

Spaced learning and melodic memory

Psychology of Music

1–19

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/03057356251401906

journals.sagepub.com/home/pomJoel J Katz^{1,2}  and Melody Wiseheart^{1,2} 

Abstract

Distributed learning is a powerful tool for optimizing retention of verbal materials. We examined the effect of distributing learning on long-term memory for a melody and found strong evidence of better recall in the spaced conditions. In the current study, music students were taught a four-phrase melody in learning sessions that were massed, spaced at 2 days, or spaced at 1 week. Three weeks later, they were tested for recall. Performances were evaluated for note omissions, number of incorrect notes and intervals, and number of correct notes and intervals. Results indicated strong evidence for a spacing effect for melody learning between the massed and spaced conditions at a retention interval of 3 weeks, and no evidence of difference between the two spaced conditions. Unlike most spacing studies, memory did not improve with longer spacing between learning episodes. These results suggest that memory for a melody may rely primarily on structural constraints within the material itself. Once these constraints are understood and associated with the cue, the performance unspools. Results have implications for best practices in melodic learning and for the role of constraining cues in the retrieval of structured non-verbal material.

Keywords

spacing effect, distributed practice, learning, long-term memory, melody learning

Melody has been part of human life since prehistory. It accompanies us from our very earliest moments. In every culture, mothers soothe infants with a lullaby. Newborn babies may recognize tunes heard while still in the womb (Partanen et al., 2013). We learn to speak through “motherese,” a wordless crooning by the parent that transmits the prosody and affective significance of language to the child (Saint-Georges et al., 2013). Given the ubiquity of melody in human activity, it is surprising to find that the best practice for learning and retaining a new melody has not yet been established. Our current study is the first to look at the effect of distributed learning on sung melody and one of the few to examine spacing effects in music. We join

¹Department of Psychology, York University, Toronto, ON, Canada

²LaMarsh Centre for Child and Youth Research, York University, Toronto, ON, Canada

Corresponding author:

Joel J Katz, Department of Psychology, York University, 4700 Keele Street, Toronto, ON M3J1P3, Canada.

Email: joel.j.katz@gmail.com

a small body of music studies where the stimulus materials are reproduced (i.e., sung) and not simply recognized (Halpern & Bartlett, 2010).

The central theme of this research is that melodic syntax, or the rules necessary to generate melody, creates structure that can be used to cue memory in performance. As such, it is important to understand the primary constituents of musical structure. Both tonal structure and melodic contour influence the perception of melody and its subsequent recall (Dowling, 1978; Schmuckler, 2014). Musical tonality is “the hierarchical organization of the chromatic set around a single reference pitch” (Schmuckler, 2014, p. 144). Simply put, it allows for the organization of melody according to the notes and chords of a specific key. In Western tonal music, melodies imply an underlying harmony (Patel, 2010). When an implied harmony is based on the simplest tonal formulas, transposed melodies are recognized more accurately (Cuddy et al., 1981). Scale and contour both influence melody retention (Dowling, 1978). Melodies that outline tonic triads (a chord formed of the first, third, and fifth notes of the scale) at phrase endings are remembered more accurately than melodies that do not (Boltz, 1991). The effect is shown only when the triads are used with regular patterns of accent. Modal melodies (melodies based on the notes of a scale, arranged into patterns other than major or minor) can be difficult to learn compared with melodies based on more conventional harmonic progressions (Oura & Hatano, 1988). Memory for traditional songs or ballads may be influenced by multiple constraints in the materials (Rubin, 1995, 2006; Wallace & Rubin, 1991). Recall is serial; what is sung cues what is to come. Although memory for a tune will not offer the poetic constraints offered by a song with words, melodic structure, with its regular phrase length, cadential formulas, and tonal resolution will still have a constraining effect on the choice of notes. Taken together, these constraints act as ongoing associative cues which help to determine the subsequent choice of notes.

There are several factors in melodic recall that distinguish it from word or prose recall. There is evidence that both melody learning and text learning require syntactical processing, which overlaps in the neural resources required for activation and integration (Patel, 2010). Melody learning may evoke semantic processing through activation of extramusical meaning (Koelsch, 2011), although this activation is weaker and of a much more general nature than in the learning and retrieval of text (Slevc & Patel, 2011). Memory for melody requires continuous recall according to an imposed rhythmic and melodic pattern. Furthermore, in sung recall, the procedural aspects of singing technique and the demands of music processing are engaged, and the rhythmic and pitch constraints of the melody function as a framework for retrieval (Rubin, 1995, 2006; Wallace & Rubin, 1991). This framework is presented complete at the first learning episode. Once the intervallic and rhythmic structure of the melody is learned it may be retrieved as a continuous associative unspooling where each pitch cues the next according to the learned patterns (Chaffin et al., 2015).

Singing a melody from memory involves retrieval of pitch and temporal information from long-term memory. This information, stored as episodic traces, is then sent to motor planning and implementation mechanisms that execute the articulatory and phonatory instructions (Dalla Bella, 2015; Dalla Bella & Berkowska, 2009). These traces represent exposure to the melody during the learning phase (Glenberg, 1979). Whatever cues are available at testing provide access to those traces. The cue activates components in the episodic trace that are identical to those in the cue. In more complex materials, retrieval is aided by implicit perceptual representation (Mandler, 2004; Snyder, 2016). This refers to unconscious statistical learning that structures unconscious expectation about environmental events and keeps a record of regularities in the environment (Reber, 1989). There is evidence that complex harmonic, melodic and rhythmic features may be acquired by both expert and non-expert musicians through exposure alone, without additional learning strategies (Rohrmeier & Rebuschat, 2012). Declarative cues, termed performance cues, have been described in certain types of expert music performance (Chaffin & Imreh, 2002;

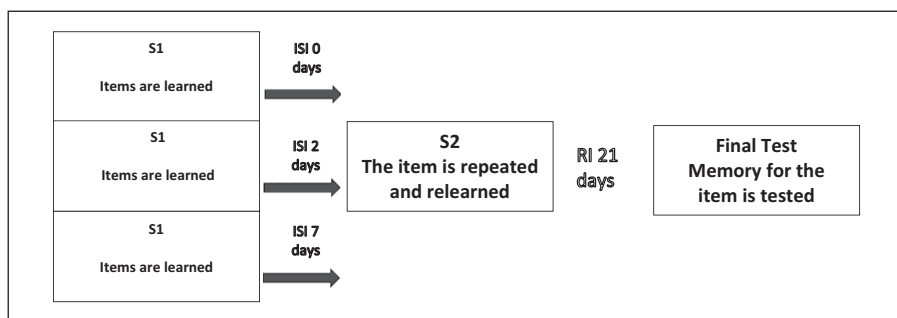


Figure 1. A Basic Distributed Practice Design.

Note. During the first study event, new material is introduced and learned. During the second study event, the material is relearned. During the test event, memory for the material is assessed. The time interval that passes between the first and second study events is referred to as the interstudy interval (ISI). The time interval that passes between the second study event and the test event is referred to as the retention interval (RI). Some researchers refer to interstudy interval as gap or lag.

Ginsborg & Chaffin, 2011). Their use is primarily limited to materials that require declarative cues internalized during the learning phase to facilitate retrieval (Bach Italian Concerto, Presto; Stravinsky Ricercar, etc.). In these performances, content addressability (the ability to restart at any point in the material) is of the utmost importance to ensure memory security. Serial cuing, whereby an ongoing chain of association allows for the unspooling of the material, is not sufficient to assure smooth performance (Chaffin et al., 2015).

The spacing effect

The spacing effect refers to the boost in recall gained from distributing learning over minutes, hours, days, or months, compared with learning materials presented in one massed session (Cepeda et al., 2008, 2009; Donovan & Radosevich, 1999; Kim et al., 2019). There is considerable research looking at the effects of distributed learning on memory for word lists (Cepeda et al., 2006) and there is evidence for the positive effects of varying both the intervals between study sessions and retention intervals for memory of relatively simple cognitive tasks (Carpenter et al., 2012). In addition to the purely verbal domain, spacing effects have been shown in a variety of other modalities (Lee & Genovese, 1988), including surgical training (Spruit et al., 2015), sports (Dail & Christina, 2004), mathematics and statistics (Ebersbach & Nazari, 2020; Lyle et al., 2020), visual acuity (Kornmeier et al., 2014), and singing (Katz et al., 2021). In a typical spacing study (Figure 1), participants are presented with an initial learning session, where the material is learned to some criterion of accuracy (often 95% correct) to ensure an equal standard of learning across conditions. A review session follows an interstudy interval (ISI) with a fixed number of learning trials. The final test is at a fixed retention interval (RI), which allows the interstudy interval to be examined as a single independent variable. For convenience, the terms spacing effect and distributed practice effect are often used interchangeably.

Music spacing studies

Spacing effect studies using musical materials are rare. Several studies have looked at the effect of varying retention interval on memory consolidation for simple keyboard sequences (Allen, 2007; Cash, 2009; Simmons & Duke, 2006). These all used massed sessions for training.

Simmons (2011) looked at memory for a nine-note keyboard figure among trained keyboard players. There were three conditions, massed and spaced at 6 and 12 hr, but the retention interval varied between experimental conditions, thereby confounding the results. As such, no conclusions regarding the effect of spacing on memory for music can be drawn. The study was also underpowered (29 participants in three conditions).

Song, with its unique cognitive profile of coordinated production of both notes and words, shows a large spacing effect in sung performance (Katz et al., 2021). Three groups of undergraduates were taught to sing a two-verse song set to a newly written melody following traditional rules of musical syntax. Lyrics followed the repetitive rhyme and metre of the traditional ballad (Wallace & Rubin, 1991). After a 3-week retention interval there was evidence for a spacing effect for syllables for the 2-day and 1-week condition compared with the massed, with evidence of no difference between the two spaced conditions. Results were consistent with the theory of multiple constraints in complex materials. The authors theorized that once the complex structure of the song was understood and associated with the initial cue, the predictable internal structure would allow for retrieval without resorting to the usual form of contextual cuing that we would expect to find in verbal materials (Cepeda et al., 2008; Glenberg, 1979). Because the notes and words were taught together, any measure of the retrieved notes would reflect word learning as well. The effect of distributed learning on melody alone thus cannot be determined from the song study. For this question, a separate melody study is required.

Spacing effect: theoretical context

The spacing effect has been explained as an interaction between encoding variability and study-phase retrieval. Most other theories (e.g., deficient processing) do not fit well with spacing gaps of days or weeks, as used in the present study (Cepeda et al., 2008, 2009). Encoding variability refers to the variability in cues available at retrieval when materials are learned over time. There are three types of cues considered to be significant in recall: contextual, structural-associative, and descriptive (Glenberg, 1979). Of these, contextual cues function largely unconsciously, based on the context of the learning environment, and are automatically encoded. Structural cues are based on the structural relationships discovered or imposed on the items to be remembered during the learning phase. They are more specific than contextual cues which are typically general in nature. Finally, descriptive cues are meanings consciously attached to the material to be remembered and are the most specific in nature. More time between subsequent presentations of the same material increases contextual variability, which in turn leads to enhanced retrieval due to the greater possibility of overlap between learning and retrieval context. A greater number of unconsciously recognized contextual cues when present at testing increase the odds of retrieving the previously learned material. Study-phase retrieval contributes to the effect by strengthening memory for items that are successfully retrieved with some effort during the learning phase (Thios & D'Agostino, 1976). Items that are easily retrieved and those that are not retrieved receive no such boost. Taken together, these factors are thought to produce the spacing effect (Delaney et al., 2010).

Current study

In this study we examine the distribution of learning events for long-term retention of a sung melody. Our theoretical perspective draws on the theory of multiple constraints as a cuing mechanism derived from the study of traditional ballads (Rubin, 1995) and encoding variability from the spacing literature. In the case of a melody, both contextual components unconsciously absorbed in the

learning phase and structural components created by the musical framework may be activated when the first notes are heard. Access to the memory is predominantly controlled by the most specific components in the trace (Watkins & Watkins, 1975). As structural components are more specific than contextual components, retrieval of the melody will be predominantly controlled by whatever structural components are implied by the initial cue and the ongoing structural cues generated by the performance as it unfolds. The various structural aspects of melody (tonality, motivic patterns, rhythm, etc.) act as constraints that limit the choice of possible notes in reconstruction of the melody from memory. Globally, coherent rhythmic patterns supported by cadential structures in balanced phrases allow the notes to be chunked for recall. With massed learning, study-phase retrieval will rely on the easy retrieval of information from the first occurrence. As the gap between learning sessions increases, a reconstructive process (Glenberg, 1979) takes hold that is closer to the demands of the final memory test. This strengthens the memory trace and in verbal materials at least, makes retrieval of the material more likely for longer spacing intervals at the final test. In the present tune study, we predict that structural cues at the final test will be enough to overtake the effect of contextual variability at the more distant spacing interval. Although the absence of words structured in a poetic form may diminish the effect of structural cuing, it will not eliminate it. For this reason, contrary to the usual findings in the verbal spacing literature, where increased spacing usually leads to better memory at the final test, structural cuing will lead to an equal enhancement to memory between spaced intervals at the final test.

Hypotheses

In this study, we tested memory for a melody under three learning conditions: massed learning, where there is a minimal gap between learning sessions; spaced learning, where there is a 2-day gap; and spaced learning, where there is a 1-week gap. Our dependent variables for all hypotheses were correctly and incorrectly sung notes and intervals.

- Hypothesis 1: A spacing effect between massed and spaced sessions at final retrieval. Because of the easy retrieval of the melody at the review session, and the reliance on working memory instead of long-term memory for retrieval, structural elements in the melody will be only weakly associated with the cue in the massed condition. This will lead to impaired retrieval at the final session, compared with the spaced conditions.
- Hypothesis 2: No difference in memory between the two spaced conditions. Either of the two spaced conditions will allow for sufficient association between structural elements in the melody and the initial cue to enable retrieval of the melody. In this study, structural cues will be enough to overtake the effect of contextual variability at the more distant spacing intervals. Although the absence of words structured in a poetic form may diminish the effect of structural cuing, it will not eliminate it.
- Hypothesis 3: Increased forgetting for the spaced conditions at the second session, when the massed and spaced conditions are compared. There are fewer shared contextual cues between initial learning and the spaced conditions. This makes retrieval more difficult for the spaced conditions.
- Hypothesis 4: A difference in forgetting between the two spaced conditions, with increased forgetting for the longer gap condition. Because structural cues in the melody are only weakly associated with the stimulus material, prior to the review session, contextual variability will largely determine retrieval. At the beginning of the second learning session, where forgetting is measured, the only contextual cues associated with the target material are from the first learning session. The longer interstudy intervals in the spaced

conditions mean an increasing mismatch between the contextual cues learned at the first session and the contextual cues available at retrieval the beginning of the second session, hence diminished retrieval.

Method

Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data are available at <https://osf.io/f4hjp/>. Data were analyzed using JASP, version 0.18.3 (JASP Team, 2024), with default priors (Wagenmakers et al., 2018; for *t*-tests, a Cauchy distribution with a location of 0 and scale of 0.707; for ANOVAs, Jeffreys-Zellner-Siow prior, *r* scale fixed effects = .5; *r* scale random effects = 1). We chose to use default priors because they “have desirable properties, are broadly applicable, and are computationally convenient” (Rouder & Morey, 2012, p. 883). This study was approved by the York University Human Participants Research Committee, e2020-291.

Participants

When we planned our study, few or no accessible methods for a priori sample size estimation for Bayesian analyses existed (e.g., Fu et al., 2022). Thus, we aimed to run enough participants to ensure that all our major hypotheses had at least moderate evidence for the null or experimental hypothesis (i.e., $0.33 > BF_{10} > 3$). Our prior study on song learning (Katz et al., 2021) suggested that a sample size of $n = 90$ would be sufficient. We planned to add additional participants after the initial round of data collection, if we were unable to find at least moderate evidence for or against our hypotheses. This is known as an open-ended sequential design (Schönbrodt & Wagenmakers, 2018). The experiment compared participants ($n = 90$) drawn from conservatory music students and undergraduate university students of varying levels of musical expertise. A total of 143 participants began the experiment: undergraduates from a second-year music class ($n = 97$) who received course credit for participation; members of a first-year ear training class in a BA program at a music conservatory ($n = 11$) who also received course credit for participation; undergraduates from the general university population ($n = 25$) who were paid participants; and candidates from a pool of exam candidates in a standardized North American music curriculum ($n = 10$) who were paid participants. The criterion for learning was a correct performance of the notes as determined by testers at the end of the first session. This high standard was chosen as partial compensation for the inexact nature of pitch reproduction in singing. Of the 143 participants who completed the study, 91 reached the initial learning criterion and 52 did not. In addition, sound files from one of the criterion candidates were too distorted to be coded, leaving a final total of 90 participants for analysis. The high exclusion rate (36%) may have been due to the necessarily high criterion standard.

A range of demographics were collected, which included age, gender, guardian education, results of the Profile of Music Perception Skills (PROMS) test of musical perceptual ability, Royal Conservatory of Music (RCM) grade level, and level of music performance experience. Participants were on average 20 years old ($SD = 2.2$, range = 16–30), 47 were female, and 43 were male, and all were native English speakers or had spoken English for 5 years or more. Participants' guardians had 16 years of education, on average. Participants were Caucasian ($n = 41$), East Asian ($n = 24$), South Asian ($n = 10$), Black ($n = 4$), Hispanic ($n = 4$), or Other ($n = 6$). Bayesian ANOVA or Bayesian contingency table tests (as appropriate) indicated no

evidence of difference between groups for individual demographics (Appendix 1, Table 3). Participants were matched between Zoom-based internet sessions and lab-based in person sessions, so that there were approximately equal ratios of participants tested on Zoom and in person for each experimental condition. A Bayesian contingency test for location indicated evidence of no difference between conditions ($BF_{10} = 0.10$). Pitch performance in online singing tests has been shown to function comparably to in-person singing tests (Tan et al., 2021). The mixed model of Zoom plus in-person testing was selected for practical considerations. In-person testing was not allowed in 2022, when testing began, due to the COVID-19 pandemic. By 2023, classes were taught in-person and we moved to in-person testing to accommodate.

Design

Participants were randomly assigned to one of the three conditions: massed, spaced at 2 days, or spaced at 1 week. They were then trained to criterion at the initial study session, followed by a review session that was either immediate (massed), or 2 days or 1 week later (spaced conditions). The final test was 3 weeks after the review session (Figure 1). All tests were cued by the first four notes of the melody, which were then dropped from analysis, leaving 48 individual notes. Measures of note accuracy (number of correctly sung notes; number of incorrectly sung notes), interval accuracy (number of correct intervals; number of incorrect intervals), and number of note omissions were the dependent variables for analysis for all hypotheses. Cents off pitch and absolute value of cents off pitch for both notes and intervals were tracked, but as a broad measure that included all notes (both correct and incorrect), they were too general to be of use. Hesitations were also recorded, although there were too few hesitations to allow for further analysis.

Materials

A four-phrase melody was newly composed by the lead researcher for testing purposes (Appendix 2). To avoid floor or ceiling effects, the melody was composed to be easy enough to learn in a single session for most participants, but difficult enough to show differences in recall after 3 weeks. Efforts were made to incorporate features in the melody that make for memorability, including a simple underlying tonal harmony (Cuddy et al., 1981; Patel, 2010); frequent use of pitches outlining basic triads at phrase endings within a regular pattern of accents (Boltz, 1991); and a completely regular phrase structure derived from traditional ballads (Rubin, 1995). Materials contained no modulations to other keys and only a brief excursion to the relative minor in the third phrase (Oura & Hatano, 1988). Stimulus materials were prepared as scores and MIDI recordings in four different keys to match the voice types of the participants. Scores were prepared using Noteflight, an online music transcription software, in the keys of F for soprano and D for alto in the treble clef, F for tenor in the tenor treble clef, and D for bass in the bass clef. In all cases, scores were notated with the syllable “doo” on each note to simplify subsequent pitch identification. MIDI recordings of the melody were generated in Noteflight in the various keys at a constant tempo of quarter note = 100 beats per second, a tempo chosen to be engaging and comfortably singable. Four different PowerPoint presentations were then prepared using the stimulus recordings and scores, according to a uniform presentation script. The script was then recorded by the lead researcher and embedded in the slides for use as instructional materials.

Procedure

The participants were all tested for voice type on presentation in the lab or Zoom session. If they were unable to match the reference keyboard pitch, they were given course credit but disqualified from further participation. Data from participants who could not recognizably sing the melody (according to the judgment of the researcher) was eliminated from analysis. Participants were first exposed to the entire melody played in the appropriate key for their voice type. Every stimulus exposure was with recording and score in the appropriate keys. They were then given three groups of learning trials: the first two phrases, the last two phrases, and all four phrases together. In the first two groups of trials, two phrases of the four-phrase stimulus were presented three times followed by a memory test cued by the starting pitch. At this point, if the participant reached criterion on the first two phrases, they were advanced to the final two phrases. If not, they continued with a maximum of three more presentations of the first two phrases alternating with cued tests. Once four tests had been given on the first two phrases, participants were advanced to the final two phrases, which were presented in the same way: three learning trials followed by a test, and then a maximum of three presentations of the stimulus, each followed by a cued test. The final group of learning trials began with a presentation of the whole melody, followed by a test cued by the first four notes of the melody. This was then followed by a maximum of three more presentations of the whole melody, each followed by a cued test. At any point in this final group of trials, if the participant reached criterion on a test, they were passed to the end of the session. Mistakes were pointed out by the researcher after every test, so that, if possible, participants would not repeat the same errors. At the end of the session, participants in the spaced conditions were thanked for their time and asked not to practice or otherwise think about the melody. Those in the massed condition went on to the second session immediately. Post hoc testing determined both equality and accuracy between all criterion tests (see Appendix 1, Table 4).

The second (review) session had a fixed number of learning trials across all conditions. The session began with a memory test to establish forgetting. All participants were then given three exposures to the stimulus and another test. Dates for the final session were confirmed, and participants were once again asked not to practice or otherwise think about the melody. The third session consisted only of the final test. All participants were directed to an online demographic questionnaire and the mini-PROMS test (Profile of Music Perception Skills; Zentner & Strauss, 2017), which they were asked to complete on their own outside the testing lab. The mini-PROMS is a well-validated 15-min version of the original PROMS battery of tests (test-retest reliability, $r = .83$. Criterion validity, $r = .61$). Results indicated evidence of no difference between groups for musical perceptual ability ($BF_{10} = 0.12$). Although testing in our study was conducted by three researchers who were aware of the experimental condition, efforts were made to ensure freedom from bias and to establish equivalence between participant groups other than for the experimental manipulation. Of the three sessions in the lab, sessions 2 and 3 followed a strict protocol determined by the slide presentation. The initial session necessitated individualized coaching to reach the criterion learning goal.

Data coding and analyses

We coded notes and intervals that were 50 cents or less off the ideal note or interval as on-pitch (i.e., correct). Being more than 50 cents off was coded as off-pitch (i.e., incorrect). As nearly as possible, objective methods of measurement were used to determine the experimental values for analysis. Participant recordings were masked by the head of the research team. Recorded

sound files were analyzed in Melodyne, a commercially available sound editing program. Melodyne uses Fast Fourier Transform analysis of the waveform for acoustic segmentation, separating each audio file into separate pitches. An average pitch is then determined from stable note segments, ignoring onsets and offsets, as these are the points of maximum instability in a vocal performance (Dalla Bella, 2015; Neubäcker, 2011). Pitches were reviewed and corrected for onset time by the lead researcher, who was unaware of the experimental condition. Pitch values in note names and cents and length of any hesitation in hundredths of a second were then transcribed from the Melodyne files and entered into spreadsheets.

Results were analyzed with Bayesian ANOVAs and post hoc tests for unimodal data (number of correctly and incorrectly pitched notes, number of correctly and incorrectly pitched intervals between notes, and number of omitted notes). Of all these measures, the most relevant to our hypotheses are the number of correctly pitched intervals. Number of notes omitted is a catch-all measure, in that all notes that were not sung were part of the total. This tended to diminish the difference between experimental groups by removing notes sung incorrectly. Number of correct notes was more specific for our purposes but discounted those notes that strayed more than 50 cents from the “ideal” pitch. This meant that notes were judged incorrect where participants drifted from the underlying tonality of the melody. Number of correct intervals was the best of the measures available to us, because it compensated for pitch drift. Although this dependent variable still included a number of pitches that were more than 50 cents out of melody but judged “correct” by the researchers, it did allow out-of-melody notes where the interval relations were no more than 50 cents off. We included number of incorrect notes and number of incorrect intervals in the tables of results, although those measures did not allow for the notes and intervals that were incorrect because they were omitted.

Results

Baseline

We examined the final test of the first session to establish that all participants had learned the material equally well (Figure 2; Appendix 1, Table 4). There was evidence of no difference between groups for note omissions ($BF_{10}=0.14$), number of correct pitches ($BF_{10}=0.33$), number of incorrect pitches ($BF_{10}=0.30$), number of correct intervals ($BF_{10}=0.11$), and number of incorrect intervals ($BF_{10}=0.23$).

Forgetting

We examined forgetting using the test at the beginning of session 2 (Figure 2). We expected that the massed group would show an advantage for note retrieval and that forgetting would be greater in the 1-week spaced condition over the 2-day condition. Results partially confirmed our predictions (Table 1). Participants omitted fewer notes and sang more correct notes and intervals in the massed condition than the spaced conditions, although evidence of increased forgetting between the two spaced conditions was inconclusive. Overall, differences between the massed and spaced conditions were due almost entirely to the difference between the massed and the 1-week condition.

We ran ANOVAs with spacing (massed, 2 day, and 1 week) as a factor. There was strong evidence overall for increased note omissions ($BF_{10}=26.9$), decreased number of correct notes ($BF_{10}=80.3$), and decreased number of correct intervals ($BF_{10}=34.0$) between conditions. There was inconclusive evidence of forgetting at the 2-day condition compared with the massed

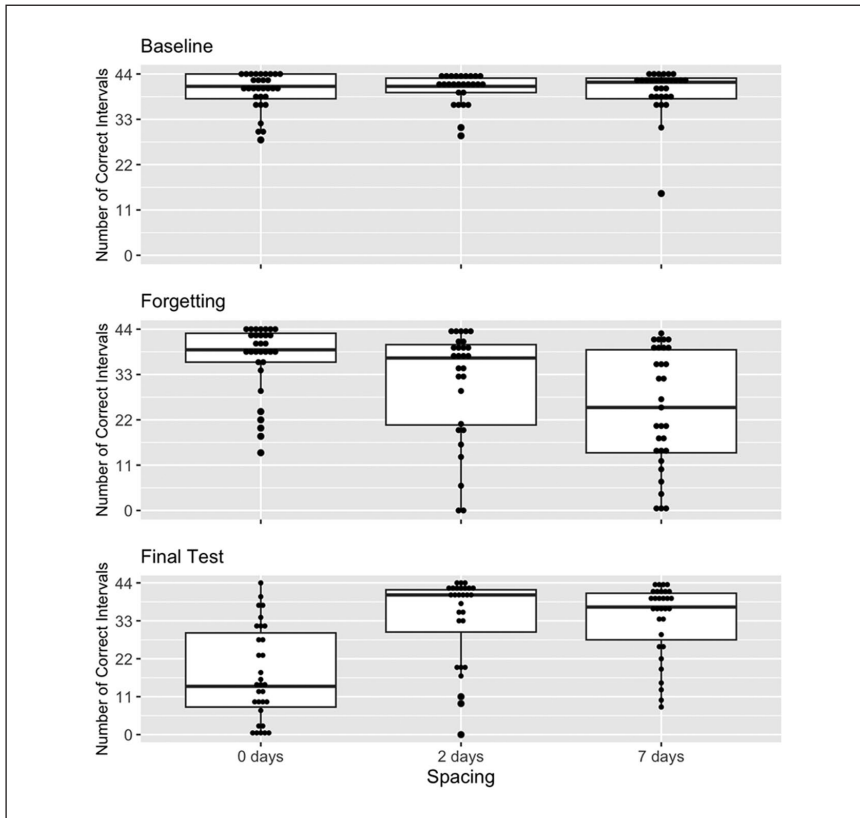


Figure 2. Number of Correct Intervals by Condition by Session.

Table 1. Measures of Forgetting at Session 2, Test 1.

Measure	Massed <i>M</i> (<i>SD</i>)	2 day <i>M</i> (<i>SD</i>)	1 week <i>M</i> (<i>SD</i>)	BF_{10} overall	Posterior odds massed vs. 2 day	Posterior odds massed vs. 1 week	Posterior odds 2 day vs. 1 week
Number of omitted notes (ANOVA)	2.1 (6.8)	6.7 (13.8)	14.3 (16.0)	26.9	0.82	104	1.3
Number of correct notes (ANOVA)	38.7 (12.8)	32.2 (16.0)	22.9 (15.9)	80.3	0.89	316	2.1
Number of incorrect notes (ANOVA)	7.2 (11.3)	7.2 (9.9)	10.7 (12.0)	0.23	0.26	0.48	0.49
Number of correct intervals (ANOVA)	36.8 (8.5)	30.9 (13.6)	24.7 (14.4)	34	1.5	163	0.89
Number of incorrect intervals (ANOVA)	5.1 (5.8)	5.0 (4.9)	5.9 (4.7)	0.12	0.26	0.3	0.32

Table 2. Spacing Effects at Session 3, Test 1.

Measure	Massed <i>M (SD)</i>	2 day <i>M (SD)</i>	1 week <i>M (SD)</i>	BF ₁₀ overall	Posterior odds massed vs. 2 day	Posterior odds massed vs. 1 week	Posterior odds 2 day vs. 1 week
Number of omitted notes (ANOVA)	19.9 (17.3)	4.6 (10.1)	5.0 (10.8)	2005	166	164	0.27
Number of correct notes (ANOVA)	18.7 (14.2)	33.6 (15.5)	30.2 (2.7)	55.6	87	11.8	0.36
Number of incorrect notes (ANOVA)	9.2 (9.3)	7.9 (10.7)	12.6 (12.5)	3.17	0.29	0.49	0.7
Number of correct intervals (ANOVA)	17.5 (13.5)	33.7 (12.6)	33.5 (10.8)	29349	1135	4624	0.26
Number of incorrect intervals (ANOVA)	7.5 (7.3)	4.3 (5.4)	5.4 (4.8)	0.58	1.18	0.54	0.36

condition (note omission, $BF_{10} = 0.82$; correct notes, $BF_{10} = 0.89$; correct intervals, $BF_{10} = 1.5$) and strong evidence of forgetting for the 1-week condition compared with the massed condition (note omission, $BF_{10} = 104$; correct notes, $BF_{10} = 316$; correct intervals, $BF_{10} = 163$). There was inconclusive evidence of a difference in forgetting between the 2-day and 1-week condition (note omission, $BF_{10} = 1.3$; correct notes, $BF_{10} = 2.1$; correct intervals, $BF_{10} = 0.89$).

Final test

Our main goal was to show the effect of three different spacing intervals on the final test performance of a melody at a 3-week retention interval (Figure 2). We predicted that there would be a retrieval advantage for both spaced conditions over the massed and that there would be no difference between the two spaced conditions. Results confirmed our predictions (Table 2). Participants recalled fewer correct notes in the massed condition than the spaced conditions, with no increase in forgetting between the two spaced conditions. We ran ANOVAs with spacing (10 min, 2 day, and 1 week) as a factor. There was strong evidence overall for increased note omissions ($BF_{10} = 2005$), decreased number of correct notes ($BF_{10} = 55.6$), and decreased number of correct intervals ($BF_{10} = 29349$) across conditions. There was strong evidence of a spacing effect at the 2-day condition compared with the massed condition (note omission, $BF_{10} = 166$; correct notes, $BF_{10} = 87$; correct intervals, $BF_{10} = 1135$), and strong evidence of a spacing effect for the 1-week condition compared with the massed condition (note omission, $BF_{10} = 164$; correct notes, $BF_{10} = 11.8$; correct intervals, $BF_{10} = 4624$). There was evidence of no difference between the 2-day and 1-week condition (note omission, $BF_{10} = 0.27$; correct notes, $BF_{10} = 0.36$; correct intervals, $BF_{10} = 0.26$).

Discussion

The findings of the present study suggest that spaced practice is an effective way to learn a melody without words. This extends the results of our previous song study (Katz et al., 2021), which also showed the positive effect of spaced learning on song. We theorized that the structural cues offered by the melody may be sufficient to allow for a comparable level of retrieval for

both spaced conditions, regardless of the gap. This was borne out by the data. We did not find in either study a benefit to extending the lag between spaced practice, a finding contrary to comparable studies of verbal learning (Cepeda et al., 2006). Comparing the two spaced conditions with the 1-week condition in the melody study, there was no recall advantage for the 1-week gap over the 2-day gap. This does not mean that contextual cues played no part in recall. The phrase-by-phrase learning approach we used certainly allowed for plenty of opportunity to form contextual cues based on the learning environment. The key issue, however, is that structural cues, when present in sufficient quantity, are more specific than contextual cues and will take precedence over them in determining retrieval (Glenberg, 1979).

Where the song study exhibited a bimodal distribution of results for forgetting and for final recall, results in the melody study were unimodal. This can be understood as a further indication that the greater number of cues in the song materials allowed some participants to retrieve the entire song, where in the melody study the notes were retrieved on a phrase-by-phrase basis. As in the song study, it appears that there will be a discernable improvement in recall once some threshold between the massed and 2-day condition has been reached. It is not unreasonable to expect on further testing, that a sleep consolidation gap may account for this (Bell et al., 2014), although the exact length of the critical gap is not yet known. Because these two studies did not have a control condition using music without structural cues, we cannot claim that we have established the relative importance of structural cues in determining retrieval in highly structured materials. We have however, presented evidence that, given the limitations of our forced learning procedure, music based on traditional materials is remembered more accurately under spaced conditions. This does suggest that structural cues can be important in determining retrieval in highly structured materials, an avenue for retrieval indicated but not tested in the previous spacing literature (Glenberg, 1979).

The results for forgetting allowed us another comparison between the presence or absence of words in the stimulus. We reasoned that a melody alone would offer fewer structural cues for recall than a melody with words. As a result, in the massed condition, common contextual cues and the recency of the material in working memory led to excellent recall for both types of materials at the beginning of the second session. In the spaced conditions, the materials were only weakly associated with the initial cue, and recency was not a factor in recall. Although we expected that the greater contextual mismatch between the initial session and the increased lag should lead to more forgetting at the 1-week interval, this was only partially borne out by the evidence. While a comparison of means shows greater recall in the 2-day condition over the 1 week, it was not enough to show evidence of difference. Retrieval of structural cues in both spaced conditions may have been strong enough to allow for similar limited retrieval. A subsequent post hoc analysis on a phrase-by-phrase basis did show evidence of a difference between the two spaced conditions for correct notes in phrase four ($BF_{10} = 8.3$). Overall, when the two spaced conditions were compared, forgetting increased in a linear fashion from phrases one to four in both note omission and number of correct notes (Tables 1 and 2 Supplemental materials).

The four-phrase melody written for our study has strong internal cues related to phrase structure, tonality, metre, and internal rhythm.¹ As in the song study, there was no increase in retrieval when the 2-day and 1-week gaps were compared. This may be characteristic of highly structured materials. Once the patterns are learned and sufficiently associated with the material, much of it can be retrieved. In both studies, the melody contains a single cue (the initial four notes), which establishes rhythmic pattern, harmony, and basic tonality. This initial cue acts to initiate a series of ongoing cues that enable retrieval.

To give a more vivid comparison, we can examine spacing effects between verbal materials, song materials, and melodic materials. For verbal materials, we have chosen experiment 1 from

Cepeda et al. (2009). In this study, a list of Swahili-English words was learned with gaps of 0, 1, 2, 4, 7, and 14 days, with a retention interval of 10 days. Although the retention intervals differed between the verbal and musical materials, the results still illustrate the difference in the spacing effect for low structure (associative memory for foreign-language vocabulary) and high structure (verbatim memory for song and melody) materials. The vocabulary study showed an increase in retention to a certain optimal gap, and a falling away in retention from there. This pattern has been characterized as a non-monotonic relationship (Cepeda et al., 2009), and it is consistent with findings in Ausubel (1966; text recall), Childers and Tomasello (2002; object recall), Edwards (1917; fact recall), and Glenberg and Lehmann (1980; free recall of word lists). In contrast, a non-monotonic relationship is absent in the spacing gaps of both the song and melody studies.

Our study suggests that spaced practice of melismatic or melodic music may offer some benefit to musicians. A singer given new cadenzas for a Handel aria on short notice, an instrumentalist learning new ornamentation, or a choral singer preparing a primarily melismatic Renaissance mass or motet will probably not retain the material well if they “cram” rehearsal into a single session. As with song, if there is an interval between learning sessions, recall will be substantially improved. The length of the critical spacing interval has yet to be determined, but we know that it must be somewhere between 10 min and 2 days, which were the approximate lengths of our massed and (shortest) spaced ISIs. It may also require an interval of sleep for consolidation (Bell et al., 2014). This research may also have more general implications for verbatim memory of all types, wherever materials are characterized by complex structural relations between elements.

Constraints on generality

We have mentioned the high criterion of 100% correct pitches from the end of the first session that may have contributed to the high exclusion rate for our study. The higher rate in the melody study (36%) over the song study (13%) may also be due to the greater difficulty of learning a melody without words. We have some support for this from the song study, where the first 20 min of learning were devoted to learning the melody alone. Over half of the participants were unable to learn the melody in this time. Once the words were given, most of the same participants were able to correct their melody errors. It is possible that the difficulty of learning a melody without words may be found more generally, a phenomenon worthy of further study. In addition, there may have been more inadvertent rehearsal between spaced sessions (including earworms) limiting the generalizability of the results. Finally, the method of learning (two phrases at a time) may have allowed for the increased formation of structural cues over a more ecologically valid approach.

Conclusion

This study showed that the spacing effect can be used to help memory for a melody based on traditional materials, when taught on a phrase-by-phrase basis. Like the song spacing study (Katz et al., 2021), but unlike most verbal learning studies, we did not find the inverse-U pattern where retention improves with greater spacing intervals and then decreases once an optimal ISI has been reached (Cepeda et al., 2008). Future studies should examine whether this result reflects our choice of inter-study intervals, or whether the large number of cues present in a traditional melody contributes to better memory from spacing regardless of ISI. In the future, studies of complex materials using different modalities (i.e., visual, gestural, aural, verbal)

would help to identify the influence of structure on retention. A more ecologically valid learning strategy could be tested. In addition, the results could be extended to different musical genres and the analysis stratified to indicate the effect of varying levels of musical expertise. A follow-up study of song lyrics without a musical setting is also clearly indicated, examining whether the spacing effect improves memory for lyric poetry.

Author contributions

M.W. acquired funding for the project; J.J.K. and M.W. conceptualized the study; J.J.K. administered the project and conducted the research; J.J.K. and M.W. supervised the research team, curated the data, and conducted data analyses; J.J.K. wrote the original paper draft and prepared data tables; J.J.K. and M.W. reviewed and edited the paper.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Social Sciences and Humanities Research Council of Canada (grant number 435-2020-1256).

Ethical considerations

This study was approved by the York University Human Participants Research Committee (e2020-291). All participants provided written informed consent prior to participating. All data are available at <https://osf.io/f4hjp/>.

ORCID iDs

Joel J Katz  <https://orcid.org/0009-0001-0287-1307>

Melody Wiseheart  <https://orcid.org/0000-0001-8332-6775>

Supplemental material

Supplemental material for this article is available online.

Note

1. We do not intend to imply a direct comparison with traditional music, which may draw on other factors to ensure memorability (Rubin, 1995).

References

- Allen, S. E. (2007). *Procedural memory consolidation in musicians* (Publication No. 3274728) [Doctoral dissertation, University of Texas at Austin]. ProQuest Dissertations & Theses Global.
- Ausubel, D. P. (1966). Early versus delayed review in meaningful learning. *Psychology in the Schools*, 3(3), 195–198. [https://doi.org/10.1002/1520-6807\(196607\)3:3<195::AID-PITS2310030302>3.0.CO;2-X](https://doi.org/10.1002/1520-6807(196607)3:3<195::AID-PITS2310030302>3.0.CO;2-X)
- Bell, M. C., Kawadri, N., Simone, P. M., & Wiseheart, M. (2014). Long-term memory, sleep, and the spacing effect. *Memory*, 22(3), 276–283. <https://doi.org/10.1080/09658211.2013.778294>
- Boltz, M. (1991). Some structural determinants of melody recall. *Memory & Cognition*, 19, 239–251. <https://doi.org/10.3758/BF03211148>
- Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, 24(3), 369–378. <https://doi.org/10.1007/s10648-012-9205-z>
- Cash, C. D. (2009). Effects of early and late rest intervals on performance and overnight consolidation of a keyboard sequence. *Journal of Research in Music Education*, 57(3), 252–266. <https://doi.org/10.1177/0022429409343470>

- Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, 56(4), 236–246. <https://doi.org/10.1027/1618-3169.56.4.236>
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354–380. <https://doi.org/10.1037/0033-2909.132.3.354>
- Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning a temporal ridge of optimal retention. *Psychological Science*, 19(11), 1095–1102. <https://doi.org/10.1111/j.1467-9280.2008.02209.x>
- Chaffin, R., Demos, A. P., & Logan, T. R. (2015). Performing from memory. In S. Hallam, I. Cross & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed., online). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198722946.013.34>
- Chaffin, R., & Imreh, G. (2002). Practicing perfection: Piano performance as expert memory. *Psychological Science*, 13(4), 342–349. <https://doi.org/10.1111/j.0956-7976.2002.00462.x>
- Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, 38(6), 967. <https://doi.org/10.1037/0012-1649.38.6.967>
- Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 7(4), 869. <https://doi.org/10.1037//0096-1523.7.4.869>
- Dail, T. K., & Christina, R. W. (2004). Distribution of practice and metacognition in learning and long-term retention of a discrete motor task. *Research Quarterly for Exercise and Sport*, 75(2), 148–155. <https://doi.org/10.1080/02701367.2004.10609146>
- Dalla Bella, S. (2015). Defining poor-pitch singing: A problem of measurement and sensitivity. *Music Perception: An Interdisciplinary Journal*, 32(3), 272–282. <https://doi.org/10.1525/mp.2015.32.3.272>
- Dalla Bella, S., & Berkowska, M. (2009). Singing and its neuronal substrates: Evidence from the general population. *Contemporary Music Review*, 28(3), 279–291. <https://doi.org/10.1080/07494460903404386>
- Delaney, P. F., Verkoeijen, P. P., & Spigel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory: Vol. 53* (pp. 63–147). Academic Press. [https://doi.org/10.1016/S0079-7421\(10\)53003-2](https://doi.org/10.1016/S0079-7421(10)53003-2)
- Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, 84(5), 795. <https://doi.org/10.1037/0021-9010.84.5.795>
- Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 85(4), 341. <https://doi.org/10.1037//0033-295x.85.4.341>
- Ebersbach, M., & Nazari, K. B. (2020). Implementing distributed practice in statistics courses: Benefits for retention and transfer. *Journal of Applied Research in Memory and Cognition*, 9(4), 532–541. <https://doi.org/10.1016/j.jarmac.2020.08.014>
- Edwards, A. S. (1917). The distribution of time in learning small amounts of material. In Colleagues and former students of Edward Bradford Titchener (Eds.), *Studies in psychology, contributed by colleagues and former students of Edward Bradford Titchener* (pp. 209–213). Louis N. Wilson. <https://doi.org/10.1037/11008-012>
- Fu, Q., Moerbeek, M., & Hoijsink, H. (2022). Sample size determination for Bayesian ANOVAs with informative hypotheses. *Frontiers in Psychology*, 13, Article 947768. <https://doi.org/10.3389/fpsyg.2022.947768>
- Ginsborg, J., & Chaffin, R. (2011). Performance cues in singing: Evidence from practice and recall. In I. Deliège & J. W. Davidson (Eds.), *Music and the mind: Essays in honour of John Sloboda* (pp. 339–360). Oxford University Press. <https://doi.org/10.1093/acprof:osobl/9780199581566.003.0017>
- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, 7(2), 95–112. <https://doi.org/10.3758/BF03197590>

- Glenberg, A. M., & Lehmann, T. S. (1980). Spacing repetitions over 1 week. *Memory & Cognition*, 8, 528–538. <https://doi.org/10.3758/BF03213772>
- Halpern, A. R., & Bartlett, J. C. (2010). Memory for melodies. In M. Reiss Jones, R. Fay & A. Popper (Eds.), *Music perception* (pp. 233–258). Springer. https://doi.org/10.1007/978-1-4419-6114-3_8
- JASP Team. (2024). JASP (Version 0.18.3) [Computer software]. <https://jasp-stats.org/previous-versions/>
- Katz, J. J., Ando, M., & Wiseheart, M. (2021). Optimizing song retention through the spacing effect. *Cognitive Research: Principles and Implications*, 6, Article 79. <https://doi.org/10.1186/s41235-021-00345-7>
- Kim, A., Wong-Kee-You, A., Wiseheart, M., & Rosenbaum, R. (2019). The spacing effect stands up to big data. *Behavior Research Methods*, 51, 1485–1497. <https://doi.org/10.3758/s13428-018-1184-7>
- Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, 8(2), 89–105. <https://doi.org/10.1016/j.plrev.2011.04.004>
- Kornmeier, J., Spitzer, M., & Sosic-Vasic, Z. (2014). Very similar spacing-effect patterns in very different learning/practice domains. *PLOS ONE*, 9(3), Article e90656. <https://doi.org/10.1371/journal.pone.0090656>
- Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport*, 59(4), 277–287. <https://doi.org/10.1080/02701367.1988.10609373>
- Lyle, K. B., Bego, C. R., Hopkins, R. F., Hieb, J. L., & Ralston, P. A. (2020). How the amount and spacing of retrieval practice affect the short-and long-term retention of mathematics knowledge. *Educational Psychology Review*, 32, 277–295. <https://doi.org/10.1007/s10648-019-09489-x>
- Mandler, J. M. (2004). Thought before language. *Trends in Cognitive Sciences*, 8(11), 508–513. <https://doi.org/10.1016/j.tics.2004.09.004>
- Neubäcker, P. (2011). *Sound-object oriented analysis and note-object oriented processing of polyphonic sound recordings*. Google Patents. <https://patents.google.com/patent/US8022286B2/en>
- Oura, Y., & Hatano, G. (1988). Memory for melodies among subjects differing in age and experience in music. *Psychology of Music*, 16(2), 91–109. <https://doi.org/10.1177/0305735688162001>
- Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences*, 110(37), 15145–15150. <https://doi.org/10.1073/pnas.1302159110>
- Patel, A. D. (2010). *Music, language, and the brain*. Oxford University Press.
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118(3), 219. <https://doi.org/10.1037/0096-3445.118.3.219>
- Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science*, 4(4), 525–553. <https://doi.org/10.1111/j.1756-8765.2012.01223>
- Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, 47(6), 877–903. <https://doi.org/10.1080/00273171.2012.734737>
- Rubin, D. C. (1995). *Memory in oral traditions: The cognitive psychology of epic, ballads, and counting-out rhymes*. Oxford University Press. <https://doi.org/10.1093/oso/9780195082111.001.0001>
- Rubin, D. C. (2006). The basic-systems model of episodic memory. *Perspectives on Psychological Science*, 1(4), 277–311. <https://doi.org/10.1111/j.1745-6916.2006.00017.x>
- Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., & Cohen, D. (2013). Motherese in interaction: At the cross-road of emotion and cognition? (A systematic review). *PLOS ONE*, 8(10), Article e78103. <https://doi.org/10.1371/journal.pone.0078103>
- Schmuckler, M. A. (2014). Components of melodic processing. In S. Hallam, I. Cross & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed., pp. 143–165). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198722946.013.14>
- Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 25(1), 128–142. <https://doi.org/10.3758/s13423-017-1230-y>

- Simmons, A. L. (2011). Distributed practice and procedural memory consolidation in musicians' skill learning. *Journal of Research in Music Education*, 59(4), 357–368. <https://doi.org/10.1177/0022429411424798>
- Simmons, A. L., & Duke, R. A. (2006). Effects of sleep on performance of a keyboard melody. *Journal of Research in Music Education*, 54(3), 257–269. <https://doi.org/10.1177/002242940605400308>
- Slevc, L. R., & Patel, A. D. (2011). Meaning in music and language: Three key differences: Comment on "Towards a neural basis of processing musical semantics" by Stefan Koelsch. *Physics of Life Reviews*, 8(2), 110–111. <https://doi.org/10.1016/j.plrev.2011.05.003>
- Snyder, B. (2016). Music and memory. In S. Hallam, I. Cross & M. Thaut (Eds.), *Oxford handbook of music psychology* (pp. 167–180). Oxford University Press. <https://doi.org/10.1093/oxfordhob/9780198722946.013.15>
- Spruit, E. N., Band, G. P., & Hamming, J. F. (2015). Increasing efficiency of surgical training: Effects of spacing practice on skill acquisition and retention in laparoscopy training. *Surgical Endoscopy*, 29, 2235–2243. <https://doi.org/10.1007/s00464-014-3931-x>
- Tan, Y. T., Peretz, I., McPherson, G. E., & Wilson, S. J. (2021). Establishing the reliability and validity of web-based singing research. *Music Perception: An Interdisciplinary Journal*, 38(4), 386–405. <https://doi.org/10.1525/mp.2021.38.4.386>
- Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, 15(5), 529–536. [https://doi.org/10.1016/0022-5371\(76\)90047-5](https://doi.org/10.1016/0022-5371(76)90047-5)
- Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35–57. <https://doi.org/10.3758/s13423-017-1343-3>
- Wallace, W. T., & Rubin, D. C. (1991). Characteristics and constraints in ballads and their effects on memory. *Discourse Processes*, 14(2), 181–202. <https://doi.org/10.1080/01638539109544781>
- Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, 1(4), 442–452. <https://doi.org/10.1037//0278-7393.1.4.442>
- Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS. *Annals of the New York Academy of Sciences*, 1400(1), 33–45. <https://doi.org/10.1111/nyas.13410>

Appendix I

Table 3. Demographic Characteristics of Experimental Groups.

Measure	Massed <i>M (SD)</i>	2 day <i>M (SD)</i>	1 week <i>M (SD)</i>	BF ₁₀ overall
Age (years)	20.6 (2.9)	20.0 (1.8)	20.1 (1.6)	0.17
Gender	14F/17M	17F/11M	16F/15M	0.27
Guardian 1 years of education	15.6 (3.5)	16.3 (3.0)	15.8 (4.1)	0.13
Guardian 2 years of education	15.8 (3.8)	16.8 (2.9)	14.9 (4.3)	0.45
Bilingual (10 = fully bilingual)	6.0 (3.2)	4.5 (3.2)	5.0 (2.9)	0.43
Hours per week of singing	4.4 (3.2)	3.9 (3.2)	3.6 (2.8)	0.15
Sheet music reading proficiency (10 = most proficient)	6.8 (2.3)	7.9 (1.7)	7.0 (2.0)	0.73
Number of songs performed from memory	52.0 (37.2)	40.4 (36.3)	55.9 (39.0)	0.29
PROMS test of music perception ability	24.2 (3.9)	25.1 (4.5)	24.5 (5.0)	0.12
Location of testing (Zoom or in-person)	8Z/23IP	10Z/18IP	9Z/22IP	0.10

Mean and *SD* are shown.

Table 4. Learning Outcomes (Session 1, Test 1): Massed, 2 Day, and 1 Week ISIs.

Measure	Massed <i>M (SD)</i>	2 day <i>M (SD)</i>	1 week <i>M (SD)</i>	BF ₁₀ overall
Number of notes omitted	0.03	0	0.03	0.14
Number of correct notes	42.6 (8.9)	44.8 (4.5)	40.7 (11.5)	0.33
Number of incorrect notes	5.3 (8.8)	2.8 (4.4)	6.4 (10.0)	0.30
Number of correct intervals	39.9 (4.6)	40.4 (3.8)	40.0 (5.6)	0.11
Number of incorrect intervals	3.9 (4.3)	2.9 (3.3)	2.7 (2.4)	0.23

Appendix 2

The image shows a musical score for a piece titled "Doo Doo". It consists of four staves of music, each with a line of lyrics underneath. The tempo is marked as $\text{♩} = 100$. The key signature has one flat (Bb) and the time signature is 2/4. The melody is simple, consisting of quarter and eighth notes. The lyrics are "Doo Doo" repeated throughout. The score is enclosed in a rectangular box.

$\text{♩} = 100$

Doo Doo Doo Doo Doo Doo Doo Doo Doo Doo Doo Doo

9

Doo Doo Doo Doo Doo Doo Doo Doo Doo Doo Doo Doo

17

Doo Doo Doo Doo Doo Doo Doo Doo Doo Doo Doo Doo

25

Doo Doo Doo Doo Doo Doo Doo Doo Doo Doo Doo Doo

Figure 3. Stimulus Materials.