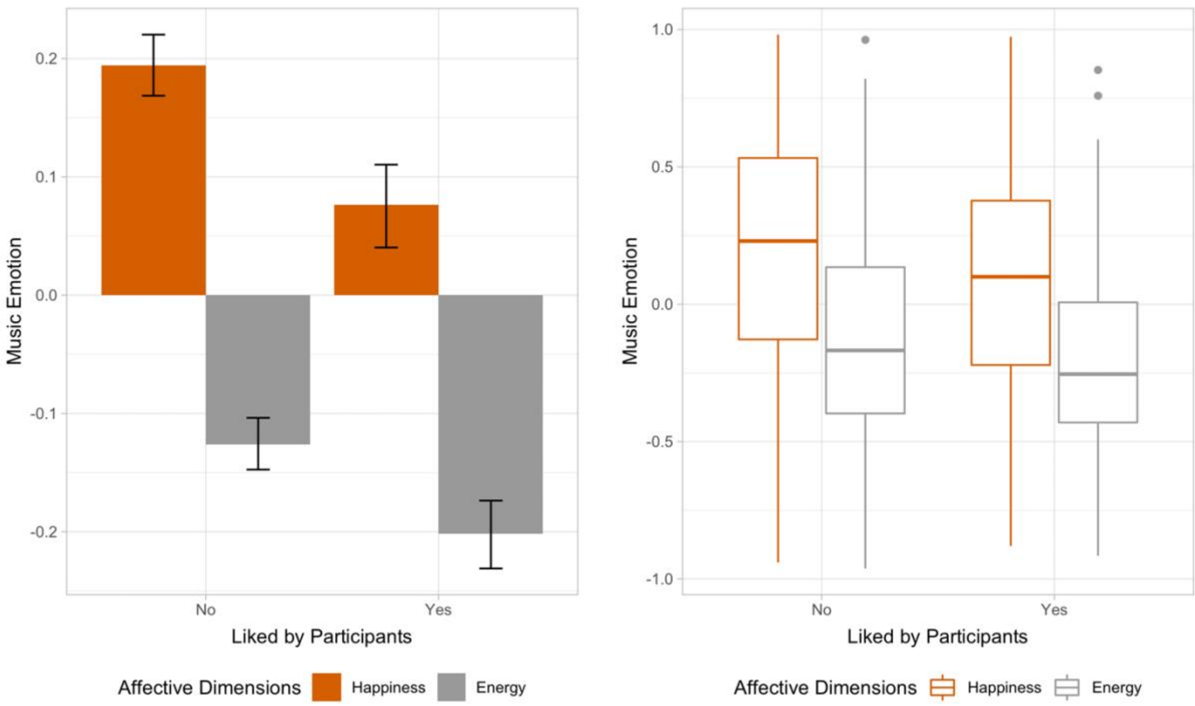


Figure 11. Error bar chart showing the mean of music emotion of non-playlist and playlist tracks (left); Boxplots of music emotion distribution (right)

A set of univariate analyses and discriminant function analysis (DFA) were conducted as the follow-up analysis of MANOVA. As shown in Table 6, Wilcoxon rank-sum tests revealed statistically significant difference both in music happiness and music energy.

Table 6

Comparison of Music Emotion Between Playlist Tracks and Non-playlist Tracks

Music Emotion	Non-playlist Tracks		Playlist Tracks		W	p
	M	SD	M	SD		
Happiness	0.194	0.431	0.076	0.417	289220***	<.001
Energy	-0.126	0.367	-0.201	0.317	278049***	<.001

Note. a) The valid range of happiness and energy were both -1~1. b) W denote the Wilcoxon rank-sum test statistic. c) Multiple comparison correction was applied based on Benjamini-Hochberg procedure (FDR=0.1). * $p < .05$, ** $p < .01$, *** $p<.001$

Compared to the non-playlist tracks, the happiness level of participants' liked music tended to be slightly positive but relatively neutral (M=0.076, SEM=0.019). The mean energy level of participants' liked music was also found to be relatively lower (M=-0.201, SEM=0.015) as against the non-playlist tracks from their session listening history. The above observations suggest that, pleasant and soothing music are more likely to suit learners' music preference, as arousal-mood-hypothesis (Husain et al., 2002) would dictate.

Last but not least, the discriminant function analysis also revealed some interesting information on the current topic. As indicated by the coefficients of linear discriminants computed from DFA, music happiness (b=-2.441) showed larger effect than music energy (b=0.156), which seems plausible given that, compared to music happiness, learners' preference on music energy is more likely to vary across the particular learning task at hand, their state of arousal at the moment (e.g., sleepy versus nervous), etc.

Low Level Music Features

Apart from music emotion, the information-load characteristics of music (e.g., tempo, loudness) also play a theoretically important role in the interplay between music and learning (Kiger, 1989; Rey, 2012). To investigate whether dynamical, rhythmic and timbral features could possibly account for learners' inclination of music selection, a set of Wilcoxon rank-sum test comparisons were applied between playlist tracks and non-playlist tracks as well (Table 7). And Benjamini-Hochberg procedure at a false discovery rate (FDR) of 0.1 was applied for controlling Type I error.

Table 7

Comparison of Low-Level Music Features Between Playlist Tracks and Non-playlist Tracks

Music Features	Non-playlist Tracks		Playlist Tracks		W
	M	SD	M	SD	
Loudness					
rmse_mean	4.320	2.059	3.641	1.580	297364***
rmse_std	2.020	0.868	1.707	0.028	303475***
Rhythm					
tempo	119.96	25.849	120.86	26.984	248007
avg_onset_frq	3.451	1.286	3.302	1.105	266767*
Timbre					
rolloff_mean	3669.8	1207.0	3334.3	1103.2	288640***
rolloff_std	1662.0	548.78	1589.7	522.20	264593
centroid_mean	1774.7	544.79	1612.8	457.07	291256***
centroid_std	765.24	268.65	712.07	240.43	272878**
flatness_mean	0.015	0.014	0.010	0.009	299502***
flatness_std	0.035	0.025	0.031	0.024	270535**
zcr_mean	0.078	0.032	0.070	0.025	286948***
zcr_std	0.055	0.026	0.050	0.022	276366***

Note. a) W denote the Wilcoxon rank-sum test statistic. b) Multiple comparison correction was applied based on Benjamini-Hochberg procedure (FDR=0.1). * p < .05, ** p < .01, *** p<.001

As shown in Table *7*, some low-level music features did reveal statistically significant difference between songs selected into participants' playlists and those did not. Particularly, as suggested by Kiger (1989)'s music information load hypothesis, playlist tracks' loudness features (i.e., rmse_mean, rmse_std), which represents the dynamics of a music piece, were found to be significantly lower than those of non-playlist tracks (at p<.001). Besides, other features represent the timbre quality and noisiness of music (e.g., roll-off, spectral centroid, flatness, zero crossing rate) were also found to be significantly different between the playlist and non-playlist tracks. However, interestingly, the rhythm features seemed to play a less important role in distinguishing the two music groups: the average onset frequency (i.e., the number of notes per second) only showed trivial effect (W=266767, p=.035, r=-.054); the tempo, on the other hand, showed no statistically significant difference between playlist and non-playlist tracks, which indicate that there might be individual or contextual difference speaking of the preference on the speed of the music.

4.1.2 Group-Wise Music Preference Profiling

Based on the assumption that learners' preference might vary across individuals, we further compared the 529 playlist tracks in regard to the user groups the track belong to. Some interesting patterns manifested itself when participants' cognitive ability, personality, prior music listening frequency during learning were incorporated in the analysis. Relevant findings were discussed in detail in this section.

Music Types and Music Styles

In terms of the study music preference of participants of different level of cognitive ability, though no significant association between learners' multitasking ability and preferred music types/styles was observed, the Chi-square tests revealed that participants' inclination of music type (i.e., instrumental or vocal; $\chi^2(1)=23.928$, $p<.001$) and preference on the mellow music style ($\chi^2(1)=15.935$, $p<.001$) might be related to their working memory capacity (WMC). As shown in Figure 12, significantly more mellow music (e.g., electronic, world, new age) were liked by participants in the high working memory capacity group (z=2.248) instead of the lower one (z=-2.154), which makes sense, given that the electronic music, one of the typical genre of the mellow style, was reported to possess high timbre complexity compared to other genres (Parmer & Ahn, 2019) and thus might not be suitable for low-WMC learners. And high-WMC participants' preference for mellow music might also, in a sense, account for their inclination of instrumental music, as the genres in mellow style are primarily instrumental ones.

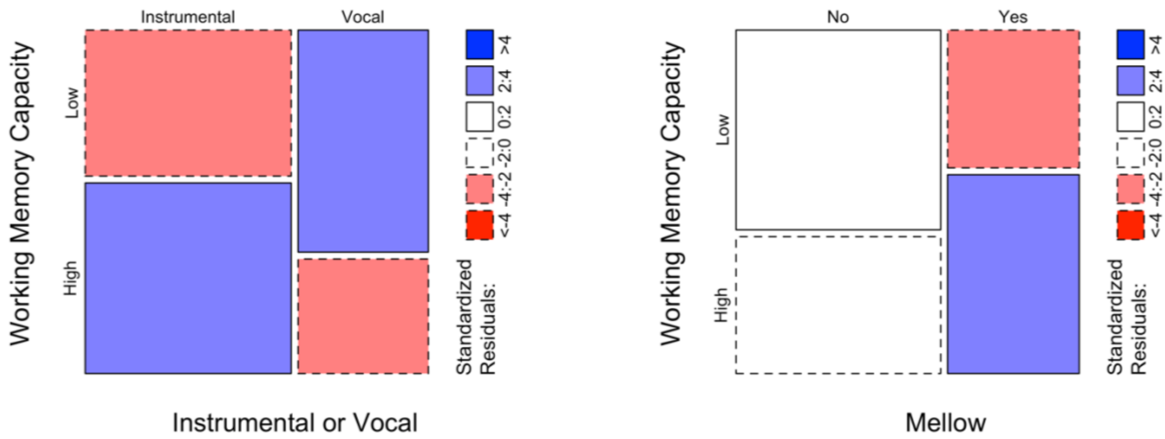


Figure 12. Mosaic plots of working memory capacity and preference on music type as well as the mellow music style. a) Tracks liked by both high & low WMC groups were temporarily excluded from the above analyses. b) The resulting number of tracks belong to each group were: high WMC (n=236), low WMC (n=257).

Apart from the cognitive ability, participants' personality was also found to play a role in their preference on music types and music styles.

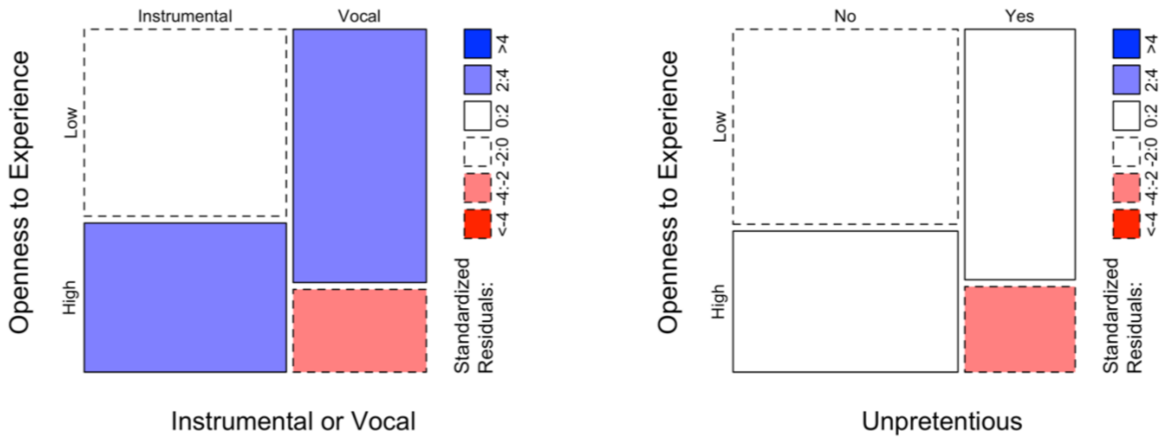


Figure 13. Mosaic plots of openness to experience and preference on music type as well as the unpretentious music style. a) Tracks liked by both high & low groups were temporarily excluded from the above analyses. b) The resulting number of tracks belong to each group were: high openness (n=183), low openness (n=318).

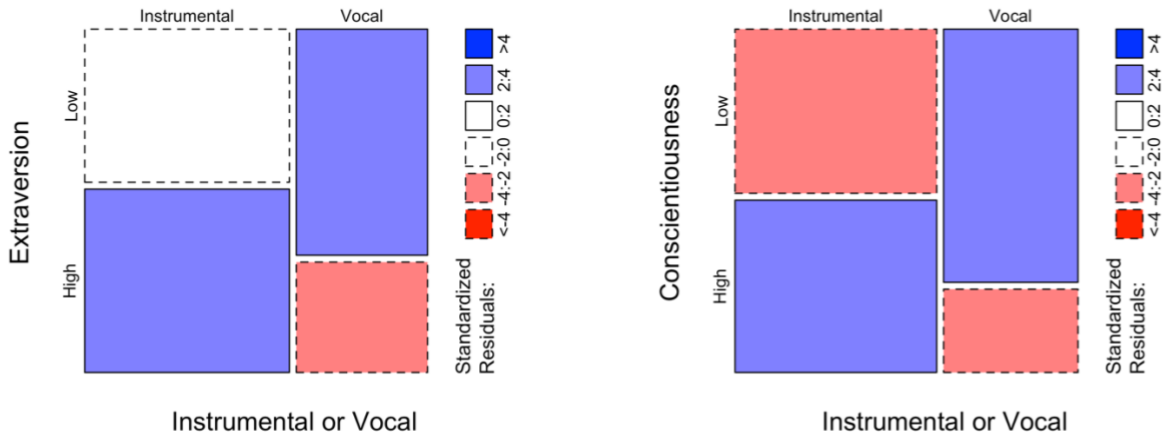


Figure 14. Mosaic plots of extraversion/conscientiousness and preference on music type. a) Tracks liked by both high & low groups were temporarily excluded from the above analyses. b) The resulting number of tracks belong to each group were: high extraversion (n=219), low extraversion (n=257); high conscientiousness (n=197), low conscientiousness (n=288).

Specifically, participants' personality trait on the *openness to experience* dimension showed significant association with their preference on music types ($\chi^2(1)=20.177$, $p<.001$) and significant negative association with the unpretentious music styles ($\chi^2(1)=13.010$, $p<.001$). As indicated by Figure 13, participants who scored high in *openness to new experience* were less stick to the mainstream unpretentious music (e.g., pop, country) but were more willing to try instrumental music in other genre for listening while learning. Finally, participants' personality traits in *extraversion* ($\chi^2(1)=21.456$, $p<.001$) and *conscientiousness* ($\chi^2(1)=33.789$, $p<.001$) were also found to be related to their preference on music types, i.e., significantly more vocal music were collected by introverts/low-conscientiousness participants instead of extroverts/high-conscientiousness participants (Figure 14).

The final learners' trait involved in the analysis was participants' prior music listening frequency during learning (i.e., habitually versus never/rarely study with background music). The Chi-square test revealed a statistically significant association between participants' preference for instrumental music and this personal factor ($\chi^2(1)=9.586$, $p=.002$). As shown in Figure 15, participants who rarely/never study with background music were more likely to prefer instrumental music to vocal music (*instrumental:* z=1.508; *vocal:* z=-1.870).

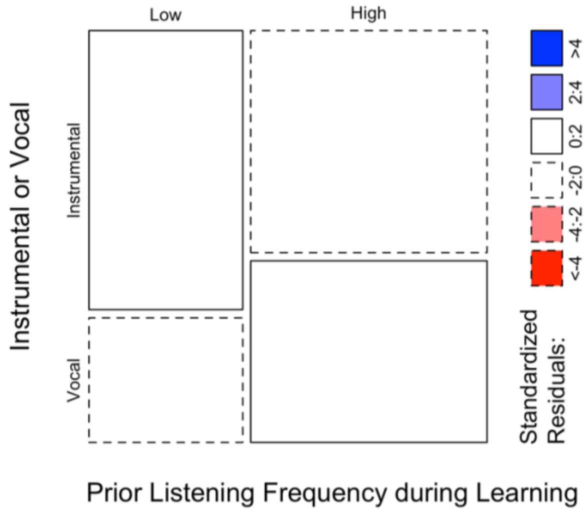


Figure 15. Mosaic plot of prior music listening frequency during learning and preference for instrumental music. a) Tracks liked by both high & low groups were temporarily excluded from the above analyses. b) The resulting number of tracks belong to each group were: high (n=289), low (n=188).

Music Emotion

With respect to the individual difference in terms of preference on music emotion, a set of MANOVAs revealed potential effect of some personality trait dimensions (*extraversion:* F=6.65, $p=.001$; *openness to experience:* F=9.355, $p=.001$) and participants' prior habit of studying with background music (F=6.749, $p=.005$).

As shown in Table 8, the music preferred by introvert, low-openness participants and participants who never/rarely study with background music tended to be relatively neutral, while those favoured by extroverts, high-openness participants and participants who habitually study with background music were slightly more positive. As for the music energy, participants who scored low on each trait dimension were inclined to lower-energy music, while the pieces preferred by their counterpart were found to have relatively higher energy.

Recalling that, as reported in section 4.1.1, music happiness showed larger effect when differentiating playlist and non-playlist tracks, this affective dimension once again showed larger effect in terms of differentiating music collected by habitual music listener as against those collected by participants who never/rarely study with background music (*happiness: b*=3.211, *energy: b*=-1.768). However, interestingly, according to the coefficients of linear discriminants computed from DFA, music energy was found to contribute more in terms of differentiating music collected by introverts versus extroverts (*happiness: b*=0.332, *energy: b*=2.820) and distinguishing music collected by high versus low openness participants (*happiness: b*=-1.234, *energy: b*=4.257).

Table 8

Comparison of Preference on Music Emotion Across User Groups

User Group	F	Happiness			Energy		
		M	SD	W	M	SD	W
Extraversion	6.65^{**}						
Low		0.028	0.429	23863^{**}	-0.249	0.308	22533^{***}
High		0.145	0.422		-0.144	0.323	
Openness	9.36^{**}						
Low		0.057	0.420	25881^{*}	-0.237	0.321	22763^{***}
High		0.134	0.435		-0.132	0.307	
Study with BGM	6.75^{**}						
Never/rarely		0.014	0.432	22936^{**}	-0.226	0.313	24183^{*}
Habitually		0.135	0.420		-0.179	0.322	

Note. a) F = MANOVA test statistic; W = Wilcoxon rank-sum test statistic; BGM = background music. b) Multiple comparison correction was applied based on Benjamini-Hochberg procedure (FDR=0.1). * $p < .05$, ** $p < .01$, *** $p<.001$

Low Level Music Features

Lastly, based on a set of Wilcoxon rank-sum tests, the dynamical and timbral differences of music were only found to be statistically significant, when comparing tracks collected by habitual music listener with tracks collected by participants who normally study without background music.

As shown in Table 9, some loudness and timbre features again showed statistical significance in differentiating music liked by the two user groups. Specifically, recalling that the RMSE features represent the volume of a sound (rmse_mean) and the variation thereof (rmse_std), the presented results suggest that the music selected by participants who never/rarely study with background music were slightly softer and showed less dynamical variation. Timbrally, statistically significant difference could also be found in the spectral distribution of the audio. As specified in section 3.4.3, the roll-off is generally used as an estimation of the amount of high frequency consists in the audio signal, and the spectral centroid is the means of the spectral distribution within each time frame (Tzanetakis & Cook, 2002). Our results suggest that the spectral centroid and roll-off of tracks preferred by participants who normally study without background music were significantly lower, which makes sense, given that human hearing system is sensitive to moderately high frequency band (i.e., 2 to 4 kHz[8]) (Müller, 2015a, p. 26) and participants who were not accustomed to studying with background music might deem music centered around relatively lower frequency band less distracting. Nonetheless, similar to the findings reported in section 4.1.1, participants' rhythmic preference again showed no significant difference across the two user groups.

Table 9

Comparison of Dynamical, Rhythmic, and Timbral Preference Across User Group

Music Features	Study with background music: Never/rarely		Habitually		W
	M	SD	M	SD	
Loudness					
rmse_mean	3.463	1.562	3.862	1.582	22784^{**}
rmse_std	1.633	0.607	1.805	0.648	23004^{**}
Rhythm					
tempo	121.95	29.083	118.66	25.061	28684
avg_onset_frq	3.189	1.168	3.407	1.063	24601
Timbre					
rolloff_mean	3166.3	1135.5	3492.5	1110.8	22787^{**}
rolloff_std	1535.6	600.47	1649.0	502.69	23327^{**}
centroid_mean	1557.3	478.37	1672.3	463.97	23573^{*}
centroid_std	693.56	271.89	739.36	241.85	24056^{*}
flatness_mean	0.010	0.009	0.012	0.011	24367
flatness_std	0.031	0.027	0.031	0.023	26291
zcr_mean	0.071	0.027	0.071	0.025	26972
zcr_std	0.049	0.022	0.051	0.024	26365

Note. a) W denote the Wilcoxon rank-sum test statistic. b) Multiple comparison correction was applied based on Benjamini-Hochberg procedure (FDR=0.1). * p < .05, ** p < .01, *** p<.001

[8] In terms of musical notes, approximately pitch higher than C7.

To sum up, as one may assume, music pieces selected by participants who normally study without background music were indeed found to have lower information load, which implies that tolerance to the information-load characteristics of music might be developed through ones' listening experiences.

4.2 The Effect of Music on Learning

As an extension to the above discussions on learners' music preference, this section further examined the effect of music on learning (i.e., enhance versus distract) and, particularly, its effect on learning engagement.

4.2.1 The Role of Music Characteristics

As the tracks played for each session were normally selected by the MIR system based on certain user-specified music filtering criteria, the music pieces in each session generally possessed relatively consistent characteristics. This allows us to analyse how the general characteristics of music listened during each session correlate with participants' task engagement as well as their perceived learning effect.

Music Genre

The dominant genre (i.e., the genre tag that appeared most in each session) was identified to represent the general music style associated with each session. Table 10 briefly summarized how participants' perception on the learning effect of music (i.e., enhance versus distract) distributed across six major genres.

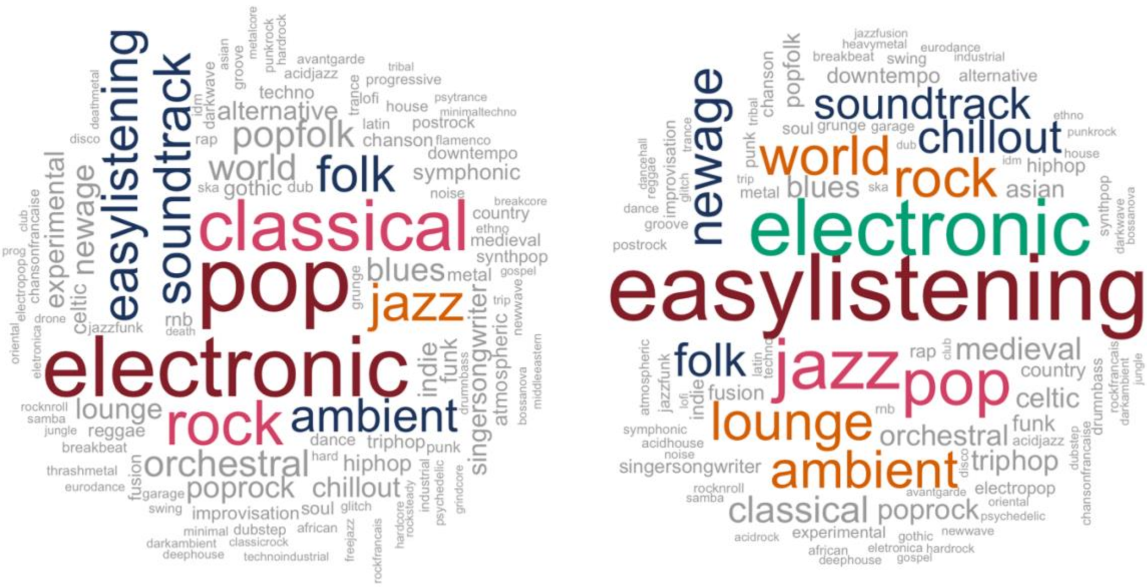
Table 10

Perceived Learning Effect Across Six Major Genres

Genre	Perceived learning effect		
	Distract (1-3)	No effect (4)	Enhance (5-7)
Easy listening	13.89%	13.89%	72.22%
Rock	41.18%	17.65%	41.18%
Pop	45.00%	25.00%	30.00%
Jazz	44.44%	16.67%	38.89%
Electronic	50.00%	27.27%	22.73%
Classical	70.59%	17.65%	11.77%

Note. N=30, L=195 (Genres with frequencies less than 15 were not included). Easy listening: n=36, rock: n=17, pop: n=40, jazz: n=18, electronic: n=22, classical: n=17.

Besides, to triangulate the above statistics derived from the aggregated data, the genre tags word clouds generated from "enhanced" versus "distracted" sessions were plotted to provide some complementary qualitative insights as well (Figure 16).



(a) (b)

Figure 16. (a) Genre tags word cloud of distracted sessions; (b) Genre tags word cloud of enhanced sessions

As shown in Table 10 and Figure 16, unlike the diversified reports from sessions with other genres (i.e., rock, pop, jazz, electronic), sessions with easy listening music mostly received positive report on perceived learning effect (*enhance:* 72.22%, *distract:* 13.89%). This observation is consistent with the common impression that easy listening music is generally preferable for listening while studying.

Interestingly, despite of its popularity in our user-generated playlists and the so-called Mozart Effect reported in the research literature (Schellenberg, 2012), among sessions with classical music as primary genre, the negative effect of music on learning performance was reported at a higher percentage (*enhance:* 11.77%, *distract:* 70.59%). Some possible explanations of this phenomenon might be: 1) Classical music normally possesses more complexity in its musical structure (Wright, 2011, p. 84) which might increase the cognitive load of learners; 2) Learners' personal music taste also plays an important role in the interaction between background music and learning. Quantitatively, the repeated measures correlation revealed a significant positive relationship between the averaged music ratings of each session and participants' perceived learning effect (r=.450, p<.001). Qualitatively, just as a participant stated in his post-interview: *"I did try some classical music. But I have to admit that it does not fit my taste, and I don't know why … I just can't immerse myself in it." (P25)*

Nonetheless, it is noteworthy that the above discussions are primarily qualitative, as our unbalanced repeated measures study design restricted the available choices of inferential statistical analysis. Further research is needed to ascertain whether the observed phenomenon implies significant relationship, preferably with larger sample size.

Music Emotion

The correlation analysis between the learning outcome measures and the averaged happiness and energy level of music associated with each session revealed some interesting results as well. Particularly, the repeated measures correlation was applied using Bakdash and Marusich (2017)'s R package implementation to model the intra-individual association, with multiple comparison correction conducted based on Benjamini-Hochberg procedure (FDR=0.1).

Table 11

The Effect of Music Emotion on Learning

Music Emotion	Repeated Measurers Correlation Coefficient		
	Perceived learning effect	Altered sense of time	Concentration
Happiness	.091	.159*	.152*
Energy	.127	.196*	.183*

Note. N=30, L=195. a) Multiple comparison correction was applied based on Benjamini-Hochberg procedure (FDR=0.1). * $p < .05$, ** $p < .01$, *** $p<.001$. b) Concentration = total concentration on the task at hand

As shown in Table 11, the happiness and energy of music were both found to be positively related to the engagement-related constructs (i.e., altered sense of time, total concentration on the task at hand)[9], while neither showed significant relationship with participants' perceived learning effect (i.e., enhance versus distract). The above observations suggest that the effect of music emotion might be primarily engagement-related rather than a direct effect on the learning performance per se. Just as a participant's comment on the reason for studying with background music: *"Honestly, I believe that music would more or less distract me and decrease my efficiency anyway. Actually, the reason I prefer studying with background music is it can help me enter the state of learning and sustainably engage in learning." (P29)*

[9] Recalling that both engagement-related constructs depict a special absorbing experience (see section 3.4.2), the *total concentration on the task at hand* referred to ones' intense focus on their task being performed at the moment, while the *altered sense of time* generally refers to the distortion of temporal awareness due to the intense involvement in the task at hand (Csikszentmihalyi, 2014, p. 240).

Besides, it’s noteworthy that the magnitude of correlation coefficients also indicate that the strength of the relationships is relatively weak. One possible reason might be, as suggested by the Yerkes and Dodson (1908)’s law, the relationship under discussion might not be a perfect linear one but probably a quadratic one (e.g., an inverted U pattern). Nonetheless, future work is needed to testify this assumption.

Low Level Music Features

Lastly, the mean values of low-level music features were calculated to represent the central tendency of music characteristics of each session as well. However, unlike music emotion, the dynamical, rhythmic, and timbral music features did not reveal significant relationship with participants’ task engagement or perceived learning effect (Table 12).

Table 12

The Effect of Dynamical, Rhythmic, and Timbral Music Features on Learning

Music Features	Repeated Measurers Correlation Coefficient		
	Perceived learning effect	Altered sense of time	Concentration
Loudness			
rmse_mean	.060	.157	.040
rmse_std	.023	.130	.035
Rhythm			
tempo	-.154	-.116	-.018
avg_onset_frq	-.028	.028	-.003
Timbre			
rolloff_mean	.098	.168	.166
rolloff_std	-.064	-.079	-.050
centroid_mean	.089	.151	.159
centroid_std	-.092	-.084	-.049
flatness_mean	.080	.090	.120
flatness_std	-.035	-.045	-.022
zcr_mean	.063	.103	.121
zcr_std	-.081	-.087	-.032

Note. N=30, L=195. a) Multiple comparison correction was applied based on Benjamini-Hochberg procedure (FDR=0.1). * p < .05, ** p < .01, *** p<.001. b) Concentration = total concentration on the task at hand

Based on the theoretical assumption that learners' task load and cognitive ability might confound the effect of information-load music characteristics on learning, this study further investigate the correlation between low-level music features and learning outcome measures on the partitioned datasets in the subsequent sections.

4.2.2 The Role of Learning Context

As specified in the conceptual framework (section 2.3), learners' emotional status and task load were identified as the major contextual factors under investigation. Specifically, participants' emotional status was measured on two affective dimensions (i.e., valence and arousal) using a set of continuums (-1~1) (Russell et al., 1989). And the task load was operationalized as the mental demand[10] and temporal demand[11] of learning task which was adapted from the NASA task load index (Hart & Staveland, 1988). Table 13 summarized the role of the aforementioned contextual factors in general.

Table 13

The General Role of Contextual Factors

Contextual Factors	Repeated Measurers Correlation Coefficient		
	Perceived learning effect	Altered sense of time	Concentration
chg_valence	.248**	.072	.124
chg_arousal	.067	.169*	.080
mental demand	.056	.155*	.023
temporal demand	.044	.156*	.029

Note. N=30, L=195. a) Multiple comparison correction was applied based on Benjamini-Hochberg procedure (FDR=0.1). * p < .05, ** p < .01, *** p<.001. b) Concentration = total concentration on the task at hand, chg_valence/chg_arousal = change of valence/arousal before and after each session.

As shown in Table 13, participants' self-reported learning effect was found to positively correlate with the change of valence before and after the learning session (r=.248, p=.001), while no significant relationship was observed between participants' task load and their perceived effect of music on learning (i.e., enhance versus distract). The above observations indicate that: 1) as suggested by the arousal-mood-hypothesis, one of the most important ways

[10] Mental demand refers to the amount of mental or perceptual activity required (e.g., calculating, remembering, thinking).

[11] Temporal demand refers to the time pressure felt due to the task (i.e., slow and leisurely versus rapid and hurried)

in which music benefits learning is closely related to mood modulation and mood enhancement; 2) music might not necessarily exert detrimental effect on learning even if the learning task was mentally or temporally demanding.

Besides, the analysis on the *altered sense of time* construct also revealed some interesting findings. As shown in Table 13, this engagement-related construct was found to positively correlate with the mental and temporal demand of learning task as well as the increase of arousal. Recalling our discussions on the challenge-skill balance hypothesis in flow theory (section 2.2.2), the above results also suggest that mentally or temporally demanding task (i.e., high level of challenge) might lead to high level of arousal and, ultimately, high level of involvement in the task at hand. We will review this hypothesis in the subsequent discussions.

Based on the assumption that learners' task load might exert moderating effect on the interaction between music and learning, the correlation analysis between music features and learning outcome measures was further conducted on the dataset partitioned by the mental and temporal demand of the learning task as well (Table 14 and Table 15).

Table 14

Repeated Measures Correlation on Dataset Partitioned by Mental Demand

Music Features	Low mental demand (L=65)			High mental demand (L=84)		
	PLE	AST	CONC	PLE	AST	CONC
Loudness						
rmse_mean	-.016	.204	.197	.152	.122	-.045
rmse_std	-.039	.218	.218	.140	.047	.055
Rhythm						
tempo	$-.344^{*}$	-.266	-.203	-.161	$-.383^{**}$	-.046
avg_onset_frq	-.169	-.006	-.032	.014	.119	.158
Timbre						
rolloff_mean	-.112	.259	.198	.284	.176	.105
rolloff_std	$-.355^{*}$	-.153	-.196	.232	-.028	.215
centroid_mean	-.156	.204	.149	.280	.164	.089
centroid_std	$-.458^{**}$	-.142	-.172	.211	-.063	.127
flatness_mean	-.228	.101	.128	.315	.112	.063
flatness_std	$-.373^{*}$	-.109	-.084	.114	-.094	-.092
zcr_mean	-.212	.023	-.031	.204	.154	.015
zcr_std	$-.426^{**}$	-.112	-.134	.129	-.143	-.022

Note. N=30. a) Benjamini-Hochberg procedure (FDR=0.1) was applied. * $p < .05$, ** $p < .01$, *** $p<.001$. b) PLE = perceived learning effect, AST = altered sense of time, CONC = total concentration on the task at hand

Table 15

Repeated Measures Correlation on Dataset Partitioned by Temporal Demand

Music Features	Low temporal demand (L=100)			High temporal demand (L=95)		
	PLE	AST	CONC	PLE	AST	CONC
Loudness						
rmse_mean	.045	.150	.141	.167	.203	.066
rmse_std	-.048	.186	.107	.155	.134	.072
Rhythm						
tempo	-.212	.032	.018	-.102	-.151	-.021
avg_onset_frq	-.029	-.012	-.042	.050	-.038	.034
Timbre						
rolloff_mean	.045	.333**	.248	.176	.013	.125
rolloff_std	-.040	-.029	-.125	.074	-.181	.082
centroid_mean	.032	.320**	.256	.171	.002	.108
centroid_std	-.140	.012	-.096	.088	-.215	.026
flatness_mean	-.020	.216	.204	.223	.034	.107
flatness_std	-.182	-.003	-.035	.167	-.027	.003
zcr_mean	-.002	.264*	.230	.114	-.005	.044
zcr_std	-.213	.067	-.029	.082	-.217	-.021

Note. N=30. a) Benjamini-Hochberg procedure (FDR=0.1) was applied. * $p < .05$, ** $p < .01$, *** $p<.001$. b) PLE = perceived learning effect, AST = altered sense of time, CONC = total concentration on the task at hand

As shown in Table 14 and Table 15, interestingly, under high mental or temporal demand, most music features did not show significant effect, except that the tempo was found to negatively correlate with the *altered sense of time* when the task was mentally demanding (r=-.383, p=.002). On the contrary, under low task load, the correlation analysis revealed significant association between several timbre features and the learning outcome measures.

Recalling our previous discussions on the challenge-arousal-involvement hypothesis, one possible explanation of this phenomenon might be: 1) Tasks of high mental or temporal demand might increase learners' arousal level and task involvement. Learners' attention might mainly be on the task per se and the effect of music would, consequently, be less clear. 2) Conversely, under low mental or temporal demand, leaners might pay more attention to the background music and thus would be more likely to be influenced by the timbre quality of music. One participant's comment on the moderating effect of task type also, in a sense, corroborated this assumption: *"I thought that we might as well categorize the tasks into three types…I mean…something like a pyramid. First, you got two extremes: for some top-level tasks like coding or painting, you could easily get yourself involved and the external factors, including music, also couldn't easily distract you; for some low-level tasks in another extreme, like filling a form or other routine affairs, they are less brain-intensive and background music might even have some positive effect for stuff like these. However, for some middle-level tasks…something requires certain amount of brain-work and could easily be distracted, background music could be really distracting. "(P15)* Besides, just as another participant stated in his post-interview: *"You know, when you were immersed in your task, you would literally leave the music in the background … you just forgot that there's a song playing." (P21)*

Nonetheless, more work and more rigorous experiment design is needed to testify the above assumptions. For instance, as suggested by the challenge-skill hypothesis in flow theory, to enter the state of flow, ones' skill have to match the challenge they face (Figure 17). However, as a participant pointed out in the post-interview: *"I have to admit that, though I was trying to help, but, when I got a really demanding task and I believed that I must use 100% focus to get it done, I wouldn't start a session on this music App." (P7)* Hence, as an experiment in naturalistic setting, it's possible that participants might avoid conducting a session when the task challenge is high (e.g., high mental demand) while their skill level did not match. This tendency could thus decrease the chance of observing the effect of music in the high-challenge-low-skill situation. Besides, it's also worthwhile investigating the effect of music with consideration of the actual mental activity involved, e.g., the effect of music on creative versus comprehensive task might be different.

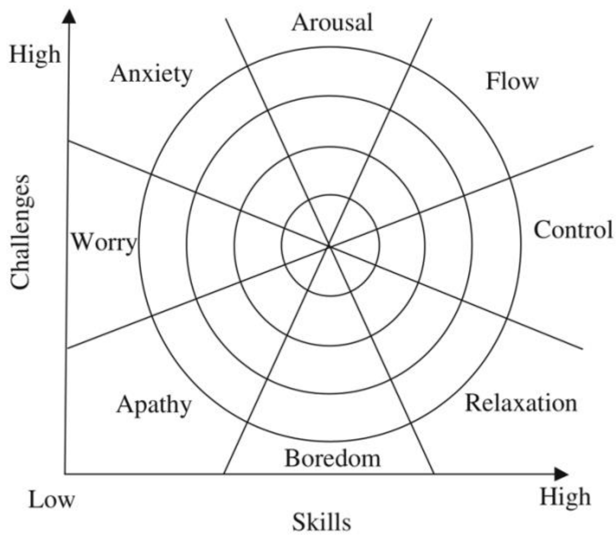


Figure 17. Quality of experience as a function of the challenge-skill balance. (Csikszentmihalyi, 2014, p. 248)

4.2.3 The Role of Learners' Traits

To further probe the potential moderating effect of learners' cognitive ability, the correlation analysis between music features and learning outcome measures was conducted on the dataset partitioned by participants' working memory capacity and multitasking ability as well (Table 16 and Table 17).

Table 16

Repeated Measures Correlation on Dataset Partitioned by Working Memory Capacity

Music Features	Working Memory Capacity					
	Low (N=15, L=95)			High (N=15, L=100)		
	PLE	AST	CONC	PLE	AST	CONC
Loudness						
rmse_mean	.007	.304**	-.001	.105	.068	.078
rmse_std	-.002	.171	-.003	.043	.108	.069
Rhythm						
tempo	-.030	.125	.183	-.261	-.283**	-.211
avg_onset_frq	-.132	-.039	-.024	.057	.073	.017
Timbre						
rolloff_mean	.142	.334**	.267**	.061	.062	.049
rolloff_std	.012	-.094	-.021	-.129	-.072	-.077
centroid_mean	.163	.340**	.304**	.023	.023	.012
centroid_std	.031	-.022	.050	-.203	-.133	-.150
flatness_mean	.176	.233**	.276*	-.027	-.032	-.083
flatness_std	.088	.022	.139	-.133	-.089	-.168
zcr_mean	.221	.359***	.326**	-.088	-.087	-.102
zcr_std	.116	.057	.143	-.255	-.191	-.206

Note. a) Benjamini-Hochberg procedure (FDR=0.1) was applied. * $p < .05$, ** $p < .01$, *** $p<.001$.
b) PLE = perceived learning effect, AST = altered sense of time, CONC = total concentration on the task at hand

As shown in Table 16, for the sessions conducted by participants with high working memory capacity (WMC), most music features did not show significant effect, while, for their counterpart, several timbre features consistently showed significant relationship with the two engagement-related constructs (i.e., altered sense of time, total concentration on the task at hand). The above observations imply that learners with low WMC might be more sensitive to the timbre quality of music, compared to those with high WMC.

Table 17

Repeated Measures Correlation on Dataset Partitioned by Multitasking Ability

Music Features	Multitasking Ability					
	Low (N=15, L=79)			High (N=15, L=116)		
	PLE	AST	CONC	PLE	AST	CONC
Loudness						
rmse_mean	-.017	.328**	.073	.098	.082	.018
rmse_std	.021	.255	.152	.024	.070	-.052
Rhythm						
tempo	-.003	.085	.205	-.239	-.220	-.183
avg_onset_frq	.064	.068	.110	-.090	.006	-.104
Timbre						
rolloff_mean	.179	.362**	.372**	.060	.084	.036
rolloff_std	.086	-.016	-.005	-.153	-.116	-.085
centroid_mean	.199	.360**	.387**	.037	.060	.013
centroid_std	.077	.045	.070	-.190	-.154	-.141
flatness_mean	.234	.249*	.312*	-.022	-.007	-.060
flatness_std	.067	.026	.072	-.090	-.081	-.090
zcr_mean	.239	.345**	.411***	-.016	.002	-.055
zcr_std	.097	.076	.165	-.175	-.168	-.172

Note. a) Benjamini-Hochberg procedure (FDR=0.1) was applied. * $p < .05$, ** $p < .01$, *** $p<.001$.
b) PLE = perceived learning effect, AST = altered sense of time, CONC = total concentration on the task at hand

Similarly, as shown in Table 17, the aforementioned pattern was also found when controlling for the effect of multitasking ability. In the low multitasking ability group, several timbre features were again found to significantly correlate with the engagement-related constructs, while, in high multitasking ability group, none of the music features showed significant effect.

To sum up, the timbre feature showed significant effect for participants who scored low in the cognitive ability assessment (i.e., n-back test, multitasking test), but no significant effect was observed for participants with high cognitive ability. However, such pattern was not observed in terms of rhythm and dynamics.

Chapter 5: Conclusion

5.1 Summary and Key Findings

This study aims to explore how background music befits learning through an experiment in naturalistic setting. A one-week field experiment was conducted in participants' own study places. During the experiment, participants were asked to conduct learning sessions with music in the background and collect the tracks they deemed suitable for learning using a novel mobile music app (i.e., Moody App). A set of participant-related, context-related, and music-related data were collected via the pre-experiment survey as well as the logging system and survey system of the music app.

Our findings revealed that, generally, learners might prefer unpretentious and sophisticated music (e.g., pop, classical, folk) to intense and contemporary music (e.g., rock, punk, hip-hop/rap), though participants who scored high in *openness to new experience* were found to less stick to the mainstream unpretentious music (e.g., pop, country) but were more willing to try instrumental music in other genre. Besides, there's no general tendency of the preference for mellow music (e.g., electronic, new age, world), though the genre tags cloud indicated its popularity in participants' personal study music playlist and preference for such style was found to be significantly associated with the level of working memory capacity. Finally, in terms of the learning effect of music genre, our findings suggest that easy listening music was generally deemed as enhancing learning, while, among sessions with classical music as primary genre, the negative effect of music on learning performance was reported at a higher percentage. However, it's noteworthy that classical music was actually one of the top 3 popular genres in participants study music playlist, which indicate the possibility that only a subgroup (sub-genre) of classical music was deemed preferable for learning. Future study is needed to further examine this phenomenon. On the other hand, there was no general tendency in terms of the preference of music type (i.e., instrumental or vocal), though some interesting pattern emerged when participants' personality traits, etc. were considered.

Emotionally, our findings revealed that pleasant and soothing music are more likely to suit learners' music preference, which is generally consistent with the arousal-mood-hypothesis. With respect to the individual difference, personality traits on extraversion and openness dimensions as well as prior habit of studying with background music was found to moderate participants' preference for the happiness and energy level of music. Most importantly, our findings corroborated that one of the most important ways in which music benefits learning is closely related to mood modulation and mood enhancement, and both affective dimensions of music emotion were found to significantly correlate with learning engagement.

In regard to the low-level music features, generally, the playlist and non-playlist tracks were found to be dynamically and timbrally different, but the rhythm features seemed to play a less important role in distinguishing the two music groups. Moreover, our results also suggest that music pieces selected by participants who normally study without background music were indeed found to have lower information load, which implies that tolerance to the information-load characteristics of music might be developed through ones' listening experiences. In terms of how dynamical, rhythmic, and timbral features play a role in the interaction between background music and listeners' learning experience, as consistent with the literature, no overall effect was observed. However, in view of the moderating effect of task load and learners' cognitive ability, the timbre quality of music showed significant effect in low mental/temporal condition and in participant groups with low working memory capacity/multitasking ability. The skill-challenge hypothesis in flow theory and the irrelevant sound effect might be able to account for these phenomena, respectively.

5.2 Implications

Our findings suggest that there might be complex interactions among music, learners, and learning task. For future research on relevant topic, our suggestion is that systematic integration and manipulation of such participant-related, task-related, and music-related factors would be important so as to control for the confounding effect and produce more generalizable results.

Moreover, the findings also provide a series of implications for designing music recommendation system in educational setting, including: 1) the music styles and genres that generally deemed suitable for learning; 2) the general characteristics of preferable study music in terms of music emotion as well as dynamical, rhythmic, and timbral characteristics. 3) the importance of personalized and context-aware music recommendation.

5.3 Limitations and Future Work

Nonetheless, this study still has some limitations to be noted:

1) *Sample size and sample representativeness*. Overall, the sample size of this study is still relatively small, and this issue is even more worthy of attention in the group-wise analyses. Besides, there is also a lack of diversity in terms of participants' cultural background. For instance, the contemporary music (e.g., soul, R&B, funk) are

primarily African American music, and hence it’s possible that its popularity in Asian population could, in a sense, bias our results. In other word, the findings on style preference might not be able to be generalized to learners with other cultural background.

2) *Trustworthiness of self-reported measures*. The measurement of learning effect of music, learning engagement, emotional status, etc. were mostly based on participants’ self-report measures and caution is needed for its subjectivity and reliability.
3) *Interpretability of low-level music features*. Though our analysis on the proposed acoustic features did reveal some interesting results, its nature of being low-level representation of audio samples still limits its interpretability. Future work might as well further incorporate higher-level music complexity measures such as structural repetitiveness, homogeneity, and novelty distribution through music structure analysis (Müller, 2015b).

Given the large scale of longitudinal data, for future work, interactions among music characteristics, task load, learners’ traits could be further probed through association rule mining and predictive modeling in a data-intensive manner. Besides, the present study only investigated the association between music characteristics and listeners’ learning experience in general using aggregated data. For more fine-grained analysis, future study could, for instance, investigate the association between timbral/rhythmic novelty and skip event in a frame-wise manner. Finally, as mentioned in our previous discussion, it’s also worthwhile investigating the effect of music with consideration of 1) the eight challenge-skill balance conditions (e.g., anxiety, arousal, flow, relaxation, boredom); 2) the actual mental activity involved, e.g., the effect of music on creative versus comprehensive task might be different.

References

Admiraal, W., Huizenga, J., Akkerman, S., & Dam, G. t. (2011). The concept of flow in collaborative game-based learning. *Computers in Human Behavior, 27*(3), 1185-1194. doi:https://doi.org/10.1016/j.chb.2010.12.013

Angel, L., Polzella, D., & Elvers, G. (2010). Background Music and Cognitive Performance. *Perceptual and Motor Skills, 110*(3), 1059.

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated Measures Correlation. *Frontiers in Psychology, 8*, 456-456. doi:10.3389/fpsyg.2017.00456

Balch, W. R., & Lewis, B. S. (1996). Music-Dependent Memory: The Roles of Tempo Change and Mood Mediation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1354-1363. doi:10.1037/0278-7393.22.6.1354

Beaman, C. P. (2005). Auditory distraction from low-intensity noise: a review of the consequences for learning and workplace environments. *Applied Cognitive Psychology, 19*(8), 1041-1064. doi:10.1002/acp.1134

Bonneville-Roussy, A., Rentfrow, P., Xu, M., & Potter, J. (2013). Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood. *Journal of Personality and Social Psychology, 105(4)*, 703-717.

Boyle, R., & Coltheart, V. (1996). Effects of irrelevant sounds on phonological coding in reading comprehension and short-term memory. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 49*(2), 398-416.

Burunat, I., Alluri, V., Toiviainen, P., Numminen, J., & Brattico, E. (2014). Dynamics of brain activity underlying working memory for music in a naturalistic condition. *Cortex, 57*, 254-269.

Christopher, E. A., & Shelton, J. T. (2017). Individual Differences in Working Memory Predict the Effect of Music on Student Performance. *Journal of Applied Research in Memory and Cognition, 6*(2), 167-173.

Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life*. NY, US: Basic Books.

Csikszentmihalyi, M. (2014). *Flow and the foundations of positive psychology: the collected works of Mihaly Csikszentmihalyi*. Berlin: Springer.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual Differences in Working Memory Capacity and What They Tell Us About Controlled Attention, General Fluid Intelligence, and Functions of the Prefrontal Cortex. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 102-134). Cambridge: Cambridge University Press.

Etaugh, C., & Ptasnik, P. (1982). Effects of studying to music and post-study relaxation on reading comprehension. *Perceptual and Motor Skills, 55*(1), 141-142.

Ferreri, L., & Verga, L. (2016). Benefits of Music on Verbal Learning and Memory: How and When Does It Work? *Music Perception: An Interdisciplinary Journal, 34*(2), 167-182.

Fitzgerald, D., & Paulus, J. (2006). *Unpitched Percussion Transcription*. Boston, MA: Boston, MA: Springer US.

Furnham, A., & Strbac, L. (2002). Music is as distracting as noise: the differential distraction of background music and noise on the cognitive test performance of introverts and extraverts. *Ergonomics, 45*(3), 203-217. doi:10.1080/00140130210121932

Gosling, S. D., Rentfrow, P. J., & Potter, J. (2014). *Norms for the Ten Item Personality Inventory*. Unpublished Data.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality, 37*(6), 504-528.

Hainsworth, S. (2006). *Beat Tracking and Musical Metre Analysis*. Boston, MA: Boston, MA: Springer US.

Hallam, S., Cross, I., & Thaut, M. (2016). *The Oxford handbook of music psychology* (2nd ed.). Oxford: Oxford University Press.

Hallam, S., Price, J., & Katsarou, G. (2002). The Effects of Background Music on Primary School Pupils' Task Performance. *Educational Studies, 28*(2), 111-122. doi:10.1080/03055690220124551

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170-179. doi:https://doi.org/10.1016/j.chb.2015.07.045

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139-183): North-Holland.

Herrera-Boyer, P., Klapuri, A., & Davy, M. (2006). *Automatic Classification of Pitched Musical Instrument Sounds*. Boston, MA: Boston, MA: Springer US.

Hu, X., Lee, J. H., Bainbridge, D., Choi, K., Organisciak, P., & Downie, J. S. (2017). The MIREX grand challenge: A framework of holistic user-experience evaluation in music information retrieval. *Journal of the Association for Information Science Technology, 68*(1), 97-112.

Hu, X., Li, F., & Kong, R. (2019). *Can Background Music Facilitate Learning?: Preliminary Results on Reading Comprehension*. Paper presented at the Proceedings of the 9th International Conference on Learning Analytics & Knowledge, Tempe, AZ, USA.

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of Musical Tempo and Mode on Arousal, Mood, and Spatial Abilities. *Music Perception: An Interdisciplinary Journal, 20*(2), 151-171. doi:10.1525/mp.2002.20.2.151

Jackson, S., Eklund, B., & Martin, A. (2012). The Manual for the Flow Scales.

Jäncke, L., Brügger, E., Brummer, M., Scherrer, S., & Alahmadi, N. (2014). Verbal learning in the context of background music: no influence of vocals and instrumentals on verbal learning. *Behavioral and brain functions : BBF, 10*(1), 10-10. doi:10.1186/1744-9081-10-10

Jäncke, L., & Sandmann, P. (2010). Music listening while you learn: No influence of background music on verbal learning. *Behavioral and Brain Functions, 6*(1), 3.

Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2010). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music, 39*(4), 424-448. doi:10.1177/0305735610376261

Kantner, J. (2009). Studying with music: Is the irrelevant speech effect relevant? In *Applied Memory* (pp. 19-40). NY, US: Nova Science Publishers.

Kiger, D. M. (1989). Effects of Music Information Load on a Reading Comprehension Task. *Perceptual and Motor Skills, 69*(2), 531-534.

Küssner, M. B. (2017). Eysenck's Theory of Personality and the Role of Background Music in Cognitive Task Performance: A Mini-Review of Conflicting Findings and a New Perspective. *Frontiers in Psychology, 8*(1991). doi:10.3389/fpsyg.2017.01991

Lehmann, J. A. M., & Seufert, T. (2017). The Influence of Background Music on Learning in the Light of Different Theoretical Perspectives and the Role of Working Memory Capacity. *8*(1902). doi:10.3389/fpsyg.2017.01902

Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). *Neuropsychological assessment*: Oxford University Press, USA.

Matney, B. (2017). The effect of specific music instrumentation on anxiety reduction in university music students: A feasibility study. *Arts in Psychotherapy, 54*, 47-55. doi:10.1016/j.aip.2017.02.006

Mayfield, C., & Moss, S. (1989). Effect of Music Tempo on Task Performance. *Psychological Reports, 65*(3_suppl2), 1283-1290.

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015). *librosa: Audio and music signal analysis in python.* Paper presented at the Proceedings of the 14th python in science conference.

Müller, M. (2015a). *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications* (2015 ed.). Cham: Cham: Springer International Publishing.

Müller, M. (2015b). Music Structure Analysis. In M. Müller (Ed.), *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications* (pp. 167-236). Cham: Springer International Publishing.

Munzel, U., & Brunner, E. (2000). Nonparametric Tests in the Unbalanced Multivariate One-Way Design. *Biometrical Journal, 42*(7), 837-854.

Parmer, T., & Ahn, Y.-Y. (2019). *Evolution of the Informational Complexity of Contemporary Western Music*. Paper presented at the the 20th annual conference of the International Society for Music Information Retrieval (ISMIR), Delft, The Netherlands.

Perham, N., & Currie, H. (2014). Does listening to preferred music improve reading comprehension performance? *Applied Cognitive Psychology, 28*(2), 279-284.

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review, 7*(3), 216-237.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: a single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57*(3), 493.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition, 34*(2), 344-355. doi:10.3758/BF03193412

Schellenberg, E. G. (2012). Cognitive performance after listening to music: A review of the Mozart effect. *Music, health, and wellbeing*, 324-338.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*(4), 695-729. doi:10.1177/0539018405058216

Schnotz, W., & Kürschner, C. (2007). A Reconsideration of Cognitive Load Theory. *Educational Psychology Review, 19*(4), 469-508. doi:10.1007/s10648-007-9053-4

Shernoff, D. J., & Csikszentmihalyi, M. (2009). Cultivating engaged learners and optimal learning environments. *Handbook of positive psychology in schools*, 131-145.

Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E. S. (2014). Student Engagement in High School Classrooms from the Perspective of Flow Theory. In M. Csikszentmihalyi (Ed.), *Applications of Flow in Human Development and Education: The Collected Works of Mihaly Csikszentmihalyi* (pp. 475-494). Dordrecht: Springer Netherlands.

Stoet, G., O'Connor, D. B., Conner, M., & Laws, K. R. (2013). Are women better than men at multi-tasking? *BMC Psychology, 1*(1), 18. doi:10.1186/2050-7283-1-18

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing, 10*(5), 293-302. doi:10.1109/TSA.2002.800560

Webster, G., & Weir, C. (2005). Emotional Responses to Music: Interactive Effects of Mode, Texture, and Tempo. *Motivation and Emotion, 29*(1), 19-39. doi:10.1007/s11031-005-4414-0

Whitely, P. L. (1934). The Influence of Music on Memory. *The Journal of General Psychology, 10*(1), 137-151. doi:10.1080/00221309.1934.9917718

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4*, 433-433. doi:10.3389/fpsyg.2013.00433

Wright, C. M. (2011). *Listening to music* (6th ed.). Boston, Mass.: Schirmer/Cengage Learning.

Xiao, H., & Yi-Hsuan, Y. (2017). Cross-Dataset and Cross-Cultural Music Mood Prediction: A Case on Western and Chinese Pop Songs. *IEEE Transactions on Affective Computing, 8*(2), 228-240. doi:10.1109/TAFFC.2016.2523503

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*(5), 459-482. doi:10.1002/cne.920180503

Appendices

Appendix 1: Pre-experiment Questionnaire

1. Name (Chinese or English): ___________
2. Gender: ____________________________
3. HKU Email: ________________________
4. Contact Number: ____________________
5. Year of Birth: _______________________
6. Major: _____________________________

(Please list all that apply and feel free to use Chinese or English. E.g., “Master of Education”)

7. How frequently do you listen to music in your daily life?

   A. Seldom

   B. Less than once a month

   C. Approximately once a month

   D. Several times a month

   E. Several times a week

   F. Approximately once a day

   G. Several times a day

   H. Almost always

8. How frequently do you listen to music while *learning*?

   A. Never

   B. Rarely

   C. Sometimes

   D. Usually

   E. Almost always

9. I would like to listen to music when I am ______ (Check all that apply)

   ☐ Excited

   ☐ Happy

   ☐ Pleased

   ☐ Sleepy

   ☐ Bored

   ☐ Depressed

☐ Frustrated

☐ Annoyed

☐ Other: __________________

10. According to your experience, does music change your mood?

A. Never

B. Rarely

C. Sometimes

D. Usually

E. Almost always

11. Have you ever received formal music training (e.g., vocal or instrumental music lessons)?

A. Yes
B. No
C. Other: __________________

12. Which genre do you prefer in daily music listening? (Check all that apply)

☐ Alternative (另類搖滾)

☐ Blues (藍調)

☐ Children's (兒童音樂)

☐ Classical (古典音樂)

☐ Country (鄉村音樂)

☐ Dance (舞曲)

☐ Easy listening (輕音樂)

☐ Electronica (電子音樂)

☐ Folk (民謠)

☐ Gospel (福音音樂)

☐ Hard Rock/Heavy Metal (重金屬搖滾)

☐ Hip hop/Rap (嘻哈說唱)

☐ Jazz (爵士樂)

☐ Latin (拉丁音樂)

☐ New age (新世紀音樂)

☐ Oldies (懷舊音樂)

☐ Pop (流行音樂)

☐ Reggae (雷鬼)

☐ R&B/Soul (節奏布魯斯/靈魂樂)

☐ Rock (搖滾)

☐ Opera/Vocal (歌劇)

☐ World (世界音樂)

☐ Other

Appendix 2: Ten Item Personality Inventory (TIPI)

Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

1 = Disagree strongly

2 = Disagree moderately

3 = Disagree a little

4 = Neither agree nor disagree

5 = Agree a little

6 = Agree moderately

7 = Agree strongly

I see myself as:

1. _____ Extraverted, enthusiastic.
2. _____ Critical, quarrelsome.
3. _____ Dependable, self-disciplined.
4. _____ Anxious, easily upset.
5. _____ Open to new experiences, complex.
6. _____ Reserved, quiet.
7. _____ Sympathetic, warm.
8. _____ Disorganized, careless.
9. _____ Calm, emotionally stable.
10. ____ Conventional, uncreative.

TIPI scale scoring ("R" denotes reverse-scored items):
Extraversion: 1, 6R; Agreeableness: 2R, 7; Conscientiousness; 3, 8R; Emotional Stability: 4R, 9; Openness to Experiences: 5, 10R.

(Samuel D Gosling et al., 2003)

Appendix 3: Post-Interview Protocol

1. How do you like studying with music in the background?

2. Which types of music do you prefer while learning (in terms of genre, tempo, music emotion, etc.)? Why?
   (Present participants' playlist)
   Why did you think these pieces are preferable for learning?

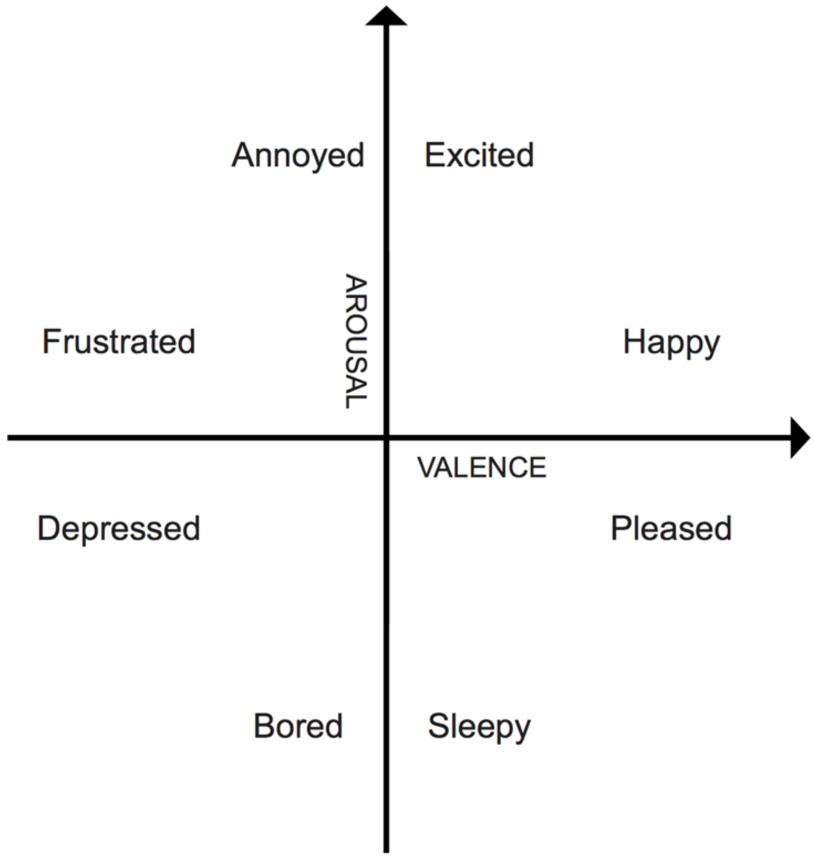
3. Do you prefer to different types of music in different listening scenario (e.g., walking vs. learning)? How does the music you preferred while learning differ from the music you liked for other listening scenarios?

4. Do you have any suggestions for the Moody app?

5. Do you have any suggestions for the experiment procedure?

Appendix 4: Pop-Up Survey

❖ **Emotional Status**



(click to choose a mood category in the affect grid.)

level of valence

Very unpleasant **Very pleasant**

-1 1

level of arousal

Very calm **Very energetic**

-1 1

❖ **Task Load**

1) Please briefly describe your learning task.

2) How mentally demanding is this task?

Very low **Very high**

1 7

2) How hurried or rushed is the pace of this task?

Very low **Very high**

1 7

❖ **Post-Task Survey**

Total concentration on the task at hand

1) My attention was focused entirely on what I was doing

Strongly disagree **Strongly agree**

1 7

2) I had total concentration.

Strongly disagree **Strongly agree**

1 7

3) I was completely focused on the task at hand

Strongly disagree **Strongly agree**

1 7

Altered sense of time

1) Time seemed to alter (either slowed down or speeded up)

Strongly disagree						**Strongly agree**
1						7

2) It felt like time went by quickly

Strongly disagree						**Strongly agree**
1						7

3) I lost my normal awareness of time

Strongly disagree						**Strongly agree**
1						7

Perceived Learning Effect of Music

To what extent did the music affect your performance on this task?

1						7

1 = Very much distracted me

2 = Moderately distracted me

3 = Slightly distracted me

4 = Had no effect

5 = Slightly enhanced my work

6 = Moderately enhanced my work

7 = Very much enhanced my work

Appendix 5: Genre Options for Music Filtering

☐ Classical	☐ Orchestral
☐ New Age	☐ World
☐ Jazz	☐ Bossanova
☐ Swing	☐ Easy Listening
☐ Lounge	☐ Ambient
☐ Atmospheric	☐ Electronic
☐ Chill-out	☐ House
☐ Downtempo	☐ Drum & Bass
☐ Folk	☐ Country
☐ Pop	☐ Rock
☐ Blues	☐ Rhythm & Blues
☐ Hip hop/Rap	☐ Metal
☐ Punk	☐ Reggae
☐ Alternative	☐ Experimental