Exploring the Effect of Personalized Background Music on Reading Comprehension

Ying Que
Faculty of Education
University of Hong Kong
Hong Kong S.A.R
yingque@connect.hku.hk

Yueyuan Zheng
Department of Psychology
University of Hong Kong
Hong Kong S.A.R
u3514160@connect.hku.hk

Janet H. Hsiao
Department of Psychology
University of Hong Kong
Hong Kong S.A.R
jhsiao@hku.hk

Xiao Hu
University of Hong Kong Shenzhen
Institute of Research and Innovation
P. R. China
xiaoxhu@hku.hk

ABSTRACT

It is a common phenomenon that many students study with background music, but the influence of background music on learning is still an open question, with inconclusive findings in the literature. Inspired by the research gap, we conducted a controlled user experiment on reading with 100 students from a comprehensive university. The participants were tasked to read nine academic passages. In the meantime, those who were randomly allocated to the experiment group listened to their self-provided music in the background during the reading task, while those in the control group did not have background music during reading. During the experiment, participants' reading logs, self-reported meta-cognition and emotion status were recorded. This paper reports the results of comparing measures on reading performance, meta-cognition and emotion changes between the two groups. In addition, the relationships between participants' personal traits and their preferred background music type were investigated. Findings indicated that learning with background music of one's own choice could be beneficial for maintaining positive emotion, with no cost on reading performance. Through providing empirical evidence on the effect of background music on reading, this study contributes to furthering our understanding on human behaviors in multi-channel learning settings and rendering design implications for personalized music services for facilitating reading and self-learning.

CCS CONCEPTS

• **Applied computing ~ Arts and Humanities** ~ Sound and music computing • **Applied computing** ~ Education • **Human-centered computing ~** Human computer interaction (HCI)

KEYWORDS

Background music, User characteristics, Music characteristics, Academic reading, Learning performance, Emotion regulation

*Article Title Footnote needs to be captured as Title Note
†Author Footnote to be captured as Author Note

ACM/IEEE, June 19-23, 2020, Wuhan, Hubei, P.R. China

https://doi.org/10.1145/1234567890

ACM Reference format:

Ying Que, Yueyuan Zheng, Janet H. Hsiao and Xiao Hu. 2020. Exploring the Effect of Personalized Background Music on Reading Comprehension. In *JCDL' 20: Joint Conference on Digital Library, June 19-23, 2020. ACM, Wuhan, Hubei, P.R. China, 10 pages*. https://doi.org/10.1145/1234567890

1 INTRODUCTION

It is not difficult to observe that numerous students enjoy studying with background music in daily routine. As a medium for regulating emotion, music is widely used in daily contexts, for entertaining, relaxing or as background for the purpose of working and learning [1]. To meet the need, online music services including music digital libraries/repositories have rendered digital and/or streaming music to the general public for daily use [2]. As a recent study indicates, background music service has been offered in intelligent libraries, yet users' acceptance for such service varied across different genders, ages and preferred music types [3]. It points us in the direction of personalization for improving services related to background music in various scenarios including in the libraries.

On the other hand, research has been carried out to explore the effects of background music on self-learning and/or reading. Some existing literature explained that music could relieve anxiety that students often experience during learning, promote learning motivation and enhance learning performance [4]. However, other studies found that music could distract students' concentration and attention during learning which might result in decrease of learning efficiency [5]. Regarding these inconclusive findings, hypotheses have been raised by researchers to attribute the influence of background music on learning to multiple possible factors, including the diverse types of background music [25], complicatedness of learning tasks [26] and personal traits of listeners [27]. Therefore, how to provide students with suitable background music that matches personalized needs and facilitates learning is worthy of research. In addition, the inconsistent findings in the literature also call for more evidence on the effects of background music on learning. This study attempts to respond to this challenge through conducting a controlled user experiment with personalized music. Specifically, 100 participants were randomly divided into an experiment group who listened to their preferred music during reading and a control group who performed the same reading task without background music (see section 3.3). By comparing reading performance, meta-cognition and emotion changes between the two groups, this study aims to answer the following research questions:

RQ1. Is there any difference in reading performance between students who read with and without preferred background music?

RQ2. Is there any difference in emotion changes between students who read with and without preferred background music?

RQ3. Is there any difference in meta-cognition between students who read with and without preferred background music?

Besides, we further probed possible interactions among music characteristics, user characteristics, as well as students' learning performance and emotion changes during the experiment, answering the following two research questions:

RQ4. To what extent are students' preferred music types related to their personal traits?

RQ5. What kinds of music characteristics are related to reading performance or emotion changes?

Findings of this study will shed light on the influences of background music on students' learning, providing new empirical evidence that may help verify or refine theoretical foundations of background music and learning such as the arousal-mood hypothesis [9]. This study is also expected to contribute to the field of music information retrieval (MIR) and music digital libraries (MDL) by revealing the relationships between music characteristics, listeners' personal traits, emotion changes and learning performances. Based on findings of this study, future research can hopefully be inspired to investigate how to provide personalized background music in the right time for facilitating and optimizing reading and learning.

2 RELATED WORK

The impacts of music on learning performance and mood regulation have been studied in multiple disciplines. However, many findings remain controversial and inconclusive. Some studies found that background music played a role in promoting second language acquisition [6], altering or maintaining mood states [31] and improving students' motivation and engagement [7]. Besides, practical implications have also been proposed, for example, to blend background music and literary work in order to improve aesthetic enjoyment in digital libraries [2]. Nonetheless, other studies indicated that music had no or negative impact on learning, specifically through distracting students' concentration or attention during the learning process [8].

One explanation of the different effects of music on learning is the arousal-mood hypothesis [9] which states that music affects learning through regulating learners' emotion status particularly in arousal and mood dimensions. Arousal indicates the level of energy which ranges from very calm to very energetic, while mood indicates the level of happiness which ranges from very unpleasant to very pleasant. Mood is equal to valence introduced in Russel's two-dimensional model [10] which has been adopted extensively in emotion studies in the fields of psychology, education and MIR/MDL. The effects of music on valence and arousal are well recognized [11], and arousal and valence in turn can influence learning [12]. In general, positive and negative moods are regarded beneficial and harmful for learning [12][13], whereas it has been indicated that a certain level of arousal is helpful for learning performance but too much arousal could be detrimental instead [9][12].

As aforementioned research revealed, the arousal-mood hypothesis is mainly relevant to the influence of background music from learners' emotional benefits. Nonetheless, it is inadequate in explaining the negative influences of background music on learning which is also reported in the literature [8]. Regarding the cognitive function in the learning process, the irrelevant sound effect (ISE) stresses that listening to background music increases students' cognitive load and thus impairs learning [28][29]. More specifically, the ISE clarifies the reason that auditory reception is an inherent function of our brain, and listening to music during the learning process would inevitably expend the limited cognitive resources in the brain, thereby adding extra cognitive burdens to students [30].

Grounded on the above hypotheses, certain mediators related to the influence of background music on learning were consequently formulated, comprising task complexity [33], personal traits [34], and the information load characteristics in background music [32]. For instance, one experiment on reading comprehension [35] reported that students who were accustomed to listening to music while learning performed better on verbal learning tasks than participants who were not used to studying with background music. In addition, with respect to personality traits, the results from one mini-review [27] presented that introverted students were more impressionable to the adverse effects of music due to relatively higher cortical arousal than extroverted students.

Previous studies have been undertaken to determine what kind of music deemed as enhancing or distracting reading. The user experiment conducted in [5] compared three background music conditions (i.e. hip-hop music, light classical music, without music). The results showed that hip-hop had more distracting effects on cognitive performance than light classical music, yet light classical music was more distractive than the condition without music. Another study [9] played four types of background music with combined tempo (slow & fast) and intensity (low & high) when students performed reading tasks. Results indicated that, fast and loud music mostly disrupted students' reading comprehension. Similarly, a user experiment in [14] compared five audio conditions with balanced mode (minor, major) and tempo (slow, fast), as well as natural environmental sound. The results, however, showed no significant differences on reading performance across varied audio conditions. A user study was designed to exam the effects of lyrical and non-lyrical music on reading comprehension between two groups of students [38]. The results showed that although students in the condition of no lyrics achieved higher scores on reading comprehension than those in the lyric condition, the difference was not significant. Although music features (e.g., genre, tempo, mode, intensity, lyric) were probed and controlled in aforementioned experiments, the same background music was applied to all participants. It has been recognized that the influence of music may differ across people [21], and acceptance of listening to background music in digital libraries also varies across listeners' gender, age and music preference [3]. Thus, personalized background music is worthy of consideration in further research.

3 RESEARCH DESIN

3.1 Reading Task

In this study, participants were tasked to read nine English passages and answer two questions on the content of each passage. The passages were selected from academic reading samples from the Graduate Record Examinations (GRE) which are considered similar to academic content encountered in typical learning contexts at the university level. To quantitively gauge the difficulty level of the passages, the software, readable.io, was used to calculate Flesch-Kincaid Grade Level [15] for each passage. This grade is a widely used text readability rating score that measures the difficulty level of reading a piece of text against the Grade Level of the United States education system. Based on the grade, the nine passages were evenly divided into three sub-categories with different difficulty levels: easy, medium, and hard. The mean grade level and the standard deviation (in parenthesis) of passages in each category was: 13.2 (0.35) for easy, 17.4 (0.46) for medium, and 21.0 (0.47) for hard.

To minimize possible influence of passage content on users' emotion, the passages were chosen to be emotionally neutral and with low arousal. They were selected from a variety of topics including archaeology, astronomy, biology, science, literature, history and sociology. All of the passages expressed complete ideas or meanings and contained similar number of words, with an average word count of 218 (standard deviation 9.84). To assess participants' reading comprehension performance, we designed two multiple-choice questions (MCQs) for each passage according to the method introduced in [16] where one text/fact-based question and one inference-based question were designed for each passage.

3.2 Background Music

Before participant came to the experiment, they were requested to provide a music piece or playlist that they often listened to during self-learning (if they tended to listen to background music while studying) or a music piece or playlist they would like to listen to during passage reading (if they did not tend to listen to background music while studying). Before each participant came to the experiment, the music provided by the participant was processed and standardized. Specifically, the volume of music was normalized to range between 65 dB(A) and 75 dB(A) when played in the headphone worn by the participants, which was an acceptable volume for learning as suggested in the literature [17]. Besides, necessary audio effect pre-processing was performed to ensure comfortable listening experience such as fading in and fading out in transition to the next songs and shortening segments of prolonged silence.

3.3 Procedure

Before joining this experiment, a pre-experiment questionnaire was filled by participants collecting their demographic information, self-assessed personal traits, music training, attitude, preference and habits on studying with background music. Participants were randomly divided into groups with or without background music when they read the English passages. When a participant came to the experiment, he/she was requested to take an online English test LexTale [18] to measure general English proficiency. Before the reading task started, the participant went through a practice block under the guidance of the experiment facilitator, which helped them become familiar with this experiment.

The formal experiment consisted of three blocks each of which was assigned with three English passages with one in each difficulty level. The orders of the blocks were randomized across participants for counter-balance. At the beginning of each block, participants were requested to answer two questions about their current emotion states, one on valence and the other on arousal, which were the two dimensions in the Affect Grid [10]. After that, all participants were prompted to listen to the self-provided music for one minute, without any reading task. After the one-minute music, the participants answered the valence and arousal questions again. This was to measure the effect of music on participants' emotional changes at rest (without reading).

During reading, the music continued for the experiment group, while for control group, the background was in silence. Upon finishing reading each passage, the audio (for the experiment group) paused. Three meta-cognition questions on engagement, difficulty, and understanding [16] (see more in Section 3.4) were presented to both groups. After that, two MCQs on the content of the passage were presented and solicited participants' responses, which was for measuring the accuracy of reading comprehension. The participants could take as much time as needed read and to answer questions. This is to simulate a typical learning context rather than that of examinations. Upon completing the MCQs, the next passage would start, and (for the experimental group) the background music resumed. The order of the three passages in one block was randomized across participants for counter-balance.

After reading three passages, a block was about to finish. Participants were asked to answer valence and arousal questions again to measure their emotional states after reading. Before starting the next block, participants had a break of 2 minutes, to help alleviate possible fatigue in the experiment.

After completing all three blocks of reading tasks, an exit interview was conducted. This was to collect participants' thoughts on the experiment, e.g., their attitudes towards study with background music, thoughts and feelings on the experiment procedure, the difficulty of the read passages and answered MCQs. At the end of experiment, each participant was paid 150 Hong Kong dollars as a nominal remuneration.

3.4 Measures

The following learner-centered measures were used to answer research questions proposed in this study.

Reading performance was measured by three metrics.

(a) Reading Accuracy: the percentage of correctly responded questions among all questions. As a metric of finer granularity, accuracies on text-based and inference-based questions were also calculated respectively.

(b) Passage Reading Time: the time difference between the time point when a passage was shown to a participant and the time point when the participant pressed the "continue" button which indicates completion of reading.

(c) Question Answering Time: the time difference between the time point when a question was shown to a participant and the time point when the participant answered the question by hitting the designated key on the computer keyboard.

These three reading performance metrics were in interval scales measured for each passage and then aggregated across all passages.

Emotion during listening to music, before and after reading was measured by the valence and arousal questions raised in the experiment procedure, corresponding to the following four metrics.

(a) Music Valence Change & (b) Music Arousal Change: the comparison of self-reported valence & arousal levels before and after listening to the one-minute music without reading.

(c) Reading Valence Change & (d) Reading Arousal Change: the comparison of self-reported valence & arousal levels before and after reading the passages.

In this study, we followed the method in [8] to operationalize the emotional dimensions into two questions with 9-point Likert scales. For the valence question, 1 indicates "very unpleasant" and 9 indicates "very pleasant". For the arousal question, 1 indicates "very calm" and 9 indicate "very energetic". To measure emotion changes, we calculated the aforementioned four metrics in each block, and then averaged each metric across the three blocks.

Meta-cognition was measured based on a self-reported meta-cognitive scale provided in [16]. In this study, the measures were used to examine whether music can reduce students' perceived difficulty or promote engagement and understanding when reading passages. Specifically, the following three self-perceived metrics were measured.

(a) Engagement: how engaged a participant was during reading.

(b) Difficulty: how difficult a participant thought the passage was.

(c) Understanding: how well a participant understood the passage.

Following [8], each of questions was operationalized as a 5-point Likert scale question, with 1 standing for the lowest rating (i.e., not engaged at all, very easy, not understand at all) and 5 standing for the highest rating (i.e., very engaged, very difficult, understand very well). For comparison, we aggregated meta-cognition ratings across passages and participants in data analysis.

3.5 Music Characteristics

Music characteristics were measured by the following three dimensions with eight metrics. It is noteworthy that the background music of each participant was a playlist with multiple music pieces, as the reading time in each block was longer than most music pieces.

(a) Acoustic Features

Rhythm includes 2 metrics, i.e. tempo and rhythm strength.

- Tempo is represented by number beats per minute.
- Rhythm strength refers to the strength of rhythm counted from the mean of onset strength.

Loudness includes 1 metric, i.e. RMSE.

- RMSE compute the root-mean-square (RMS) energy value for each frame from spectrograms [39].

Timber includes 3 metrics, i.e. cent, flatness, roll-off.

- Cent: computes the spectral centroid. The centroid (mean) is extracted at each normalized frame of a magnitude spectrogram which is treated as a distribution over frequency bins [39].
- Flatness: computes the spectral flatness. Spectral flatness is a measure to quantify a sound how much it is similar to a noise or a tone. A high flatness value means the spectrum is close to white noise [39].
- Roll-off: computes the roll-off frequency. The spectral roll-off is defined as the frequency which a percentage of the magnitude distribution is concentrated and the percentage is generally set as 85% [40].

(b) Lyric: represents whether and to what extent lyrics exist in the background music playlist for each participant. This metric includes 3 ranked values: 1 stands for no lyrics in the music playlist, 2 for several lyric pieces included in the playlist, and 3 for all music pieces having lyrics.

(c) Genre: represents the dominant genre (i.e. the genre appearing most often in a participant's playlist). This study followed the genre classification taxonomy in [36] which contains eight general music genres: Rebellious (i.e. heavy metal, punk); Classical (i.e. piano/organ); Rhythmic & Intense (i.e. hip-hop, pop); Easy Listening (country, folk); Electronic; Contemporary Christian; Jazz & Blues; Traditional Christian. It is notable that the music used in this experiment did not contain music in the genres of Contemporary Christian and Traditional Christian. Besides, this study added one category named *Mix* to represent the cases where multiple genres appeared in the playlist with roughly equal amount. The genre metric finally consists of 7 values corresponding to the 7 genre categories.

It is noteworthy that the acoustic features (i.e., rhythm, loudness, timber) were extracted employing a well-adopted audio processing library, Librosa [22]. The metrics were all in interval scales. In contrast, the other two music characteristics (i.e. genre and lyric) were determined by manual coding with a consensus protocol. That is, two researchers coded the measures independently, followed by discussion between them. Disagreed items were then judged by the third researcher and coded by majority votes.

3.6 User Characteristics

User characteristics were measured by the following five metrics. All of them were collected in the pre-questionnaire, including demographic information, music training, background music listening habits and personality assessment.

(a) Gender: it is on a binary scale (Male=1, Female=2)

(b) Age: participants recruited in this study were from 18-35 years old. The original age data were continuous. This study classified the data equally into three ranked categories (18-23=1; 24-29=2; 30-35=3).

(c) Music Training: it assesses whether the participants have passed any graded/qualification exam in music. This metric is on a binary scale (No=0, Yes=1)

(d) Background Music Listening Habits: we followed the method introduced in the literature [14] to measure the frequency of listening to background music during self-learning/reading. This metric is a 5-point ordinal variable (Never=1, Seldom=2, Sometimes=3, Often=4, Always=5).

(e) Personality Traits: we used Ten Item Personality Inventory (TIPI) scale to collect participants' self-assessed personality traits [37]. The TIPI contains 10 questions in 7-point Likert scales: extraverted & enthusiastic; critical & quarrelsome; dependable & self-disciplined; anxious & easily upset; open to new experiences & complex; reserved & quiet, sympathetic & warm; disorganized & careless; calm & emotionally stable, conventional & uncreative. For each question, 1 indicates "disagree strongly" and 7 indicate "agree strongly".

4 RESULTS

4.1 Participants

100 participants (49 males, 51 females) in a comprehensive university were recruited in this experiment. All of them were non-native English speakers. All reported no visual, hearing or learning impairment. Participants were divided into experiment group with music (50) and control group without music during reading (50). Mean age was 25.0 (SD = 4.24) for the experiment group and 22.5 (SD = 3.88) for the control group. 23 participants in the experiment group and 19 participants in the control group had passed graded/qualification exam in music. In terms of habits towards listening to music during reading/self-learning, experiment group answered "like" (19 participants), "neutral" (17) and "dislike" (14), while control group answered "like" (17), "neutral" (24) and "dislike" (9). Results using Independent sample T test indicates that the two groups were comparable in terms of English abilities as measured by the LexTale scores (p=.741). The mean and standard deviation (in parenthesis) of the LexTale scores of experiment and control groups were .707 (.131) and .699 (.119).

4.2 Effects of Music on Reading Performance

To answer RQ1, independent sample T test was used to compare the three metrics of reading performance between the two groups. Metrics include Overall Reading Accuracy, Passage Reading Time and Question Answering Time (see section 3.4). The metrics were averaged across all passages for each participant. When analyzing Accuracy and Question Answering Time, the difference between text-based and inference questions were taken into account. To control Type I error in multiple comparisons, Benjamini-Hochberg procedure was applied [23]. Results of the t-tests, means and standard deviation (in parenthesis) of the metrics are shown in Table 1.

Table 1: Results of T-test on Reading Performance

Metrics	Exp Group	Ctrl Group	p
Overall Acc	.474(.150)	.431(.151)	.150
Text Acc	.489(.159)	.407(.162)	.013*
Infer Acc	.460 (.198)	.454 (.203)	.885
PasgTime	126.2 (51.6)	106.9 (47.3)	.054
Overall QSTime	30.7 (12.6)	28.6 (10.8)	.366

N=100 (Exp=50; Ctrl=50). $^{*}p < 0.05$

As we can see in Table 1, there was no significant difference between the two groups on reading accuracy (p =.150), indicating that the existence of background music did not cause detrimental effects on the overall reading accuracy. Regarding the fine-grained Reading Accuracy, a significant difference was found on the accuracy of answers to text-based questions (p =.013), suggesting that the experiment group (Mean = 48.9%) significantly outperformed the control group (Mean = 40.7%) in reading accuracy measured by text-based questions. For inference questions, there is no significant difference between groups (p= .885). Considering the differences of the two types of questions, text-based questions may require better memorization and storing of information presented in the passage. Therefore, the results seem to indicate that participants with their preferred background music might have memorized information better than those without preferred background music during reading.

For the time spent on reading the passages, the difference between the two groups was just shy of being significant (p = .054). The averaged reading time spent by participants in the experiment group (Mean = 126.2 seconds) was approximately one fifth (18.1%) longer than that of the control group (Mean = 106.9 seconds). Relating this result and the result that the two groups had similar reading accuracy, it could be possible that background music might have added cognitive load of the participants and thus they needed more reading time to compensate for it [24]. It could also be that participants with background music were more comfortable to spend more time on reading the passages, whichmay be related to a known function of music in influencing people's temporal perception [19].

However, in terms of Question Answering Time, there was no significant difference between the two groups (p = .366). This held true for the comparison of QSTime across both text-based and inference questions (p = .215 and =.613 respectively). It is noteworthy that, in order to standardize the condition of question answering, there was no music played in background for both groups when participants answered MCQs after reading the passages. This may be the reason why there was no difference in question answering time.

4.3 Effects of Music on Emotion Change

To answer RQ2, independent sample T tests were used to compare the valence and arousal levels between groups: 1) before and after listening to the one-minute music at rest (Music Valence/Arousal Change; 2) before and after reading the passages (Reading Valence/Arousal Change (see Section 3.4). It is noteworthy that the emotion changes for the first block (when neither group had read any passage) did not present any significant differences in the change of valence (p =.146) or the change of arousal (p =.179), suggesting that the background music from ones' own choices had similar impacts on the mood regulations when initially between the two groups. After that, we averaged the differences of the emotion changes across the three blocks. Type I error was controlled through Benjamini-Hochberg procedure [23]. Results of the t-tests, means and standard deviation of the metrics are shown in Table 2.

Table 2: Results of T-test on Emotion Changes

Metrics	Exp Group	Ctrl Group	p
Music_Valence_Change	.71 (.99)	1.19 (1.13)	.024*
Music_Arousal_Change	.27 (.93)	.64 (1.58)	.150
Reading_Valence_Change	-1.16 (1.26)	-1.68 (1.25)	.041*
Reading_Arousal_Change	-.40 (1.14)	-1.06 (1.83)	.033*

N=100 (Exp=50; Ctrl=25). * $p < 0.05$

It can be seen in Table 2, there was a significant difference (p = .024*) between the groups in valence change before and after listening to the 1-min music without reading. For both groups, the valence changes were positive meaning that participants felt more pleasant after listening to music. Besides, the change of valence in control group (Mean=1.19) was to a larger extent than the experiment group (Mean=.71). It can be explained that the control group were not listening to music while reading, but the group were listening to the 1-min music at the beginning of each block, and thus the music could cause more obvious enhancement to the mood in the control group than that in the experiment group. However, the two groups did not present significant difference (p =.150) in changes in the arousal dimension, which indicates the energy levels between the two groups increased for similar levels (Mean = 0.27 and 0.64 respectively) during music listening (before reading).

This paper also analyzed emotion changes before and after reading. As can be seen in Table 2, there was a significant difference (p =.041*) between the two groups in valence change before and after reading. For both groups, the means of valence changes were negative (Mean = -1.16 and -1.68 respectively), meaning participants felt less pleasant after reading, which could be attributed to the fatigue effect since the reading materials in this study, which were selected from GRE tests, were quite challenging for most second-language learners. However, the valence change was to a smaller extent in the experiment group than that of the control group. It can thus be said that the existence of background music played a role in alleviating negative emotions in the reading process. In addition, there was a significant difference (p = .033*) between groups in the change of arousal. In fact, the arousal levels of both groups reduced (Mean = -.40 and -1.06 respectively). This result indicates that participants in general calmed down after reading, which is not surprising given that reading is a sedentary activity. It is still noteworthy that the decrease of arousal in the experiment group was less than half of that in the control group, suggesting that background music might have helped maintain arousal level of the participants [11].

Combining this finding and those from Section 4.2, we can see that, although participants in the experiment group seemed to have paid more efforts in reading (i.e., longer reading time) than those in the control group, they had more benefit on the emotion aspect. It is thus worthy of further investigation in future research whether the more reading time spent by participants in the experiment group was all about efforts or was there a hedonic factor associated with music as suggested in previous studies [20].

4.4 Effects of Music on Meta-Cognition

To answer RQ3, independent sample T tests were used to compare the meta-cognitive measures (see Section 3.4) between the two groups. The metrics were averaged across all reading passages. Results of the t-tests, means and SD (in parenthesis) are shown in Table 3. No significant difference was found in meta-cognitive metrics between the two groups, indicating that music had little influence on self-perceived difficulty of the passage, engagement in reading, and understanding level of the content.

Table 3: Results of T-test on Meta-cognition

Metrics	Exp Group	Ctrl Group	p
Difficulty	2.90 (.57)	2.85 (.68)	.712
Engagement	3.26 (.78)	3.12 (.81)	.372
Understanding	3.20 (.54)	3.28 (.64)	.503

N=100 (Exp=50; Ctrl=50). * $p < 0.05$

To further investigate RQ4, we ran linear regression with stepwise forward selection to predict reading accuracy from the meta-cognitive metrics in each group. The reading accuracy were measured by all questions, and by fine-grained measures including inference-based and text-based questions. Significant variables in the prediction models are reported in Table 4.

Table 4: Regression on Predicting Reading Accuracy

Group	Accuracy	Predictor	β	t	p
Control	Overall	understand	.498	3.982	.000*
	Inference		.442	3.415	.001*
	Text		.376	2.813	.007*
Experiment	Overall	understand	.351	2.595	.013*
	Inference		.393	2.957	.005*

N=100 (Exp=50; Ctrl=50). * $p < 0.05$; R^2= .248, .195, .142 for the three predictive models in the control group; R^2= .123, .154 for the two predictive models in the experiment group.

As we can see in Table 4, for both groups, the metrics regarding participants' perceived difficulty and engagement levels during reading were excluded in the regression model, meaning these metrics did not have predictive power in estimating the reading accuracy. However, in terms of self-perceived understanding level of the reading content, it was found statistically significant in predicting the overall accuracy and the accuracy measured by inference questions. Besides, relationship between reading accuracy and understanding level in each group was moderate positive (coefficient β > 0.3). It is also noteworthy that the predictive effects differ between control group (R^2=.248) and experiment group (R^2=.123), suggesting that approximately 24.8% of the observed variance in reading accuracy in control group could be explained by participants' self-reported understanding level, while that in experiment group was much lower (12.3%). In addition, with regard to text-based questions, it is interesting to see that in the control group without background music during reading the self-reported understanding level was still significant (p =.007*) in predicting text-based reading accuracy. In contrast, this was not the case for the experiment group (reading with background music). It could be inferred that reading under background music could have possibly interfered with participants' perceived understanding level, and thus when the participants in the experiment group answered text-based questions that required to memorize information on the read passages, the self-report understanding level seemed less matched to the actual reading accuracy. It is interesting to find that although the participants in the experiment group seems not quite confident on the understanding level to the read content, they got higher score on the accuracy from the text-based questions (see the result in section 4.2).

4.5 The Diversity of Background Music

To assess the diversity of the background music brought by the participants, we first visualized the six acoustic metrics, i.e. tempo, rhythm strength, RMSE, cent, flatness, roll-off in boxplot, after being normalized into the range from 0 to 1 (Figure 1). As we can see, the metrics present different distribution. Tempo (one of the rhythm features) clusters in the middle values (between 0.4 and 0.6), while flatness (one of the timber feature) clusters near lower values between 0 and 0.3. Despite aggregating on central regions, the metrics also dispersed on the rest areas of the boxplot.

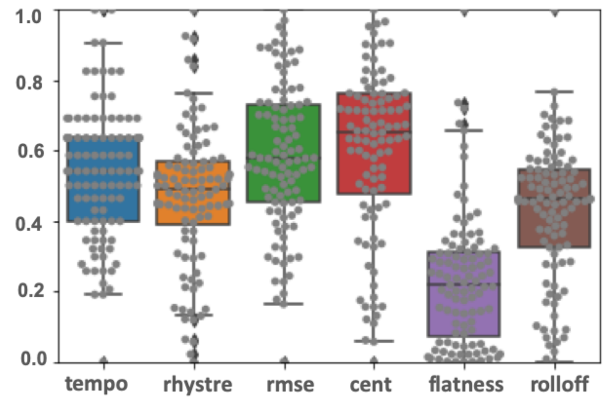

Figure 1. Distribution of Acoustic Features

For the lyric feature. playlists from nearly two thirds participants (N = 63) were with lyrics, indicting a tendency of listening to lyrical background music during self-learning/reading. Music brought by one third participants (N=31) were without lyrics. Besides, there were a small number of participants (N=6) whose playlist consisted of both lyrical and instrumental music, showing that having lyric or not was not a criterion for them in selecting background music for learning and/or reading.

Figure 2 presents the distribution of music genre where the most prevailing was Rhythmic & Intense (56%), followed by Classical (21%) and Easy Listening (11%). In contrast, the rest genres (i.e. Mix, Rebellious, Electronic, Jazz & Blues) only accounted for a small percentage each (less than 5%). The distributions of music features made it clear that preferences for background music for reading differed across participants'.

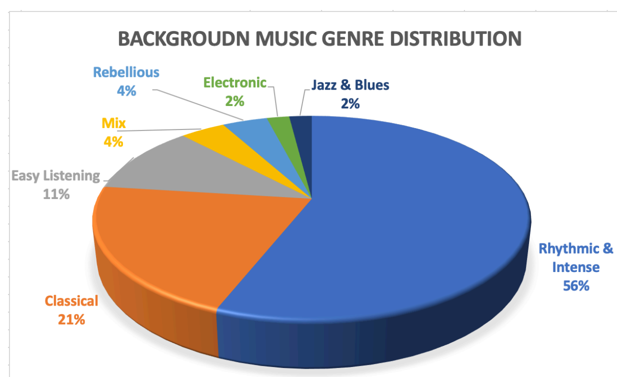

Figure 2. Distribution of Genre

4.6 Relationship between Background Music Characteristics and User Characteristics

This section aims to answer RQ4. Under RQ4, three hypotheses were formulated in consideration of three characteristics (i.e. acoustic features, lyric feature, genre) in the background music:

H1: acoustic features are related to user characteristics.

H2: lyric feature is related to user characteristics.

H3: genre is related to user characteristics.

Given that acoustic features were in interval scales, to test H1, we calculated point-biserial correlation coefficients for user characteristics metrics on binary scales: Gender (Male=1, Female=2), and Music Training (N=0, Y=1). We also ran Spearman's correlation on three ordinal variables: Age, Frequency of studying with background music, Personality Traits (see Section 3.6). Results with significant correlations after Benjamini-Hochberg correction are shown in Table 5.

Table 5: Correlation between acoustic and user characteristics

Character	Music Feature	Coefficient	*p*
Music Train	rhy_str	-.229	.022*
Music Train	flatness	-.248	.013*
Age	rmse	-.228	.022*
Careless	cent	.256	.010*
Careless	roll-off	.261	.009*

N=100. * $p < 0.05$, ** $p < 0.01$

As shown in Table 5, music training (a user characteristic) has significant and weak negative correlation with rhythm strength (a music feature related to rhythm) and spectral flatness (a music feature related to timber, indicating how much the sound is tone-like or noise-like). This result indicates that musically trained persons might prefer studying under background music with soft light type and less noise-like. Besides, there is also a negative weak correlation between age and RMSE (one loudness feature of music). In other words, the younger the listener is, the more energetic background music he/she is inclined to like to listen to during study. In addition, the personality trait *careless & disorganized* had a positive relationship with two music features related to timber, i.e. roll-off and centroid. The roll-off is normally applied to estimate the quantity of high frequency exists in the music, and the spectral centroid is the means of the spectral distribution in each frame [40]. Relating to that human hearing system is sensitive to moderately high frequency band (i.e., 2 to 4 kHz8) [41], the result might suggest that a careless person is possibly inclined to listen to music centring on relatively higher frequency band which is generally more distracting. Nevertheless, other two user characteristics (i.e., gender, frequency of studying with background music) had no significant correlation with acoustic features. Thus, our hypothesis H1 is partially supported.

Given that the lyric feature was in ordinal scale, to test H2, we ran a one-way ANOVA to measure the correlation for categorical variable: Gender and Music Training. However, the result presented no significant correlation among these metrics (at $p <$ 0.05 level). Following that, we further ran ordinal least squares (i.e. a type of linear least squares method in a linear regression for examining the relationship between two or more interval/ratio variables) in predicting the lyric feature from the 12 ordinal variables: Age, Frequency of studying with background music, and 10 variables of Personality Traits. Significant variables in the prediction models are reported in Table 6.

Table 6: Regression on Predicting Lyric Feature

Metric	Predictor	*β*	t	*p*
Lyric	Careless	.282	2.934	.004*
	Reserved	-.210	-2.190	.031*

N=100. * $p < 0.05$; R^2= .113 for the predictive model.

Two personality traits were found statistically significant in predicting the lyric feature: careless and reserved. The correlation between lyric feature and careless personality was weak positive ($\beta < 0.3$), whereas the correlation between lyric feature and reserved personality presents weak negative relationship. Besides, the predictive power s (R^2=.113) indicates that approximately 11.3% of the variance in the lyric feature could be explained by self-assessed personality traits. Other user characteristics had no significant correlation with lyric feature. Thus, our hypothesis H2 is partially supported.

Given that the feature of genre was in categorical scale, to test H3, we ran Pearson's Chi-Squared tests to measure the association between genre and other two categorical variables: Gender and Music Training. However, there was no significant association between these metrics (Gender: p = .840; Music Training: p = .599). Further, we ran multinomial logistic regression to predict the genre (categorical metric) from 12 ordinal variables: Age, Frequency of studying with background music, and 10 variables of Personality Traits. However, the logistic model did not fit significantly better than an empty model (p=1.000) indicating that the 12 variables of user characteristics could not predict the genre of students' preferred background music. Thus, H3 is not supported.

4.7 The Role of Music Characteristics

This section aims to answer RQ5. Under RQ5, three hypotheses were formulated in view of the characteristics (i.e. acoustic features, lyric feature, genre feature) in background music. It is noteworthy that the following analysis focused on the data from the experiment group (N=50), because RQ5 is about the effects of background music during reading while participants in the control group were reading in silent condition.

H4: the acoustic features are related to reading performance (i.e. reading accuracy, reading time) or emotion changes (i.e. changes of valence and arousal during reading).

H5: the lyric feature is related to reading performance or emotion changes.

H6: the genre feature is related to reading performance or emotion changes.

Given that acoustic features were in interval scales, to test H4, we used a series of linear regression to predict reading performance and emotion changes from acoustic features. Significant variables in the prediction models are reported in Table 7.

Table 7: Regression on Predicting Emotion Changes

Metric	Predictor	*β*	t	*p*
Reading_Valence _Change	RMSE	-.350	-2.589	.013*

N=50. * $p < 0.05$; R^2= .123 for the predictive model.

As we can see in Table 7, RMSE (a music feature related to loudness) was found statistically significant in predicting the change of valence during reading . The significant correlation between the two variables is moderately negative ($\beta < -0.3$), suggesting that loud background music was possible to impair students’ pleasantness level when they were involved in academic reading tasks [9]. Just as a participant mentioned in the post-experiment interview: *“Although when I was reading especially those content I was not interested, music can make me less depressed [. However], exciting music will bother me and made me less focused.”* (Participant #120). In addition, regarding the predictive effect (R^2), approximately 12.3% of the variance in valence change during reading could be explained by the audio feature RMSE. The results show that metrics of reading performance have no significant correlation with acoustic features. Thus, our hypothesis H4 was partially supported.

Given that lyric feature was in ordinal scale and genre feature was in categorical scale, to test H5 & H6, one-way ANOVA was applied to measure the correlation between reading performance and emotion changes. Boferroni test was used for Post Hoc Multiple Comparisons. Table 8 shows the results with significant effects.

Table 8: Effects on Emotion Changes during Reading

Music Ft	Emotion	F value	Post Hoc Tests
Lyric	Reading_Valence_Change	3.895*	without lyric (-0.69) vs. with lyric (-1.53)

* $p < 0.05$; values in parentheses are mean changes in that condition; Music Ft=Music Feature

From Table 8, we can see that background music with different lyric conditions had different effects on emotion change during reading ($p = .023^*$). Nonetheless, as the Post Hoc tests indicate, the difference of emotion changes during reading between the lyric condition and no lyric condition was just shy of being significant (p=.079). Specifically, the music piece without lyrics decreased valence level for 0.69 scale on average while the piece with lyrics decreased valence level for 1.53 scale on average. In consideration of the decreased scale, it seems that, compared with music without lyrics, listening to background music with lyrics had more negative influences in participants’ pleasantness level during reading. In the exit interview, some participants stated that the lyric could cause mood fluctuations, *“In the experiment, the lyric was in my mother tongue (Chinese), but what I read were in English. It was indeed disturbing, and my mood got irritable.”* (Participant #06). However, there were also different opinions, *“lyrics have little effects on my attention. Even if the songs are from my mother tongue, I cannot understand more than 50% of the lyrics while listening.”* (Participant #19). Thus, the insignificant correlation might have resulted from differences across participants, reinforcing the importance of personalization in selecting background music. However, the music pieces with and without lyrics did not show significant difference in reading performance, i.e. reading accuracy, reading time at $p < 0.05$ level. This aligned with the results in research [38]. In addition, different genre types did not show statistical difference in reading performance or emotion changes either. Thus, H5 was partially supported, whereas H6 was not supported.

5 CONCLUSIONS, IMPLICATIONS AND FUTURE WORK

This paper reports an empirical study probing the effect of background music on reading with a control user experiment. In the study, we compared reading performance, emotion changes and meta-cognitions between groups of participants with or without music during reading. One distinctive design in this research is that the background music was provided by the participants as preferred by them in the scenario of studying with background music. The relationship between user characteristics and music characteristics, as well as the effects of music features on reading performance and emotion changes were also analyzed. Results demonstrated interesting findings. First, studying with preferred music in the background did not impair performance in academic reading. On the other hand, background music could benefit learners in their emotion, through maintaining their valence and arousal levels during reading. As for the meta-cognition, self-perceived understanding level was found to be a significant predictor for the overall reading accuracy measured by all questions and the accuracy measured by inference-based questions in each group. However, regarding the reading accuracy measured by the text-based questions, the understanding level in control group was still a significant predictor for reading accuracy, whereas in the experiment group the understanding level was not statistically significant for prediction.

In terms of the relationship between music characteristics and user characteristics, several significant coefficients were indicated between acoustic features (related to rhythm, loudness and timber) and user characteristics (including age, music training and personality). With regard to the role of music played in the learning process, it was found that music with higher loudness and stronger intensity could be detrimental to the pleasantness level of students’ emotion. Besides, whether the music contains lyrics was found to have a significant correlation with emotion changes in the valence dimension, whereas the effect seemed to vary from person to person. By using participants preferred, personalized music, this study revealed findings that are different from existing studies using uniformed music. This calls for more research towards personalizing background music for facilitating learning. Future work is needed to investigate the effects of background on other learning tasks involving different cognitive abilities such as mathematics learning or writing. It is hoped that this study could help inspire more research in the MIR/MDL field to explore effective methods in retrieving or recommending suitable music for specific users in different contexts that involve learning.

ACKNOWLEDGMENTS

This study is supported by National Natural Science Foundation of China (No. 61703357) and the Research Grants Council of the Hong Kong S. A. R., China (No. HKU 176070).

REFERENCES

[1] J. H. Lee, H. Cho, and Y. S. Kim, “Users' music information needs and behaviors: Design implications for music information retrieval systems,” *Journal of the association for information science and technology,* vol. 67, no. 6, pp. 1301-1330, 2016.

[2] L. Zhuang, Z. Ye, J. Wu, F. Zhou, and J. Shao, "Towards a new reading experience via semantic fusion of text and music." pp. 149-152. *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*. ACM, 2011.

[3] Y. Liu, “Investigating users' willingness of acceptance for background music service in intelligent library,” *Library Hi Tech*, 2019.

[4] I. Eady, and J. D. Wilson, “The influence of music on core learning,” *Education,* vol. 125, no. 2, pp. 243-249, 2004.

[5] P. T.-M. Chou, “Attention Drainage Effect: How Background Music Effects Concentration in Taiwanese College Students,” *Journal of the Scholarship of Teaching and Learning,* vol. 10, no. 1, pp. 36-46, 2010.

[6] H. J. Kang, and V. J. Williamson, “Background music can aid second language learning,” *Psychology of Music,* vol. 42, no. 5, pp. 728-747, 2014.

[7] K. N. White, “The Effects of Background Music in the Classroom on the Productivity, Motivation, and Behavior of Fourth Grade Students,” *Online Submission*, 2007.

[8] P. J. Flowers, and A. A. M. O'Neill, “Self-reported distractions of middle school students in listening to music and prose,” *Journal of Research in Music Education,* vol. 53, no. 4, pp. 308-321, 2005.

[9] W. F. Thompson, E. G. Schellenberg, and A. K. Letnic, “Fast and loud background music disrupts reading comprehension,” Psychology of Music, vol. 40, no. 6, pp. 700-708, 2012.

[10] J. A. Russell, A. Weiss, and G. A. Mendelsohn, “Affect grid: a single-item scale of pleasure and arousal,” Journal of personality and social psychology, vol. 57, no. 3, pp. 493, 1989.

[11] T. Eerola, and J. K. Vuoskoski, “A review of music and emotion studies: Approaches, emotion models, and stimuli,” Music Perception: An Interdisciplinary Journal, vol. 30, no. 3, pp. 307-340, 2013.

[12] E. G. Schellenberg, “Cognitive performance after listening to music: A review of the Mozart effect,” Music, health, and wellbeing, pp. 324-338, 2012.

[13] J. W. You, and M. Kang, “The role of academic emotions in the relationship between perceived academic control and self-regulated learning in online learning,” Computers & Education, vol. 77, pp. 125-133, 2014.

[14] X. Hu, F. Li, and R. Kong, "Can Background Music Facilitate Learning?: Preliminary Results on Reading Comprehension." pp. 101-105.

[15] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom, “Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel,” 1975.

[16] A. C. Strain, R. Azevedo, and S. K. D'Mello, “Using a false biofeedback methodology to explore relationships between learners' affect, metacognition, and performance,” Contemporary Educational Psychology, vol. 38, no. 1, pp. 22-39, 2013.

[17] N. Perham, and H. Currie, “Does listening to preferred music improve reading comprehension performance?,” 2015.

[18] K. Lemhöfer, and M. Broersma, “Introducing LexTALE: A quick and valid lexical test for advanced learners of English,” Behavior research methods, vol. 44, no. 2, pp. 325-343, 2012.

[19] J. J. Kellaris, and R. J. Kent, “The influence of music on consumers' temporal perceptions: does time fly when you're having fun?,” Journal of consumer psychology, vol. 1, no. 4, pp. 365-376, 1992.

[20] X. Hu, and N. Kando, “Task complexity and difficulty in music information retrieval,” Journal of the Association for Information Science and Technology, vol. 68, no. 7, pp. 1711-1723, 2017.

[21] D. Boer, R. Fischer, H. G. Tekman, A. Abubakar, J. Njenga, and M. Zenger, “Young people's topography of musical functions: Personal, social and cultural experiences with music across genders and six societies,” *International Journal of Psychology,* vol. 47, no. 5, pp. 355-369, 2012.

[22] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python."

[23] D. Thissen, L. Steinberg, and D. Kuang. 2002. Quick and Easy Implementation of the Benjamini-Hochberg Procedure for Controlling the False Positive Rate in Multiple Comparisons. Journal of Educational and Behavioral Statistics, 27, 1 (2002), 77-83.

[24] F. Cauchard, J. E. Cane, and U. W. Weger, “Influence of background speech and music in interrupted reading: An eye‐tracking study,” *Applied Cognitive Psychology,* vol. 26, no. 3, pp. 381-390, 2012.

[25] J. Kantner. 2009. Studying with music: Is the irrelevant speech effect relevant. Applied Memory. Nova Science Publishers, NY, US. 19-40

[26] J. Kämpfe, P. Sedlmeier, and F. Renkewitz. 2010. The impact of background music on adult listeners: A meta-analysis. Psychology of Music, 39, 4 (2010), 424-448.

[27] M. B. Küssner. 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.

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

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

[30] J. A. M. Lehmann, and T. Seufert. 2017. The Influence of Background Music on Learning in the Light of Different Theoretical Perspectives and the Role of Working Memory Capacity, 8, 1902

[31] W. L. Magee, and J. W. Davidson, “The effect of music therapy on mood states in neurological patients: a pilot study,” Journal of music therapy, vol. 39, no. 1, pp. 20-29, 2002.

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

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

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

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

[36] D. George, K. Stickle, F. Rachid, and A. Wopnford, “The association between types of music enjoyed and cognitive, behavioral, and personality factors of those who listen,” Psychomusicology: A Journal of Research in Music Cognition, vol. 19, no. 2, pp. 32, 2007.

[37] E. Romero, P. Villar, J. A. Gómez-Fraguela, and L. Lopez-Romero, “Measuring personality traits with ultra-short scales: A study of the Ten Item Personality Inventory (TIPI) in a Spanish sample,” *Personality and Individual Differences,* vol. 53, no. 3, pp. 289-293, 2012.

[38] Z. Liapis, Z. Giddens, and M. Uhlenbrock, “Effects of lyrical music on reading comprehension,” *Retrieved August,* vol. 20, pp. 2010, 2008.

[39] B. McFee, M. McVicar, S. Balke, C. Thom‘e, C. Raffel, D. Lee, O. Nieto, E. Battenberg, D. Ellis, and R. Yamamoto, “librosa/librosa: 0.6. 3,” Zenodo, 2018.

[40] G. Tzanetakis, and P. Cook, “Musical genre classification of audio signals,” IEEE Transactions on speech and audio processing, vol. 10, no. 5, pp. 293-302, 2002.

[41] M. Müller, Fundamentals of music processing: Audio, analysis, algorithms, applications: Springer, 2015.