2015

The Complicated Relationship Between Music and Foreign Language Learning: Nuanced Conditions Required for Cognitive Benefits Due to Music

Talia Greenberg
Oberlin College

SCIENTIA VERITAS

Oberlin College
Department of Psychology
Honors Thesis

May 2015

Honors Candidate: *Talia Greenberg*

Thesis Title: *The Complicated Relationship Between Music and Foreign Language Learning: Nuanced Conditions Required for Cognitive Benefits Due to Music*

Honors Examination Committee

Signature

Signature

Signature

The Complicated Relationship Between Music and Foreign Language Learning: Nuanced Conditions Required for Cognitive Benefits Due to Music

Talia Greenberg

Oberlin College

Abstract

Many people enjoy listening to music while they study, but others find music distracting. Research about the effect of music on performance during a cognitive task mirrors the equivocal nature of this subjective debate. Across 3 experiments, music, either in the background or as an active encoding device, was found to have no effect on foreign language learning. In Experiment 1, participants studied foreign language vocabulary in silence, while listening to instrumental music, or while listening to music with lyrics. There was no effect of music on recall at immediate ($p = .52$) or delayed testing ($p = .80$). Participants in Experiments 2 and 3 listened to and then repeated foreign language phrases by speaking or singing them aloud. No significant differences were found in recall for phrases learned by singing and for phrases learned by speaking ($p = .827$). Experiment 3 assessed whether using a self-composed melody as a musical mnemonic device was more effective than singing a given melody in learning foreign language phrases. Recall for foreign language phrases sung to given melodies was not significantly different than recall for phrases sung to self-composed melodies at any retention interval (all *p*-values > .50). Despite finding only null results, this research sheds light on the question of when music may be successfully employed to enhance learning and suggests that familiarity of the music and difficulty of the learning task may be important factors.

The Complicated Relationship Between Music and Foreign Language Learning: Nuanced Conditions Required for Cognitive Benefits Due to Music

In studying difficult information, people can use mnemonic devices, which are learning techniques, to help trigger information about a particular concept. For example, mnemonic devices might involve rhymes, imagery, or even music to help us remember information. Music as a learning tool is something that we may have encountered at a very early age through nursery rhymes, such as using the tune of “Twinkle Twinkle, Little Star” to help us remember the sequence of the letters in the alphabet. So it is not surprising that music is thought by some to be useful in trying to remember information. However, while some people love using music as a way to focus their attention and block out external noise, others find music distracting when they are trying to study. Interestingly, this subjective debate about study preferences mirrors the equivocal nature of the research: there is conflicting evidence as to whether the presence of music helps or hinders performance on a cognitive task. The present study examined the use of music as a mnemonic technique.

Researchers have explored the impact of background music on cognitive performance. As an example, Schellenberg (2005) found that participants who sang or listened to music prior to a cognitive assessment, such as a test of spatial skills or processing speeds, performed better than participants who sat in silence. This effect was found to be mediated by mood and arousal: participants who preferred listening to a story over music showed the same cognitive benefits when listening to stories as did participants who found music more arousing when they heard music. This finding implied that while music in itself may not have a benefit, it may still yield improvements in performance by providing increased levels of arousal to listeners.

In contrast to the Schellenberg (2005) finding that the effect of music is mediated by arousal and mood, other researchers suggest that there is something unique about the rich structure of music beyond mere stimulation that leads to increased performance. An experiment by Wallace (1994) found music to enhance verbatim recall both at immediate testing and after a 20 min delay when the words being learned were presented through song. Participants who heard three verses of a ballad sung to the same melody recalled a greater percentage of the lyrics than did participants who heard these same three verses spoken. Importantly, participants in the music condition also performed better than participants who heard the text spoken rhythmically to the backdrop of a constant beat, which likely also provided stimulation; this showed that music was still beneficial even when controlling for arousal.

However, other studies have shown that not only can music be stimulating, but that this stimulation might be distracting too. Hughes, Rudin-Brown, and Young (2012) found that both listening to music and singing along to music decreased performance on a simulated driving task compared to baseline driving performance in the absence of music. Similarly, Brodsky and Slor (2013) found that novice drivers drove less safely when they listened to their preferred music in the car than when they drove without music. These experiments highlighted that music is not always beneficial in performing cognitive tasks: in fact, it might provide a severe disadvantage compared to silence. However, Brodsky and Slor (2013) also found that alternative music to the drivers' listening preferences was less distracting. Participants displayed fewer deficient driving behaviors when listening to the alternative music, which was designed to minimize distracting features such as melodies and lyrics. Thus, even seemingly minor changes in the stimuli were able to alter whether music provided cognitive benefits or whether it was distracting.

Indeed, in the same study that found music to enhance recall for ballad verses, Wallace (1994) found the opposite effect in follow-up experiments when there were seemingly only minor changes in the stimuli and procedure. In the first follow-up study, participants heard only one verse of the ballad, in contrast to the three that they had heard in the original experiment. This time, participants who had heard the ballad spoken recalled a greater percentage of the lyrics than participants who had heard it set to a melody. This might have been because participants did not have a chance to become familiar with the melody because they only heard it once, so they were not able to use it to their advantage to encode the lyrics. This explanation also makes sense in light of the second follow-up study. In the second follow-up study, participants heard three verses of a ballad just as before, but participants in the music condition heard each verse set to a different tune, rather than hearing three verses in a row that reflected the same melodic pattern, as was the case in the original experiment. Again, participants in the music condition performed worse than those who heard the verses spoken, likely due to their limited exposure to each melody. Although participants did not benefit from the musical presentation of the stimuli in these follow-up experiments, there is still more evidence that music can be helpful at any stage of the learning process.

Whereas Schellenberg (2005) found that listening to music before completing a cognitive task may improve scores, and Wallace (1994) showed that under the right conditions, the presence of music during the studying process can provide a benefit, another study highlighted that music may even provide an improvement in performance when it is presented after studying. An experiment by Judde and Rickard (2010) found that participants who listened to a musical excerpt 20 min after studying a list of 30 words performed better on a recall test a week after studying the list than did participants who had not listened to music. Thus, music presented post-learning may also have cognitive benefits, such as increased long-term memory. Yet Judde and Rickard (2010) also found that the effect was not present for participants who listened to music immediately after studying or 45 min after studying. Clearly, the findings were very mixed: while music has been found to confer a cognitive benefit, very specific conditions must be met in order to enable participants to use music as a tool to their advantage rather than finding the music distracting.

In most of the above studies, participants were passive listeners. Even in the Schellenberg (2005) and the Hughes et al. (2012) experiments where participants sang the music, the content that they were singing was not related to the criteria that they were tested on. However, a recent study by Ludke, Ferreira, and Overy (2013) had participants actively singing the content they were studying. In the Ludke et al. (2013) study, one group of participants was instructed to sing the material that they were learning, while other groups merely spoke the same material or spoke it rhythmically. Participants were then tested on the Hungarian phrases they had studied using a variety of recall measures to assess whether singing could be an effective learning tool. The researchers found that participants who studied phrases by singing them aloud were better able to recall the phrases than participants who had practiced by speaking the phrases or by speaking the phrases in rhythm. These findings suggested that singing could be an effective tool in facilitating learning.

The present research most closely aligned with the Ludke et al. (2013) methodology. In the present study, I sought to replicate and extend the Ludke et at. (2013) finding that in studying material, singing the material during study can improve recall. The present study used a series of three experiments to test the connection between using music and learning foreign languages, as one example of how music might improve performance on a cognitive task. Through this work, I sought to address three central questions: (1) does the presence of music help or hinder when performing a cognitive task? (2) is music more effective when it is used as part of the encoding process rather than merely in the background? and (3) do the effects of music on foreign language learning persist over time? The answers to these questions may inform theories of learning and effective study. The first experiment assessed the effect of listening to music on memory for foreign language vocabulary. The second and third experiments required participants to actively sing as a part of the studying process. The third experiment additionally included a composing condition, in which participants designed their own melodies to sing phrases on.

Experiment 1[1]

In Experiment 1, participants studied foreign language vocabulary in the presence or absence of background music. Some participants studied in silence; others listened to instrumental music while they studied; others listened to music with lyrics while they studied. Participants were tested on their memory for the foreign language vocabulary both immediately after they studied and the following day.

Based on the Schellenberg (2005) finding, I hypothesized that the presence of music during studying would improve recall, as it would provide stimulation, which could improve participants' endurance and focus during studying. However, I hypothesized that vocal music would be more distracting than instrumental music because the lyrics might interfere with the verbal cognitive task that participants were trying to complete. This hypothesis was also supported by the Brodsky and Slor (2013) finding that music designed not to include melodies or lyrics was less distracting than popular songs. Thus, the optimal condition would be listening to

[1] In Experiment 1, there were two research assistants: Melissa Kravets and Nora Kipnis. These research assistants helped compile stimuli and run sessions of the experiment.

instrumental music. Additionally, I hypothesized that any effects present at immediate testing would also be present and might even be more drastic at delayed testing based on the Wallace (1994) finding that music enhanced recall at both immediate and delayed testing and the Judde and Rickard (2010) finding that the benefits of music were present even a week after the initial study session.

Methods

Participants

Introductory Psychology students at Oberlin College (N = 32; 12 men, 19 women, 1 genderqueer person) participated in the study. Consent forms were collected from the participants prior to participation. Participants received course credit as compensation for their participation and were given additional course credit for the completion of an online test the day after the study session. Out of 32 participants, 23 participants (72 %) completed the follow-up online test.

Design

The study used a mixed-model design. The between participant factor was study condition: one group listened to music with lyrics while they studied (*Vocal Music*), a second group listened to instrumental music while studying (*Instrumental Music*), and a third group studied in silence (*No Music*). Participants were randomly assigned to one of these three groups. The within participant factor was retention interval. Participants were tested at two retention intervals: immediately after each study phase (*T1*), and before 5:00 PM on the day after the study session (*T2*). The dependent variable was the number of verbs participants correctly recalled on each of the recall tests.

In addition, participants completed an operation span task at the beginning of the study as a measure of individual difference.

Materials and Apparatus

The operation span task was presented through a timed PowerPoint presentation, and participants recorded responses in a paper packet. The operation span task consisted of three lists of seven English nouns (see Appendix A) with each word displayed for 1 s. Each word was followed by a mathematical equation (see Appendix A) that was either true or false; the equations were each displayed for 20 s. The paper packet included space to record whether each equation was true or false, followed by an area to record all recalled nouns from the operation span task. The paper packet included instructions on each page reminding participants not to flip forward or backward in the packet in order to prevent cheating during the recall task.

The foreign language material to study consisted of three lists of 20 Hebrew verbs (written in English letters) and their English translations (see Appendix B). The order of the lists was counterbalanced between participants. The verbs were presented in the paper packets, with all 20 words in each list shown together on one page.

The PowerPoint displayed instructions directing participants to study the verbs in their paper packets. In the *Instrumental Music* and *Vocal Music* conditions, the PowerPoint also played music during the study phases. Blank paper and blank index cards were also provided to participants as resources to create study materials.

The music played was cover songs of familiar Beatles tunes so as to emulate real-world conditions, in which students who listen to music while studying are likely to listen to music they know well. Music for the *Instrumental Music* condition was collected from karaoke versions of Beatles songs on Spotify. The music played in the *Vocal Music* condition included the same karaoke tracks overlaid with a vocal track of the melody sung by the experimenter.

The recall tests presented during the experiment were administered in the paper packets with all 20 verbs from the list presented on one page. The verbs were all followed by a blank space where participants could record translations they recalled. On the recall test page in the packet, the verbs were presented in a different order than they had been when initially presented with their English translations. The paper packets also included instructions reminding participants not to flip to another page in the packet so as to prevent cheating. The PowerPoint presented instructions during the recall test phase that directed participants to write down all recalled words. The next-day recall test was presented through Qualtrics and was sent out via email. In this test, all 60 words were presented together in one list. All recall tests were cued with the Hebrew verbs and included a blank space for the English translation, and the words on each recall test appeared in a randomized order.

A short questionnaire was used to collect demographic information, and to assess musical expertise, knowledge Hebrew, knowledge of languages besides English, and study habits. This survey was presented on the last page of the paper packets.

Procedure

All laboratory sessions were held in a room with many tables and chairs so that many participants could complete the study at once. The room was also equipped with a projector and a sound system through which to present the PowerPoint. Participants were asked to review and sign a consent form at the laboratory site before taking part in the study.

Participants first completed the operation span task. During the operation span task, participants were presented with three lists of words to study. Each list contained 7 words that were presented in quick succession through a PowerPoint, separated by mathematical equations. Participants were instructed to judge whether each equation was true or false as it appeared and to record their answer in their paper packet. After the last equation in each list, participants were given 1 min to recall as many words from the list as they could.

Next, participants began a study phase for the Hebrew verbs. They were instructed to use any methods they wished to study the 20 verbs listed in their paper packets and were given 10 min to study. Participants in the *Vocal Music* condition listened to vocal music during this phase, and participants in the *Instrumental Music* condition heard instrumental music, while the *No Music* group studied in silence. After the study phase ended, participants were given 5 min to record all of the translations they remembered on the cued recall page in their packets. No music was played during the recall test, regardless of condition. The recall test was followed by another 10 min study phase and 5 min recall test for the second list of Hebrew verbs, and then a final 10 min study phase and 5 min recall test for the third list.

After completing the final recall test, participants completed a short questionnaire detailing their demographic information and prior experience with music and foreign languages. The questionnaire was self-paced. The next morning, participants received the *T2* recall test via email and had until 5:00 PM that day to complete the test. This test included all 60 verbs from the study in a randomized order and again presented the Hebrew verbs as cues with a space to record the English translations. This final recall test did not have any time limit.

Results & Discussion

Participants' scores on each list were coded as the number of items correctly recalled when cued with Hebrew verbs from the studied lists. Total scores for *T1* and *T2* were calculated by adding the three list scores for *T1* and by counting the total number of items correctly recalled for *T2*. A one-way between groups ANOVA was used to analyze whether study condition (*No Music*, *Instrumental Music*, or *Vocal Music*) had an effect on participants' recall score. Recall score was measured through multiple dependent variables: the participants' scores on each of the three lists as well as the total scores at each of the time intervals. Surprisingly, whether participants studied in the *No music*, *Instrumental music*, or *Vocal Music* condition made no difference in the number of words they recalled on List 1, $F(2, 29) = 0.348$, $p = .71$ The same was true for the other dependent variables: there was no difference in recall on List 2, $F(2, 29) = 0.718$, $p = .50$; on List 3, $F(2, 29) = 0.920$, $p = .41$; on the total score at *T1*, $F(2, 29) = 0.677$, $p = .52$, and on the total score at *T2*, $F(2, 29) = 0.225$, $p = .80$. See Table 1 for mean recall scores on each test by study condition. This finding was contrary to my hypothesis that the presence of music in the background during study would improve recall.

A two-way mixed design ANOVA was used to analyze the interaction between study condition and retention interval. Even though the main effect of study condition on recall score was not significant, I had hypothesized that those participants who had studied with music would perform better at delayed testing than participants in the *No Music* condition. Surprisingly, the interaction was not significant: whether participants studied in the *No Music*, *Instrumental Music*, or *Vocal Music* condition made no difference in the decrease in the number of words they recalled between immediate testing and next day, $F(2, 29) = 0.68$, $MSe = 204.77$, $p = .61$. Despite past research suggesting that listening to music while completing a cognitive task may improve performance, in the present experiment, music neither helped nor hindered participants as they studied foreign language vocabulary.

A paired sample t-test showed that, as expected, there was a significant main effect of time, with a sharp decrease in memory between immediate testing (*Mean* = 42.75) and next-day testing (*Mean* = 10.50), $t(31) = 9.165$, $p = .00$. This result was not surprising as we have a tendency to forget over time, so participants were likely to recall more immediately after studying than after more time had elapsed.

Though individual difference scores had been collected about traits such as music expertise, language expertise, and study habits, no differences in recall were found between participants with high or low expertise or between participants who were accustomed or unaccustomed to listening to music while studying. Additionally, although operation span task scores had been collected in order to account for individual difference in memory, it could not be used as a covariate in these analyses because it was not correlated with most of the dependent measures. Whereas operation span task correlated with recall score on list 2, $r = .367$, $p = .04$, and with total recall score at T2, $r = .405$, $p = .02$, it did not correlate with recall score on list 1, $r = .138$, $p = .45$, with recall score on list 3, $r = .197$, $p = .28$, or with total recall score at T1, $r = .255$, $p = .16$. Moreover, random assignment was not successful in terms of the operation span task scores. A one-way ANOVA showed that the operation span task scores were significantly different between the three study conditions: *No Music* (*Mean* = 14.92), *Instrumental Music* (*Mean* = 17.60), and *Vocal Music* (*Mean* = 14.30), $F(2, 29) = 3.847$, $p = .03$. This result may point to the failure of random assignment as a confounding factor, which might explain in part why no differences were found between study conditions.

Experiment 2

Although background music had no impact on memory performance in Experiment 1, a second experiment was run in order to assess whether music could still be an effective mnemonic device when it was used actively in the encoding process. In the Ludke et al. (2013) study, singing was found to enhance recall for Hungarian phrases over merely speaking the phrases aloud or speaking them rhythmically, which showed that setting foreign language phrases to sung melodies could be an effective tool for learning them. However, these results were only present in the dependent measures that involved recalling the Hungarian phrases (the Hungarian production test and the delayed-recall Hungarian conversation task); Ludke et al (2013) failed to find this result in the dependent measures that involved recalling the English translations (the English recall test). Interestingly, in the singing condition, participants only sang the Hungarian phrases; the English counterparts were spoken normally even in the rhythmic speaking and singing conditions. The fact that singing showed benefits only in situations where participants had to recall the Hungarian motivated the question of whether singing is helpful only in recalling the language that was sung rather than providing benefits for the spoken translation of the sung phrase as well. Predictions regarding this question are based on the encoding specificity principle, which states that memory is best when conditions at retrieval match conditions at encoding (Tulving & Thomson, 1973). Here, that would manifest if participants used the same strategy, namely singing, at test that they had used during study to produce better recall. In this follow-up study to the Ludke et al. (2013) paper, I aimed to replicate the Ludke et al. (2013) findings and to test the hypothesis that the benefits of singing on recall are specific to the sung language.

In the present experiment, participants studied Hebrew phrases and their English translations. Instead of singing solely the foreign language phrases as in the Ludke et al. (2013) experiment, participants in this experiment studied phrases in each of four conditions. In one condition, they spoke both the Hebrew and the English, in another, they sang only the Hebrew and spoke the English, in a third condition, participants sang the English and spoke the Hebrew, and in a fourth condition, both languages were sung. The dependent measure was recall in both languages for the Hebrew-English phrase pairs that the participants had studied. Participants were tested on their recall twice: once immediately after studying the phrases and again after a delay.

I hypothesized that Experiment 2 would replicate the Ludke et al. (2013) experiment to show that singing does increase recall for foreign language phrases. However, I also hypothesized that this benefit of singing would be specific to recall tests in which the language being recalled was sung during the study phase. For example, if Hebrew was sung and English was spoken during study, I predicted that this would result in a boost in memory performance for recalling the Hebrew phrases when cued in English without improving performance for recalling the corresponding English phrases when cued in Hebrew due to the encoding specificity principle (Tulving & Thomson, 1973). Finally, I hypothesized that though there would be a decrease in recall over time across all conditions, the effects present at immediate testing would persist, as in the original study by Ludke et al. (2013) in which the results of the delayed-recall Hungarian conversation task mirrored those of the immediate Hungarian production test.

In the Ludke et al. (2013) experiment, Hungarian was used for three reasons: most of the participants likely had no prior exposure to Hungarian, Hungarian and English do not share cognates, and the two languages include different phonemes. Hebrew also fulfills these three criteria. Thus, Hebrew was used in the present study instead of Hungarian both for convenience and for generalizability: the expectation was that if the Ludke et al. (2013) findings were applicable to learning Hebrew, the results might be generalizable to learning other languages as well.

Methods

Participants

Introductory Psychology students at Oberlin College (N = 31; 7 men, 23 women, 1 genderqueer person) participated in the study. Consent forms were collected from the participants prior to participation. Participants received course credit as compensation for their participation. Although the original dataset included 32 participants, ultimately, one survey was eliminated due to procedural errors, resulting in a dataset of 31 participants.

Design

The study employed a 4 (Study Technique) x 2 (Retention Interval) within participant design. Each participant studied 24 transliterated Hebrew-English phrase pairs under 4 study conditions, with 6 phrase pairs in each condition. In the *Hebrew Sung* condition, participants sang the Hebrew phrase but spoke their English counterparts; in the *English Sung* condition, participants did the reverse, speaking the Hebrew and singing the English; in the *Both Sung* condition, participants sang both the Hebrew and English phrases; in the *Neither Sung* condition, participants spoke both phrases. Each list of 6 phrases appeared in each study condition between participants using a Latin square design. For example, if one participant heard the phrase "How old are you?" presented in the *English Sung* condition, another would hear it presented in the

Hebrew Sung condition, while another participant heard it presented in the *Both Sung* condition and another heard it in the *Neither Sung* condition. Participants were randomly assigned to a given presentation of each list of words.

Cued recall was assessed at two retention intervals: immediately after the study session (*T1*), and 10 min after the study session (*T2*). The dependent variable was the number of syllables from the phrases correctly recalled on each test.

Materials and Apparatus

The study stimuli consisted of 24 Hebrew-English phrase pairs with 6 in each study condition (*Hebrew Sung*, *English Sung*, *Both Sung*, and *Neither Sung*). The assignment of phrases to study conditions was counterbalanced between participants using a Latin square design so as to avoid specific item effects. Five of the phrases used were taken directly from the stimuli in the Ludke et al. (2013) study. Others did not translate well to Hebrew or resulted in phrases with too many or too few syllables. Therefore, 19 similar phrases were substituted. All of the phrases can be found in Appendix C.

The Hebrew-English phrase pairs were presented in a timed PowerPoint. Each slide included one transliterated Hebrew phrase and its corresponding English translation. The phrase pair was written on the slide (the Hebrew was written in English letters) alongside a recording of the experimenter speaking or singing the phrase pair, depending on the condition.

The experimenter who performed all recordings was a native speaker of American English who was also fluent in Hebrew. She had some training as a vocalist and felt comfortable singing. The melodies used in the singing conditions were adapted from those employed by Ludke et al. (2013) to fit the emphases of the Hebrew and English phrases, but the tonal patterns in the melodies were preserved. These melodies used diatonic intervals, and each melody stayed within the range of an octave to maximize ease of singing.

The two cued recall tests were presented on paper. The test included 12 of the phrases (with 3 phrases from each study condition) cued in English with a space for participants to fill in the Hebrew translation, followed by the other 12 phrases (with 3 from each study condition) cued in Hebrew so that participants could write down the corresponding English phrase. The cue language of each phrase was consistent for the two recall tests for each participant. For example, if "How are you?" was cued in English at *T1*, it was also cued in English at *T2*. However, the cue language for each phrase was counterbalanced between participants so that if one participant recalled "How much does it cost?" based on an English cue, another recalled that same phrase based on a Hebrew cue. In addition, the order of the words on each test was randomized.

A short questionnaire was used to collect demographic information. The questionnaire also included questions about each participant's musical expertise, knowledge of Hebrew, knowledge of languages besides English, and aptitude for foreign language learning. This survey was presented on paper.

Procedure

All laboratory sessions were held in one of two quiet rooms, with one participant per room. Participants were asked to review and sign a consent form at the laboratory site before taking part in the study. During the session, each participant viewed a timed PowerPoint presentation on a laptop computer. After beginning the PowerPoint, the experimenter left the room and waited outside the laboratory site.

Participants first completed a practice study session with Spanish-English phrase pairs (see Appendix D). During the practice session, participants studied one Spanish-English phrase pair under each of the four conditions (*Hebrew Sung*, *English Sung*, *Both Sung*, and *Neither Sung*). After the practice study session, each slide presented one Hebrew phrase (written in English letters) and its English translation with an audio recording of the phrase pair. Participants were instructed to repeat each phrase pair exactly as they heard it. All words were grouped together on the list according to study condition and were presented as one long list with no breaks between study conditions.

Each phrase pair was followed by a pause, allowing the participants time to repeat the material aloud. After the pause, the phrase was repeated in the recording, with another space for the participant to practice repeating the material before proceeding to the next phrase. This allowed participants to practice each phrase twice. The slide for each phrase was displayed for a total of 30 s, with both iterations of the audio-recorded phrase playing at 15 s intervals.

Immediately after completing the study phase, participants were given the first cued recall test. Participants were instructed to complete the portion of the test with English cues before moving on to the section with Hebrew cues. The recall tests were self-paced; participants reported back to the experimenter for further instructions after completing both tests.

After completing an unrelated cognitive task, participants took a second cued recall test. This test occurred 10 min after completion of the first test. Again, participants were not timed while completing the recall test. After completing the second cued recall test, participants completed a short questionnaire detailing their demographic information and prior experience with music and foreign languages. The questionnaire was also self-paced.

Results & Discussion

Recall scores for each condition were coded as the number of syllables correctly recalled in each phrase from that condition; each correct syllable was worth one point. Syllables that were partially correct (for example, syllables that had the right consonant sound with the wrong vowel sound or vice versa) were given half a point.

Because there were many instances of participants failing to recall any syllables of a phrase, the recall test scores were all recoded into dichotomous variables based on whether each participant had correctly recalled any syllables (coded = 1) or not (coded = 0) for a given phrase. The extreme skew of the distributions of all recall scores ruled out the use of normality-based testing; instead, the non-parametric test Cochran's Q was used.

I had predicted that I would replicate the Ludke et al. (2013) finding that participants would recall phrases learned in the singing condition better than they recalled phrases learned in the speaking condition. However, there was no significant difference in recall scores between phrases that were learned in the *Neither Sung* condition and phrases that were learned in the *Both Sung* condition at *T1*, $p = .68$, or at *T2*, $p = .55$ (see Table 2 for counts). There was also no difference in recall scores between phrases that were learned in the *Neither Sung* condition and phrases that were learned in the *Hebrew Sung* condition at *T1*, $p = .88$, or at *T2*, $p = .67$; this analysis more closely parallels the Ludke et al. (2013) study in which participants who sang only the foreign language phrases were compared to participants who had spoken the entire phrase pair.

I had also expected to find a main effect of time, because there is a tendency to forget information over time. However, the data ultimately showed no significant differences between recall scores at *T1* and *T2* in any of the study and recall conditions. Additionally, there was no significant interaction between retention interval and study condition. This surprising result is probably due to the fact that there was little recall to begin with, so there could not be much decrease in memory between *T1* and *T2*. The fact that I did not replicate the Ludke et al. (2013) findings is likely due to the large number of zeroes in the dataset. Participants performed at floor level on the recall tests, so it was difficult to statistically parse out differences due to restricted range issues. The next direction for this research must be redesigning the experiment such that participants can achieve higher recall scores.

Additionally, I had expected to find a main effect for cue language. Writing down the English translation for a given Hebrew phrase should have been an easier task than recalling the Hebrew translation for an English phrase because English was the more familiar language for the participants. Surprisingly, across all study conditions, there were no significant differences between recall scores on phrases that were cued in English compared to the recall scores for phrases cued in Hebrew.

Finally, I had expected to find higher recall scores when participants had sung the phrase in the language that they were recalling (for example, if a phrase studied in the *English Sung* condition were cued at testing in Hebrew so that participants would need to recall the English, which they had sung earlier). However, the data showed no significant differences between phrases for which participants had sung the language they recalled and phrases for which participants spoke the recall language at *T1*, $p = 1.00$, or at *T2*, $p = .75$. Additionally, there was no significant interaction between this match criterion and retention interval.

Although individual difference scores on traits such as music expertise and language expertise had been collected, no differences were detected. This lack of result was likely due in part to the large number of zeroes in the data.

It is puzzling that recall in the present study was at floor given that participants in the Ludke et al. (2013) experiment did recall many of the phrases that they had studied. This is especially surprising because the methods used in Experiment 2 closely replicated the Ludke et al. (2013) methods: participants studied approximately the same number of phrases set to the same tonal patterns for the same amount of time. The only stark difference in procedure and stimuli between these two experiments was the use of Hebrew instead of Hungarian in the present study. However, it seems unlikely that there is a language effect given that both languages met the same criteria of being foreign to participants, having no cognates with English, and using different phonemes than English. Thus, I had not expected to have such limited recall in the present study.

Experiment 3[2]

Experiment 2 was not successful in replicating the Ludke et al. (2013) findings, so in Experiment 3, I sought to provide participants with an even more effective mnemonic device. According to the generation effect, the best mnemonic devices are ones that learners create for themselves (Begg & Snider, 1987). For example, to remember that the Hebrew word *l'ashen* means "to smoke," I could tell you that this comes from the Hebrew word *esh*, meaning fire, so it therefore has a similar sound; but if you were to instead draw your own connection that makes sense to you, it would be a better mnemonic device than the one that I provided.

The Ludke et al. (2013) study used musical mnemonic devices, but this research did not incorporate the generation effect to find out whether, as with other mnemonic devices, creating

[2] Experiment 3 was designed and run in collaboration with Chris Bromberg, Jennifer Carpenter, Siena Castañares, and Jake Rivas as part of a lab group. They have all given me permission to use it as a part of this manuscript.

one's own musical mnemonic device is more effective than using a given one. In the present study, I aimed to expand on the Ludke et al. (2013) findings by incorporating the generation effect and to generalize this research by using Japanese, rather than Hungarian phrases. Here, participants studied Japanese-English phrase pairs in three conditions: speaking, singing, and composing. They were then tested on their recall of the Japanese phrases to assess whether using a self-composed melody in study provided a benefit in recall. In the original study, one of the dependent measures used was a delayed recall test, in which participants' learning was reevaluated after a 20 min delay (Ludke et al., 2013). In the present study, recall tests were administered after an even longer delay the day after the lab session in order to more closely mimic retention in a real-world learning environment, where studying might happen prior to the day of the test.

Given the ubiquity of the generation effect, I hypothesized that composing one's own melody to use as a mnemonic device and singing foreign language phrases set to this melody should be the most successful memory tool, leading to the highest recall score (Begg & Snider, 1987). If this strategy were found to be more effective than simply singing a given melody, it could be applied as a study technique for many language learners. As for the time delay, I hypothesized that although recall would decrease over time across all conditions, any enhancements that singing and composing might offer in increasing recall scores would continue to provide a benefit in the delayed testing conditions. This hypothesis was supported by the original study, in which Ludke et al. (2013) found effects both at immediate testing and after a delay.

Methods

Participants

Introductory Psychology students at Oberlin College (N = 42; 17 men, 24 women, 1 genderqueer person) participated in the study. Although there had been 52 participants to begin with, nine were eliminated due to procedural errors, such as an audio malfunction, and one participant was eliminated due to prior knowledge of Japanese. This resulted in a final dataset of 42 participants. Consent forms were collected from the participants prior to participation. Participants received course credit as compensation for their participation and were given additional course credit for the completion of an online test the day after the study session. Out of 42 participants, 23 participants (55%) completed the follow-up online test.

Design

Using a within-participant 3 (Study Technique) x 3 (Retention Interval) factorial design, participants studied transliterated Japanese-English phrase pairs in each of three conditions: *Speaking*, *Singing*, and *Composing*. The order of the conditions presented in the PowerPoint was counterbalanced across participants using a Latin square design, and participants were randomly assigned to a given order of conditions.

Cued recall was assessed at three retention intervals: immediately after the study session (*T1*), 10 min after the study session (*T2*), and before noon on the morning after the study session (*T3*). The dependent variable was the number of syllables of the Japanese phrases correctly recalled when cued with the corresponding English translations.

Materials and Apparatus

21 Japanese-English phrase pairs were used, with 7 in each study condition (*Speaking*, *Singing*, and *Composing*). Some of the phrases were taken from the Ludke et al. (2013) study. The length of each phrase (number of syllables per phrase) was matched across the three study technique conditions (see Appendix E).

The Japanese-English phrase pairs were recorded on a laptop by an experimenter. The experimenter recorded herself speaking the Japanese and English phrases for the *Speaking* and *Composing* conditions. For the *Singing* condition, the experimenter recorded herself singing the Japanese phrases and speaking the English phrases. The experimenter who performed all recordings was a native speaker of American English, with a working proficiency in Japanese. She was not formally trained as a vocalist but felt comfortable singing.

The pre-composed melodies used in the *Singing* condition were adapted from those employed by Ludke et al. (2013). The melodies in the present study used diatonic intervals, and each melody stayed within the range of an octave to maximize ease of singing. The contour of the melodies reflected the patterns of syllabic emphasis naturally present in the Japanese and English languages.

Stimuli were presented through a timed PowerPoint presentation. Each slide presented one Japanese phrase (written in English letters) and its corresponding English translation. The recording of the experimenter speaking or singing the Japanese-English phrase pair was played alongside the corresponding slide.

Three cued recall tests of the Japanese-English phrase pairs were administered to measure recall for the Japanese phrases. The cued recall test presented the English phrases with a space to write down the corresponding Japanese phrase. The order of the English phrases was randomized across study conditions for each recall test in order to mimic real-life study and test conditions in an academic setting. The first two cued recall tests were administered on paper. The third cued recall test was administered online through Google Forms.

A short questionnaire was used to collect demographic information. The questionnaire also included questions about each participant's musical expertise, knowledge of Japanese, knowledge of languages besides English, and aptitude for foreign language learning.

Procedure

Participants were asked to review and sign a consent form at the laboratory site before taking part in the study. All laboratory sessions were held in one of two quiet rooms, with one participant per room. During the session, each participant viewed a timed PowerPoint presentation on a laptop computer. Slides within the PowerPoint presented one transliterated Japanese phrase and its English translation with audio recordings of the phrase pair. The PowerPoint contained three segments corresponding to the three study conditions: *Speaking*, *Singing*, and *Composing*. Each segment began with a slide instructing the participants to repeat the following phrase pairs aloud using the study technique for that particular segment. Each segment of the PowerPoint contained seven different Japanese-English phrase pairs. After beginning the PowerPoint, the experimenter left the room and waited outside the laboratory site.

Before studying the Japanese-English phrase pairs, participants completed a practice study session with Spanish phrases, similar to the practice session for Experiment 2 (see Appendix F). Participants practiced two Spanish phrases under each of the three study conditions (*Speaking*, *Singing*, and *Composing*). During this tutorial portion of the PowerPoint, participants also received audio instructions on how to compose short melodies to correspond with each phrase pair in the *Composing* condition. Two examples were given of how to do so (see Appendix G).

In the *Speaking* condition, the audio recordings embedded in the PowerPoint consisted of an experimenter speaking each Japanese phrase and its English translation. Each Japanese-English phrase pair was followed by a pause, allowing the participants time to repeat the material aloud. After the pause, the phrase was repeated in the recording, with another space for the participant to practice speaking the material before proceeding to the next phrase. This allowed participants to practice each phrase twice. The slide for each phrase was displayed for a total of 30 s, with both iterations of the audio-recorded phrase playing at 15 s intervals.

In the *Singing* condition, the audio recordings embedded in the PowerPoint consisted of an experimenter speaking the English translations and singing the Japanese phrases to short melodic sequences. Again, each phrase was followed by a pause and then repeated in the recording followed by another pause before proceeding to the next phrase. The slide for each phrase was again displayed for a total of 30 s, with one iteration played 15 s after the last.

In the *Composing* condition, participants heard an audio recording in which an experimenter spoke the Japanese phrase and its English transliteration, just like in the speaking condition. However, participants were instructed to compose short melodies during the pauses in the audio recording and to practice singing the phrases aloud to their self-created melodies, one pair at a time. Again, each phrase was displayed for 30 seconds total, with the recording repeated at 15-second intervals.

Immediately after completing the study phase, participants were given the first cued recall test. Participants were encouraged to use any methods they chose (e.g., speaking or singing aloud) during testing. There was no time limit on the cued recall test: participants reported back to the experimenter once they finished.

After completing an unrelated cognitive task, participants took a second cued recall test. This test occurred 10 min after completion of the first test and again had no time limit. After completing the second cued recall test, participants completed a short questionnaire detailing their age, gender, musical expertise, knowledge of Japanese, knowledge of languages besides English, and language learning abilities. This questionnaire was self-paced.

The test at the next-day retention interval was an online test, administered via Google Forms. A link to the test was emailed to participants the morning after they completed the in-lab session and expired at noon of that day. Participants were again cued with the English phrases and were given unlimited time to recall the Japanese translations.

Results & Discussion

Recall scores for each condition were coded as the number of syllables correctly recalled in each phrase from that condition; each correct syllable was worth one point. Syllables that were partially correct (for example, syllables that had the right consonant sound with the wrong vowel sound or vice versa) were given half a point. No points were deducted for misspellings that resulted in the same pronunciation (such as “dokko” instead of “doko”).

Like in Experiment 2, recall scores were very low, with many instances of no syllables correctly recalled. Therefore, the recall test accuracy scores were all recoded into dichotomous variables based on whether the participant had correctly recalled any syllables (coded = 1) or not (coded = 0). The extreme skew of the distributions of all recall score variables ruled out the use of normality-based testing; instead, the non-parametric test Cochran’s Q was used. For the individual difference variables, both language expertise and musical expertise were transformed using a median split. Participants were divided into a low expertise group and a high expertise group.

I had predicted that Japanese phrases studied in the *Composing* condition would be better recalled than phrases studied in the *Speaking* and *Singing* conditions due to the generation effect (Begg & Snider, 1987). Contrary to my expectations, there was no difference between recall scores for phrases learned in the *Singing* and *Composing* conditions at *T1*, $p = 1.00$, at *T2*, $p = .65$, or at *T3*, $p = .51$ (see Table 3, Table 4, and Table 5 for respective counts). Moreover, participants recalled phrases learned in the *Speaking* condition significantly better than phrases learned in the *Singing* and *Composing* conditions at *T1*, $p = .03$ and at *T3*, $p = .00$ (see Table 3 and Table 5 for respective counts). At *T2*, there were only marginally significant differences in cued recall between the three study conditions, with most phrases recalled in the speaking condition, $p = .10$ (see Table 4 for counts). The fact that recall scores in the *Speaking* condition were higher than the scores in the *Singing* condition is in direct contrast to the Ludke et al. (2013) findings.

I had also predicted that there would be a significant main effect of time: that participants would recall the most items at *T1*, fewer items at *T2*, and the fewest items at *T3*. Although no significant differences were found across time delay for the *Speaking* condition, $p = 1.000$ or the *Composing* condition, $p = .311$, there was an effect in the *Singing* condition. Participants showed better recall at *T1* and *T2* for phrases studied in the *Singing* condition than at *T3*, $p = .001$ (see Table 6 for counts). This means that whereas participants could recall phrases learned in the *Speaking* and *Composing* conditions just as well the day after the test as they could on the day they learned them, in the *Singing* condition, there was a drop-off in recall from after the day participants learned the words. Like in Experiment 2, I might have found less of an effect of time than I had anticipated due to the low recall scores: because recall was at floor at *T1*, there was little room for a decline in recall scores at *T2* and *T3*.

I also wanted to investigate the impact of musical expertise on the cued recall of Japanese-English phrase-pairs in the *Speaking*, *Singing*, and *Composing* conditions. I found that music experts in the *Singing* condition remembered significantly fewer phrases at *T3* than at *T1* and *T2* $p = .007$ (see Table 7 for counts), while non-experts did not have significantly different recall scores between *T1*, *T2*, and *T3*. Musical expertise had no effect on recall scores in the other study conditions. I had also predicted that experts at language learning would recall more items than non-experts. However, I did not find a significant main effect for language expertise. Interestingly, language experts recalled fewer phrases learned in the *Composing* condition at *T3* than at *T1* and *T2*, $p = .039$ (see Table 8 for counts).

One limitation of Experiment 3 was the specific item effect. Because each word was always studied under the same condition (*Speaking*, *Singing*, or *Composing*), some of the results found might not be due to the study condition but rather due to the words on the list. For example, if the words on the list in the *Speaking* condition happened to be easier, this might have falsely shown a benefit of speaking over singing.

Summary and Concluding Discussion

Surprisingly, I found no benefit of music- either in the background or as an active encoding device- across the three studies. Experiment 1 showed no difference between recall scores of participants who had studied foreign language vocabulary while listening to vocal music, the scores of those participants who had studied while listening to instrumental music, and the scores of those who had studied in silence. This contradicted my hypothesis that music would enhance memory by providing stimulation. The present findings are in direct contrast to the Schellenberg (2005) study, which found that listening to background music before study improved recall. Considering the Brodsky and Slor (2013) findings might shed light on the inconsistencies. Brodsky and Slor (2013) found music to be distracting while completing the cognitive task of driving. In Experiment 1, the presence of music was not found to impair cognition, as in the Brodsky and Slor (2013) experiment; it simply failed to provide a benefit. Because Schellenberg (2005) suggested that the benefits of music were due to increased arousal, it is possible, therefore, that any benefit that the music might have been able to confer through increased arousal in the present study was counterbalanced by the distracting nature of the music. The benefits found in the Schellenberg (2005) experiment were not neutralized by distraction because music was used before the studying process, allowing participants to benefit from the stimulation that music provided without being distracted by it during the study phase.

The Judde and Rickard (2010) findings also highlighted that listening to music may be helpful only when doing so is separated from studying so as to avoid distraction during study. Judde and Rickard (2010) found that listening to music 20 min after studying boosted recall scores; however, listening to music immediately after studying did not have the same positive effect. This may be explained by the distracting nature of music: when played directly after study, music may have interfered with the rehearsal stage of committing material to memory. But once this material was encoded, as in the delayed listening condition, distraction was no longer an issue, so the benefits of music were able to outweigh the distraction.

The distraction of having music present during encoding might have been particularly problematic in Experiment 1 given the familiarity of the music. Brodsky and Slor (2013) found that familiar music was more distracting to drivers than unfamiliar music. In Experiment 1, the Beatles music used was likely to be familiar to participants. This may have increased the level of distraction for participants who studied in one of the music conditions and in turn may have minimized the benefit of music.

Experiments 2 and 3 similarly failed to support the literature and failed to support my hypotheses. I did not replicate Ludke et al.'s (2013) finding that singing foreign language phrases resulted in better recall than merely speaking these phrases. In Experiment 2, I found no difference between speaking and singing, and in Experiment 3, speaking actually resulted in better recall than singing. These findings also failed to support the Wallace (1994) study, which found better verbatim recall for ballad lyrics among participants who had learned the lyrics set to a melody than for participants who had initially heard the lyrics spoken. Furthermore, I had hypothesized in Experiment 3 that composing a melody for a foreign language phrase would lead to better recall for that phrase than did singing it with a provided tune. My hypothesis was based on the generation effect (Begg & Snider, 1987). However, this hypothesis was not supported, as no difference was found between singing and composing, and speaking was found to be more effective than composing.

In Experiment 2, the encoding specificity principle led me to hypothesize that the match between the language that was sung and the language being recalled was critical in order for the music to have a positive effect (Tulving & Thomson, 1973). Yet the hypothesis was not supported; there were no differences between recall for phrases where the cue language had been sung, phrases where the cue language was spoken, phrases where the recall language was sung, and phrases where the recall language was spoken. This might in part have been due to the low recall across all conditions, which made it difficult to detect differences.

The fact that performance on the recall tests was at floor level, even at immediate recall, was a limitation of Experiments 2 and 3. The low recall scores were surprising, given that participants in the Ludke et al. (2013) experiment studied list of approximately the same number of phrase pairs and were able to recall some of these phrases in a variety of recall tasks. Therefore, one would have expected recall in the present study to be better than what I found. My results might be explained by the Wallace (1994) findings that music is helpful in encoding only when participants have had enough time to learn the melody. Although participants in Experiments 2 and 3 listened to each sung phrase twice, perhaps they needed to hear more iterations of the melody in order for the music to become familiar enough that they could have used it as a mnemonic device. However, participants in Experiment 2 were exposed to the same musical stimuli as and heard each melody the same number of times as in the Ludke et al. (2013) study, so it is surprising that recall was so much lower in in the present study.

Alternatively, my results might be explained by the difficulty of the task. Perhaps the foreign language phrases were much more difficult to recall than the items recalled in the Wallace (1994) and other studies. Again, this seems surprising given that the Ludke et al. (2013) experiment used similar phrase pairs and did find recall. In the present study, the low recall scores at immediate testing left little room for participants' scores to drop off as time progressed, which made it difficult to detect differences between recall scores across the different retention intervals. Therefore, I was not able to find evidence supporting my hypothesis that while all scores would drop over time, singing and composing would provide comparative advantages in recall over speaking even after a delay.

The equivocal nature of the research in the field highlights that details in the procedure and in stimulus construction can produce opposite results, so perhaps it is not terribly surprising that no replication effects were found across the three experiments. Future directions for this research may focus on effect of the familiarity of the music on recall. In Experiment 1, it was possible that the lack of benefit of background music was due to the familiar nature of the musical stimuli, which might have distracted participants and negated the positive effect of musical stimulation. A replication of this experiment with less familiar music might be successful in finding a benefit of listening to background music while studying. In contrast, in Experiments 2 and 3, the musical stimuli were not familiar enough to be useful as encoding tools. Setting the foreign language phrases to more familiar melodies, such as pop songs or famous jingles, might improve memory for these phrases, which would enable participants to perform above floor level on the recall tests.

Another future direction could be to analyze how the difficulty of the task mediates the effectiveness of using music as a mnemonic device. One aspect of the present study that was truly successful was creating a difficult learning task. In Experiments 2 and 3, performance on the recall task was very low, even in the control conditions in which participants spoke the stimuli. Perhaps when the recall task itself is difficult enough, participants' cognitive load is too high to be able to use the additional stimulus of music as a mnemonic device. However when the task is too easy, music is likely not necessary in order to perform well. Extensions of this research could seek to identify optimal levels of task difficulty for implementing music as a mnemonic device.

In summary, the research about how music might enhance or interfere with performance on a cognitive task is still quite mixed. Whereas some studies have found benefits to music, others have found music to be distracting. It seems that the explanation for this inconsistency is complex: nuances of the stimuli and the situation can make a difference in whether music is helpful or harmful when learning foreign language vocabulary. The question of when music may be successfully employed as a learning tool remains an interesting pursuit of research, having important implications for application, such as developing best practices for foreign language learning.

References

Begg, I. & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(4), 553; 553-563; 563.

Brodsky, W., & Slor, Z. (2013). Background music as a risk factor for distraction among young-novice drivers. *Accident Analysis and Prevention, 59*, 382-393. doi:10.1016/j.aap.2013.06.022

Hughes, G. M., Rudin-Brown, C. M., & Young, K. L. (2013). A simulator study of the effects of singing on driving performance. *Accident Analysis and Prevention, 50*, 787-792. doi:10.1016/j.aap.2012.07.001

Judde, S., & Rickard, N. (2010). The effect of post-learning presentation of music on long-term word-list retention. *Neurobiology of Learning and Memory, 94*(1), 13-20. doi:10.1016/j.nlm.2010.03.002

Ludke, K. M., Ferreira, F., & Overy, K. (2013). Singing can facilitate foreign language learning. *Memory & Cognition,* doi:10.3758/s13421-013-0342-5

Schellenberg, E. G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science, 14*(6), 317-320. doi:10.1111/j.0963-7214.2005.00389.x

Tulving, E. & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352; 352-373; 373.

Wallace, W. T. (1994). Memory for music: Effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1471-1485. doi:10.1037/0278-7393.20.6.1471

Appendix A

Experiment 1 Operation Span Task

List 1:	List 2:	List 3:
Stick	Tax	Sale
(3+5) x 4 = 8 x (6-2)	5 + (7x2) = (6x3) - 1	(8-5) x 4 = 7 + (3-2)
Socks	Riddle	Tablet
9 + 7 - 4 = (8x2) - 6	(8+5) - 2 = (6x3) - 7	(9-2) + 4 = (7+8) - 3
College	Garden	Souvenir
(1+6) x 5 = (8x4) + 2	(7-3) x 9 = (6x8) - 5	(3+8) x 4 = (6x7) + 2
Tongue	Finance	Interlude
(7-4) x 6 = 3 + (9x2)	(6x7)/2 = (8x3) - 4	5 + (2x8) = (7x3) + 1
Bulldozer	Spoon	Bully
(8+5) - 9 = (7+1)/2	3 x (5+4) = (9+8) + 7	(7-6) + 9 = 5 + (8-3)
Laugh	Ear	Market
2 x (6+5) = (9x3) - 4	(6+4) + 9 = (7x3) - 2	(4+9) x 3 = (5x8) - 2
Sound	Prison	Pyramid
4 + (8-5) = (9+7)/2	2 x (5+3) = 6 + (2x5)	(8+7) x 2 = (9+5) x 3

Appendix B

Experiment 1 Hebrew Verbs List

List 1:

L'hitpalel – to pray

L'daber – to talk

L'ashen – to smoke

Lalechet – to walk

Le'echol – to eat

Larutz – to run

Livkot – to cry

L'hit'orer – to wake up

L'sachek – to play

L'hitrachetz – to wash

L'hikale'ach – to shower

L'vashel – to cook

Linso'a – to drive

Lishtot – to drink

L'hadlik – to light

Lichtov – to write

Lir'ot – to see

Lishmo'a – to hear

La'a lot – to ascend

L'hadpis – to print

List 2:

L'hitkasher – to call

Laredet – to descend

Lachgog – to celebrate

L'varech – to bless

L'ta'er – to describe

Lit'om – to taste

Lirkod – to dance

Lish'ol – to ask

L'histovev – to turn

L'nasot – to try

L'vakesh – to request

L'sayem – to finish

La'afot – to bake

Likro – to read

Lishlo'ach – to send

Lihiyot – to be

Linshok – to kiss

L'targem – to translate

L'hadbik – to glue

L'hakshiv – to listen

List 3:

Lashevet – to sit

Lipol – to fall

L'harim – to raise

L'hilachem – to fight

L'shalem – to pay

Lashir – to sing

Lakum – to stand up

Lachalom – to dream

L'nakot – to clean

Lasim – to put

Lischot – to swim

Livchor – to choose

L'hitlabesh – to get dressed

Lishtok – to be quiet

L'hit'aesh – to sneeze

L'tayel – to travel

L'ha'atik – to copy

L'tapes – to climb

L'taken – to fix

La'oof – to fly

Appendix C

Experiment 2 Hebrew-English Phrase Pairs

List 1:

Eyfo hasherutim? – Where's the bathroom?

Todah rabah. – Thank you very much.

Ani ra'ev. – I'm hungry.

Kol hakavod. – Well done.

Bo iti. – Come with me.

Ben kama atah? – How old are you?

List 3:

Kama zeh oleh? – How much does it cost?

Al tid'ag. – Don't worry.

Ani maskim. – I agree.

Ko'ev li haregel. – My leg hurts.

Ani mitzta'er. – I'm sorry.

Mah nishma? – What's new?

List 2:

Ech kor'im lach? – What's your name?

Boker tov. – Good morning.

B'hatzlacha! – Good luck!

L'hitra'ot. – See you later.

Yesh li she'elah. – I have a question.

Me'eyfo atah? – Where are you from?

List 4:

Mah hasha'ah? – What time is it?

Ani yachol la'azor. – I can help.

B'te'avon! – Bon appetit!

Na'im me'od. – Nice to meet you.

Gam ani. – Me too.

Mah shlomcha? – How are you?

Appendix D

Experiment 2 Spanish Phrases for Practice Study Session

Necesito ducharme – I need to shower

Tengo que irme – I have to go

Estoy cansado – I’m tired

Soy un estudiante – I am a student

Appendix E

Experiment 3 Japanese-English Phrase Pairs

List 1: Speaking Condition

Densha no eki wa doko? – Where is the railway station?

Shiranai – I don't know

Tasukerareru – I can help you

[Ann] Desu – I am [Ann]

Doko – Where is it

Byouki da – I'm sick

Mou ichidou – What was that again?

List 2: Singing Condition

Kippu wo doko de kaeru ka? – Where can I buy tickets?

Naruhodo – I see.

Ja Ne – Bye

Isha wo yonde – Please call the doctor

Douzo – Welcome

Taihen – Help!

Ikura –How much?

List 3: Composing Condition

Eigo wo hanasu – Do you speak English?

Konban Wa – Good evening

Wakaranai – I don’t understand

Mizu wo kudasai – I want water

Tanoshinde ne! – Have fun!

Doumo – Thank you very much

Genki? – How are you?

Appendix F

Experiment 3 Spanish Phrases for Practice Study Session

Necesito ducharme – I need to shower

Tengo que irme – I have to go

Estoy cansado – I'm tired

Soy un estudiante – I am a student

Hace frío – It's cold

Quiero comer – I want to eat

Appendix G

Experiment 3 Instructional Examples of How to Compose Melodies

If you have written music before or feel comfortable with your own ability to compose, please feel free to compose melodies as you normally would for each of these words. If you have never written music before, the following tutorial will give you some strategies for how to do it. Please listen carefully now to the tutorial.

Let's start with the first practice example.

Sorry = Sumimasen

[Sorry] [1s pause] [Sumimasen]

Experimenter: 'Sumimasen.' This word has five syllables: su, mi, ma, se, n. To compose a melody for this word, we need to have one note for each of the five syllables. So my melody will have five notes. Before I decide what those notes will be, I am going to choose one main note. This note will be the most important note in the melody.

For my melody, the main note will sound like this: [sing/hum]. Now that I have my main note, I will write the five notes for 'sumimasen.'

Let me think for a minute. Okay, here's my melody: 'Sorry. Su mi ma se n.' The first and last syllable are the main note. The syllables in between are other notes.

This melody is just fine. Just to give an extra example, I'll compose another one, using the same word, 'sumimasen.' This time, I'll use the meaning of the word "sumimasen" to write my melody. Since 'sumimasen' means "I'm sorry," my new melody will sound a little bit more sad.

Let me think for a minute. Okay, here's my melody: 'Sorry. Su mi ma se n.' Each of the five notes, except the starting note, was lower than the last one. I did that to make it sound a little bit sad, since 'sumimasen' means I'm sorry.

Thanks for listening to this tutorial. As you go through the word repetitions, you can compose melodies in any way you want to. You can use the strategies from the tutorial, using the same number of notes as syllables, using main notes, or using the meaning of the word to inspire a melody. Or you can compose melodies however you choose. Whatever you decide to do, please sing your melody aloud after you hear the second repetition of the Japanese word, and please do not use the same melody for all words. You will have 10 seconds to compose and sing your melody.

Let's give it a try.

[Sorry] [1s pause] [Sumimasen] [10s pause]

[Sorry] [1s pause] [Sumimasen] [10s pause]

[words on screen: Compose and sing your own melody for 'sumimasen'!]

Thank you. This practice example is similar to the rest of the test. Let's begin.

Tables

Condition	List 1	List 2	List 3	*T1* total	*T2* total
No Music	12 (4.77)	13.75 (4.54)	13.42 (6.49)	39.17 (13.59)	10.17 (13.72)
Instrumental Music	12.4 (5.60)	16 (4.24)	15.9 (4.61)	44.3 (13.73)	13.3 (20.47)
Vocal Music	13.8 (5.31)	15.4 (4.99)	16.3 (4.83)	45.5 (13.77)	8.1 (18.31)

Table 1. *Mean recall scores on each test by study condition.*

Study Condition	Language recalled at testing	Number of participants who recalled at least 1 syllable at *T1*	Number of participants who did not recall any syllables at *T1*	Number of participants who recalled at least 1 syllable at *T2*	Number of participants who did not recall any syllables at *T2*
Neither Sung	English	17	14	14	17
Neither Sung	Hebrew	15	16	12	19
Both Sung	English	16	15	12	19
Both Sung	Hebrew	13	18	9	22
Hebrew Sung	English	16	15	16	15
Hebrew Sung	Hebrew	14	17	12	19
English Sung	English	12	19	13	18
English Sung	Hebrew	12	19	10	21

Table 2. *Cell counts for recall at each time interval by study and test conditions.*

Study Condition	Number of participants who recalled at least 1 syllable	Number of participants who did not recall any syllables
Speaking	33	9
Singing	24	18
Composing	23	19

Table 3. *Cell counts for recall at T1.*

Study Condition	Number of participants who recalled at least 1 syllable	Number of participants who did not recall any syllables
Speaking	31	11
Singing	25	17
Composing	22	20

Table 4. *Cell counts for recall at T2.*

Study Condition	Number of participants who recalled at least 1 syllable	Number of participants who did not recall any syllables
Speaking	21	2
Singing	7	16
Composing	10	13

Table 5. *Cell counts for recall at T3.*

Retention Interval	Number of participants who recalled at least 1 syllable	Number of participants who did not recall any syllables
T1	24	18
T2	25	17
T3	7	16

Table 6. *Cell counts for recall of phrases studied in the singing condition*

Retention Interval	Number of participants who recalled at least 1 syllable	Number of participants who did not recall any syllables
T1	12	10
T2	13	9
T3	4	7

Table 7. *Cell counts for music experts' recall in the singing condition.*

Retention Interval	Number of participants who recalled at least 1 syllable	Number of participants who did not recall any syllables
T1	15	6
T2	13	8
T3	5	7

Table 8. *Cell counts for language experts' recall in the composing condition.*