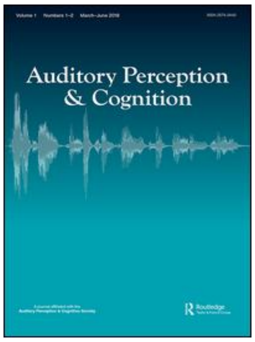

Auditory Perception & Cognition

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rpac20

The Role of Long-Term Memory in Mental Transformations of Pitch

Emma B. Greenspon , Peter Q. Pfordresher & Andrea R. Halpern

To cite this article: Emma B. Greenspon , Peter Q. Pfordresher & Andrea R. Halpern (2020): The Role of Long-Term Memory in Mental Transformations of Pitch, Auditory Perception & Cognition, DOI: 10.1080/25742442.2020.1866428

To link to this article: https://doi.org/10.1080/25742442.2020.1866428

AUDITORY PERCEPTION & COGNITION
https://doi.org/10.1080/25742442.2020.1866428

Routledge
Taylor & Francis Group

The Role of Long-Term Memory in Mental Transformations of Pitch

Emma B. Greenspon[a], Peter Q. Pfordresher[b] and Andrea R. Halpern[c]

[a]Department of Psychology, Monmouth University, West Long Branch, NJ, USA; [b]Department of Psychology, University at Buffalo, SUNY, Buffalo, NY, USA; [c]Department of Psychology, Bucknell University, Lewisburg, PA, USA

ABSTRACT

Most people can recognize and perform a musical piece under a variety of transformations such as altering the key or varying the tempo. However, we also know that other mental transformations of music can be difficult to generate and to recognize. Two factors that might affect this mental flexibility are the familiarity of the piece and musical ability of the listener, in this case singing accuracy. The current experiment addressed the accuracy and flexibility of representations of novel and traditional melodies among accurate, moderate, and inaccurate singers. Participants sang or recognized melodies in either their original form or as a transformation: a transposition, a shift of serial position, and a reversal of the melody. Participants showed an advantage for traditional melodies, but only when singing or recognizing tunes in their original form. Participants were similarly disrupted by mental transformations of traditional and novel tunes in both production and recognition tasks. Interestingly, we found that the only advantage for traditional melodies when singing repetitions of the melody occurred among the moderate singers, but all three groups showed an advantage for traditional melodies when recognizing exact repetitions.

ARTICLE HISTORY
Received 27 July 2020
Accepted 7 December 2020

KEYWORDS
Mental Transformation; familiarity; memory; music; singing; auditory Imagery

The act of conversing with a friend or singing the song "Happy Birthday to You" at a party seems effortless to most, yet these actions rely on complex memory processes. In order to engage in a conversation, you must be able to remember the beginning and end of a sentence produced by another speaker. Before you could sing "Happy Birthday to You" from memory at a birthday party, you first had to learn the song by maintaining multiple musical phrases in short-term memory (STM) while hearing the song for the first time. Through exposure, these auditory representations are eventually transferred into long-term memory (LTM). Even nonmusicians exhibit a large capacity for musical memory, as illustrated by most people's ability to recognize countless tunes on the radio (Krumhansl, 2010; Schellenberg, Iverson, & McKinnon, 1999). Despite the ubiquity of implicit musical knowledge, the cognitive processes involved in the maintenance and storage of musical information are not as well explored as memory processes for verbal and visual information (Halpern & Bartlett, 2010).

CONTACT Emma B. Greenspon egreensp@monmouth.edu Department of Psychology, Monmouth University
Supplemental data for this article can be accessed here.

In both speech and music, the sequencing of auditory units (i.e., words or pitches) is critical for interpreting the meaning of the auditory signal. In speech, changing the words in a sentence can completely alter its meaning. However, in music, the relative pitches that define intervals (relative pitch; Zatorre, Perry, Beckett, Westbury, & Evans, 1998), can be even more important than the occurrence of specific pitches, referred to as absolute pitch. For instance, most people have heard their national anthem sung by various performers across different keys and octaves. The ability to recognize the national anthem across these different instances requires a degree of flexibility in how we store pitch sequences over time and highlights the importance of relative pitch in LTM (Dowling & Bartlett, 1981).

A traditional approach to measuring memory for nonverbal information is assessing recognition memory. Dowling and Bartlett (1981) used a musical standard/comparison task with altered and preserved pitch intervals or rhythmic patterns of a familiar melody. The authors found that people were better able to identify changed relative pitches compared to changed rhythmic patterns, suggesting that representations of pitch patterns exhibit a degree of flexibility. In other words, relative pitch may act as a template for pitch sequences in memory, which may be able to be shifted either up or down in pitch space much like a stencil can be moved around on a piece of paper.

The use of LTM structures to increase working memory efficiency has been established across various domains (Charness, 1976; Takahashi, Shimizu, Saito, & Tomoyori, 2006). However, the degree to which coding of relative pitch enables representational flexibility, i.e., forming a mental template of pitch relationships, is not well understood for either pitch STM or LTM. Bartlett and Dowling (1980) addressed this question in a series of studies that presented participants with standard and comparison melodies that were either traditional children's songs or novel melodies, which relied on LTM and STM, respectively. Participants distinguished accurate from inaccurate transpositions of the standard melodies better for traditional tunes than novel tunes, suggesting that melodies stored in LTM benefit from greater working memory efficiency than melodies stored only in STM. This result aligns with the conceptualization of relative pitch functioning as a mental stencil for pitch representations in LTM. However, in these studies novel and traditional melodies were not matched for musical features, which is a critical component of the current study.

The ability to manipulate pitch relationships in working memory (or move the stencil in pitch space) like one can rotate an object in space may rely on auditory imagery. Greenspon, Pfordresher, and Halpern (2017) explored this possibility by using novel melodies. Participants performed a musical standard/comparison task in which they identified whether a comparison melody was an accurate musical transformation of a standard melody. Transformations included serial order shifts in which the melody started from the penultimate note and cycled through the melody, reverse ordering of pitches in the melody, and transpositions of the melody. Participants performed well above chance when recognizing exact repetitions, but performance was much worse in the transformation conditions. The difficulty of the transformation conditions suggests that novel auditory images demonstrate a degree of inflexibility (i.e., a crystallized representation) and may not operate like a mental template, unlike representations stored in LTM (Bartlett & Dowling, 1980; Dowling & Bartlett, 1981). Other work has found that listeners retain absolute pitch in LTM for familiar melodies, suggesting that crystallization of pitch representations, to at least some degree, is a feature of LTM as well (Halpern, 1989; Schellenberg & Trehub, 2003; Van Hedger, Heald, & Nusbaum, 2018).

However, whether access to absolute pitch representations in LTM affects one's ability to manipulate musical information in STM requires further examination.

Another way to assess memory processes is by asking participants to recall previously stored information. This is a challenge for musical sequences because these tasks rely on people's musical production abilities. The most universal form of musical production is singing, making singing the most easily available measure of musical recall abilities. However, although most people have some level of singing experience, people vary greatly in their ability to sing. Thus, in addition to measuring recognition performance for exact repetitions and transformations of novel sequences, Greenspon et al. (2017) also included a production task in which participants of varying ability to match pitch sang an exact repetition of a target melody in addition to singing transformations (serial shift, reversal, and transposition) of the melody. As was found for recognition performance, recall performance for the musical transformations indicated that novel auditory images are somewhat inflexible in that participants' performance was similarly disrupted relative to repetition performance for all three types of transformations. Interestingly, accurate singers were more disrupted by mental transformations than inaccurate singers. We interpreted this to mean that a more robust auditory image of a baseline melody had the drawback of being resistant to change.

The current study addresses whether the flexibility of an auditory image depends on one's familiarity with the melody, type of manipulation, and level of singing ability, by measuring both recall and recognition performance of musical information. A familiarization task primed LTM representations for traditional melodies. Another critical component of the familiarization task was the collection of familiarity ratings for each of the long excerpts to confirm that participants had more familiarity with traditional than novel melodies. Short melody excerpts of both types of melodies, matched for musical features, were used in the production and recognition tasks in order to have musical stimuli that were within the boundaries of STM capacity. Flexibility of auditory images was assessed by having participants sing and recognize exact repetitions and transformations of traditional and novel melodies. Transformations consisted of serial order shifts of the melody, reversed ordering of the notes in the melody, as well as transpositions of the melody (as in Greenspon et al., 2017).

One limitation of our prior work on the flexibility of auditory imagery (Greenspon et al., 2017) was that we focused only on the extreme ends of the continuum of singing ability by measuring performance for accurate and inaccurate singers while excluding moderate singers from the sample. Rather than treating pitch imitation ability as a dichotomous variable, as in our prior work (Greenspon et al., 2017), we evaluated performance in the current study across three categories of singers: accurate, moderate, and inaccurate. We included the third category of moderate singers because previous research from music education has suggested that there may be important qualitative differences across the middle and extreme ends of singing ability (Demorest, 2001; Demorest & Clements, 2007), particularly with respect to differences across production and perception tasks. In these studies, individuals falling near the middle of the continuum in production-based pitch-matching tasks tend to align more closely with accurate singers in perception-based pitch-matching tasks.

We hypothesized that having access to long-term representations would increase efficiency of imagery manipulation, resulting in better transformation performance for traditional than novel tunes in both the production (recall) and recognition tasks. We also predicted that the influence of LTM information on performance would vary by singing ability such that inaccurate singers, who are theorized to have degraded mental representations of pitch and pitch relationships and less robust auditory imagery (Pfordresher, Halpern, & Greenspon, 2015), would exhibit the largest benefit from LTM in trials with untransformed sequences (i.e., exact repetitions). In contrast, singers who have more veridical initial representations of pitch (accurate and moderate singers) would exhibit the largest benefit from LTM in trials with melody transformations. In addition, we predicted that group effects would vary by output task in line with Greenspon et al. (2017), who found larger group effects in production than recognition.

Methods

Participants

A sample (n = 83) of participants with a mean age of 19.59 years (range: 18–25 yrs.) was recruited from the Introductory Psychology subject pool at the University at Buffalo and were compensated with course credit for their participation. The ability to reproduce the correct number of notes in the production task and discriminate pitch were both preconditions of inclusion in the sample.[1]

Participants were categorized into three groups (Accurate, Moderate, and Inaccurate) based on their singing performance on an assessment task presented at the beginning of the session, which consisted of imitating eight four-note novel melodies. The a-priori classification of singing ability was determined as follows: participants who had an accuracy score of 67% or higher on the imitation trials in the assessment task were categorized as Accurate singers ($n = 35$, 13 female participants and 22 male participants), participants with a score between 33% and 66% were categorized as Moderate singers ($n = 34$, 10 female participants and 24 male participants), and participants scoring below 33% were categorized as Inaccurate singers ($n = 14$, 5 female participants and 9 male participants). Results based on different strategies for grouping participants are reported in the Supplementary Results.

Apparatus

All tasks were completed in a sound attenuated vocal booth (Whisper Room SE 2000). Stimuli were presented to the participant over Sennheiser HD 280 Pro headphones and vocalizations were recorded using a Shure WH30 headset microphone. Trials were presented using the program Matlab (MathWorks, Natick, MA) on a 3.4 GHz PC running Windows XP.

Stimuli

Assessment Task

Stimuli in the pitch imitation assessment task were four-note novel melodies. The melodies comprised prerecorded pitches sung by a male or female singer using the syllable "doo". All melodies were presented in a key that provided a good fit to the participant's vocal range, as evidenced by a single comfortable pitch participants produced at the start of assessment. Pitches within the 4-note melodies were centered around the participants' comfort pitch, which was identified as either A2, D3, F3, A3, or D4.

Familiarization Task

The familiarization task included six traditional songs and six novel melodies. The six traditional songs had been rated as being highly familiar in a pilot study. The novel melodies were constructed to share musical features with the traditional songs, which included number of notes, number of repeated pitches, and number of contour changes. In addition, traditional melodies were approximately matched to novel melodies with respect to rhythmic complexity and tempo; see Appendix A.[2]

Participants first listened to a long excerpt of a melody, which ranged between 12 and 14 notes. After the melody was presented, participants rated the melody for familiarity (1 = not at all familiar, 7 = highly familiar). They were then prompted to recall the title of the melody if they knew it and then were provided with the correct title of the melody on the computer monitor. This was done for both novel and traditional melodies. Novel melody titles were approximately matched to traditional melody titles based on the number of words, the number of repeated words, and word familiarity ratings acquired by the MRC Psycholinguistic Database (Coltheart, 1981; Wilson, 1988). For instance, "Arise, Arise, Kind Brother" was the novel match to "Twinkle, Twinkle, Little Star".

A short excerpt of the melody, which ranged from 3 to 5 notes (see Appendix A), was then presented to the participant. We adopted the procedure of presenting both the long and short excerpts in the experiment because pilot runs indicated that the long excerpts were more effective for priming LTM, whereas it is necessary to use short excerpts for imagery transformation tasks due to limitations of STM capacity. All melodies were presented in the key of C-major and consisted of prerecorded pitches that were sung by a trained male and female singer on the syllable "doo". The number of melodies that started on C (eight melodies), E (two melodies), or G (two melodies) was matched across novel and traditional melodies. Additionally, the number of melodies that ended on either A (four melodies), E (two melodies), F (two melodies), C (two melodies), or G (two melodies) were matched across familiarity type. The short excerpts were used in the production and recognition of transformation tasks.

Transformation Tasks

The experiment consisted of a vocal production task and a melody recognition task that was based on the mental transformations reported by Greenspon et al. (2017).

Production Trials. The stimuli used in the production task were the same as the short melody excerpts used in the familiarization task with the exception that the production stimuli contained a 1.7 s pause at the end of the melody followed by a cue note. The cue note was always the correct first note of the target excerpt and varied by condition. Participants were asked to sing the target melody either as an exact repetition (untransformed condition), or as a transformation of the melody (transformed condition). Transformations were shifts of serial position (i.e., rotation), reversal of the ordering of pitches in the melody (i.e., retrograde), and singing a transposition of the melody. Transformations are described in more detail in a later section.

Recognition Trials. The stimuli for the recognition task were derived from the excerpts used in the familiarization task. In the recognition task, participants heard a short melody excerpt in its original form followed by a 2.5 s pause and then either a correct or incorrect repetition, reversal, serial shift, or transposition of the melody. Incorrect melodies were never a reversal or inversion of any other sequence in the study and contained a single altered pitch, which was never the first note in the comparison melody.

Procedure

Assessment Tasks

Participants were first seated in the vocal booth and instructed on appropriate singing posture. Next, participants completed a series of vocal warm-ups in which they counted backward from 10 followed by singing the song "Happy Birthday". Participants were then asked to sing a pitch that they found comfortable to sing and this was termed the participant's comfort pitch. The comfort pitch was used to determine the key of the melodies used in the imitation trials in the assessment phase.

The imitation task consisted of eight trials, one four-note melody per trial. Following this task, participants completed an adaptive pitch discrimination task (Loui, Alsop, & Schlaug, 2009; Loui, Guenther, Mathys, & Schlaug, 2008). Participants then completed a forward and backward digit span task.

The familiarization tasks followed the assessment phase, see Figure 1.

Transformation Tasks

The production task was completed prior to the recognition task. Participants completed a practice trial before starting each of the four conditions: untransformed, reverse, serial shift, and transposition. The practice trials used a familiar melody ("Deck the Hall"), which was not used in the production nor recognition tasks. In the *untransformed* condition, participants sang the first four notes of the target melody in its original form (i.e., an exact repetition), starting on the cue note, which was the first note of the melody. For the *reverse* condition, participants repeated the notes of the melody in reverse order. In this condition, the cue note was the last note of the melody. In the *serial shift* condition, participants produced the melody starting from the penultimate note of the melody and then looped back to the beginning of the melody. For instance, if the melody comprised the pitches "A B C D" the participant would sing "C D A B".

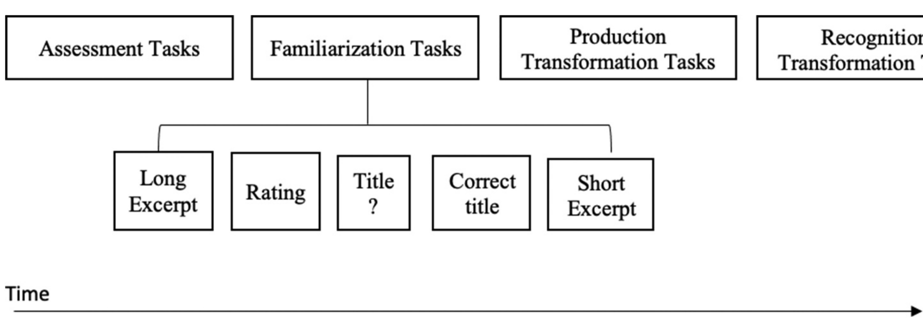

Figure 1. Order of the presentation of tasks in the experiment.

Therefore, for this condition the cue note was the penultimate note in the melody. For the *transposition* condition, the first note of the excerpt was shifted to E^b which initiated a key change to E^b for melodies starting on C (eight melodies), A^b for melodies starting on G (two melodies), or C^b for melodies starting on E (two melodies). Trials were blocked by condition such that the untransformed condition was always presented first and the three transformations were counterbalanced across participants using a Latin square design.

After the production task, participants exited the vocal booth and completed further assessment tasks which included a series of questionnaires on music and language background as well the Bucknell Auditory Imagery Scale (BAIS, Halpern, 2015), a self-report measure of auditory imagery ability. Participants then reentered the vocal booth to complete the recognition trials. As in the production task, a practice trial was presented prior to the start of each of the four conditions in the recognition task. For each trial, participants heard the target melody followed by a comparison melody. Participants then provided a yes/no response as to whether the comparison melody was a correct repetition or transformation of the first melody. Trials were blocked by condition and were presented using the same design as described for the production task.

Data Analysis

Pitch traces from the production task were extracted from the sung pitch using the YIN algorithm (De Cheveigné & Kawahara, 2002). In order to reduce the influence of pitch variability at the beginning and end of the vocalization, we looked at the median f_0 for the middle section of the sung pitch. This value was then converted to cents using C3 as the reference pitch for a male vocal range and C4 as the reference pitch for a female vocal range. The difference in cents (100 cents = 1 semitone) between the sung and target pitch was then calculated.

Performance was assessed by the proportion of correct responses in the production and recognition tasks. In the production task, a correct response (absolute pitch accuracy) was defined as a pitch deviation for a sung tone that was less than ± 50 cents from the target pitch, thus forming a range of 100-cents surrounding the target pitch (there are 100 cents in a semitone, the distance between the two closest notes of a scale). Therefore, a score of 1 represented a sung pitch that was within the categorical boundary of the target pitch according to the Western musical scale. As such, each sung pitch was scored as accurate or inaccurate dichotomously (1 = correct, 0 = error), and the proportion correct on each trial reflected the average of these scores. An estimate of chance performance was generated for each condition using a random permutation procedure based on Jacoby et al. (2019). Specifically, a set of 1,000 data sets were created by randomly permuting the association between each sung pitch and its target across all participants.

Performance on the recognition task was based on accuracy for participant judgments of "same" or "different" on each trial. Each trial was thus either scored as correct (1) or incorrect (0). Chance performance on the recognition task has an expected mean accuracy of .5.

The full factorial design for this study resulted in a considerable number of cell means for each task: 3 (participant group) × 4 (transformation condition) x 2 (familiarity condition) x 3 (sequence length) = 72 conditions. Based on previous findings from Greenspon et al. (2017) and preliminary analyses of the present data, we aggregated across some conditions to increase statistical power and simplify the presentation of results. First, because sequence length, which ranged between 3 and 5 notes, did not yield significant interactions with other factors in the production task (although length did yield a significant main effect, as may be expected), we collapsed across all levels of this factor. Second, because no reliable differences among the three transformation conditions (transposition, reversal, serial order shift) were found, see Supplementary Results, we collapsed across these conditions to yield a composite measure of performance. The design was thus reduced to a 3 (participant group) × 2 (transformation condition) x 2 (familiarity condition) design that was applied separately to analyses of production and recognition.

Results

Assessment Measures

Descriptive statistics from the pitch imitation assessment task that was used to categorize participants as well as the other assessment measures are shown in Table 1, along with pairwise *t*-tests across all three groups.

Interestingly, group differences between moderate and inaccurate singers in the assessment measures were limited to the pitch imitation task and did not generalize to measures of auditory perception, memory, or imagery measures. This was not the case for accurate singers, who demonstrated larger auditory short-term memory capacity as compared to moderate and inaccurate singers and reported experiencing more vivid auditory images than moderate singers.[3]

Familiar Melody Ratings

One participant did not follow instructions during the familiarity rating task and for this reason was dropped from this analysis.[4] Participants responded to the manipulation of familiarity as expected. Mean ratings for yoked traditional and novel melodies, as well as the significance to paired-*t*-tests on ratings, are shown in Table 2. A 3 (Accurate, Moderate, Inaccurate) × 2 (Traditional, Novel) ANOVA revealed that traditional melodies were rated as more familiar than the novel melodies in general, $F(1,79) = 1564.24$, $p < .001$, $\eta^2_p = .95$. As shown in Table 2, each traditional melody was rated as more familiar than its novel variation. Importantly, there was not a significant main effect of

Table 1. Descriptive statistics and group differences in the assessment measures.

Group (*n*)	Accurate (35) *Mean* (*SD*)	Moderate (34) *Mean* (*SD*)	Inaccurate (14) *Mean* (*SD*)	Tukey Kramer A-M	M-I	A-I
Pitch Imitation (% correct)	.86 (.09)	.50 (.11)	.23 (.07)	***	***	***
Music Experience (years)	5.67 (4.05)	3.35 (3.58)	1.82 (2.97)	*	n.s.	**
BAIS Vividness (1–7)	5.43 (0.71)	4.89 (0.98)	4.86 (1.13)	*	n.s.	n.s.
BAIS Control (1–7)	5.52 (0.74)	4.97(1.08)	4.95 (1.19)	n.s.	n.s.	n.s.
Forward Digit Span	13.03 (1.95)	10.35 (2.23)	10.21 (2.52)	***	n.s.	***
Backward Digit Span	7.63 (2.29)	6.91 (2.29)	5.64 (1.82)	n.s.	n.s.	*
Pitch Threshold (cents)	49.77 (109.24)	64.28 (90.63)	114.79 (256.16)	n.s.	n.s.	n.s.

Note. A-Accurate, M-Moderate, I–Inaccurate, *p <.05, **p <.01, ***p <.001 adjusted for multiple comparisons

Table 2. Familiarity ratings for traditional and novel tunes.

Traditional Melody Title/Novel Melody Title	Traditional *M* (*SD*)	Novel *M* (*SD*)	Difference *M* (*SD*)	*p*
Pair 1. Twinkle Twinkle Little Star/ Arise Arise Kind Brother	6.80 (0.66)	1.48 (0.95)	5.24 (1.30)	<.001
Pair 2. Old McDonald Had a Farm/ Wild Robinson Got One Boot	6.31 (1.30)	1.47 (0.93)	4.78 (1.88)	<.001
Pair 3. Yankee Doodle/Chirpy Canary	5.84 (1.47)	1.58 (0.89)	4.24 (1.69)	<.001
Pair 4. Jingle Bells/ Noble Poet	6.49 (1.08)	1.80 (1.30)	4.58 (1.94)	<.001
Pair 5. This Old Man/ Summer of Fun	5.62 (1.50)	1.84 (1.12)	3.74 (1.76)	<.001
Pair 6. Mary Had a Little Lamb/ Ah David, You Got Music	6.48 (1.20)	1.60 (0.97)	4.82 (1.56)	<.001

group or a group x familiarity interaction. Our manipulation of familiarity was thus validated by these ratings.

Production Task

Accuracy of production was analyzed using a 3 (Accurate, Moderate, Inaccurate) × 2 (Traditional, Novel) x 2 (Untransformed, Transformed) ANOVA on the proportion of correct sung notes. Means and 95% confidence intervals for each condition are shown in Figure 2, along with means and confidence intervals from a null distribution of 1,000 sample means based on randomly permuted data (see Data Analysis). All effects reported remained significant after applying a Greenhouse-Geisser correction for violations of sphericity. Post-hoc analyses using Tukey-Kramer tests indicated that the main effect of group, $F(2,80) = 23.70$, $p < .001$, $\eta^2_p = .37$, was driven by Accurate singers ($M = .48$ proportion correct) performing better than both moderate ($M = .28$) and inaccurate singers ($M = .16$), whereas moderate and inaccurate singers did not differ from one another.

Familiarity also influenced accuracy of production, $F(1,80) = 7.73$, $p < .01$, $\eta^2_p = .09$, with traditional melodies being produced more accurately on average ($M = .36$) than novel melodies ($M = .32$). As found previously (Greenspon et al., 2017), excerpts produced in their original configuration ($M = .47$) were produced more accurately than mental transformations of excerpts ($M = .21$), $F(1,80) = 167.34$, $p < .001$, $\eta^2_p = .68$. As can be seen, performance of mentally transformed pitch patterns among the moderate and inaccurate singing groups falls in the range of accuracy likely to occur by chance, thus suggesting that these groups were effectively "guessing" during their sung reproductions in these conditions.

The most critical effects involve interactions. There was a significant familiarity × transformation interaction, $F(1,80) = 7.45$, $p < .01$, $\eta^2_p = .09$. This two-way interaction validated the robustness of the transformation effect, which was significant for both novel and traditional melodies. By contrast, the interaction showed the familiarity effect was more limited, in that an advantage for traditional melodies was only observed in the untransformed condition.

The differentiating group in the current study was the moderate group; as a result, all interactions with the factor group were also significant: group × familiarity interaction, $F(2,80) = 7.45$, $p < .01$, $\eta^2_p = .13$; group × transformation interaction, $F(2,80) = 5.19$, $p < .01$, $\eta^2_p = .11$; and group × transformation x familiarity interaction, $F(2,80) = 3.68$, $p < .05$, $\eta^2_p = .08$.

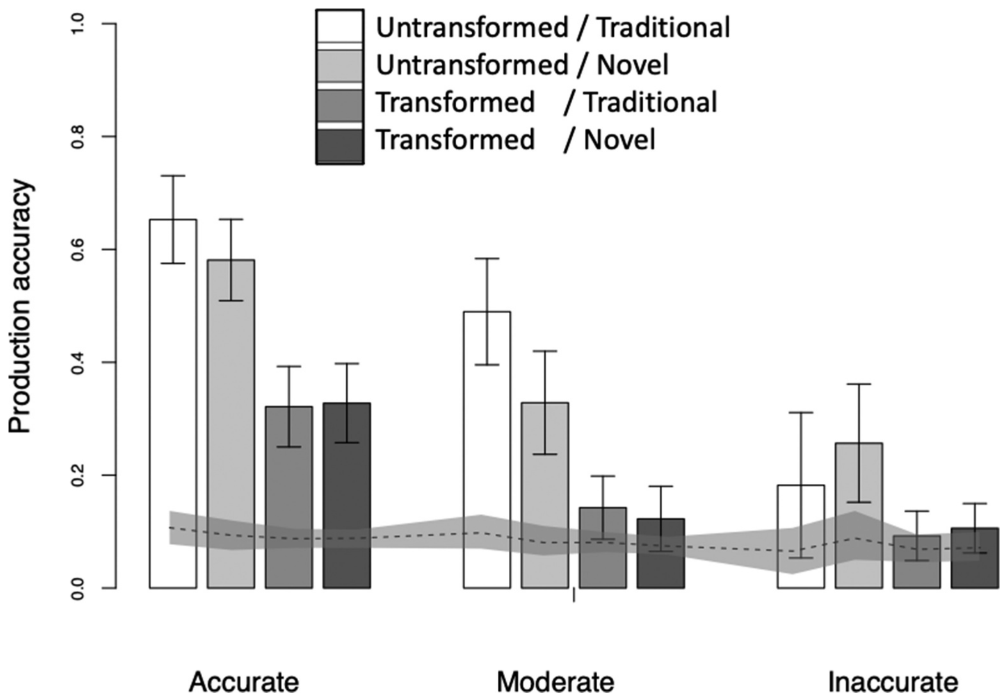

Figure 2. Mean proportion correct in the production task for each group for traditional and novel untransformed and transformed melodies. Means are plotted with 95% confidence intervals. The shaded area represents the mean (dashed line) and 95% confidence intervals from a null distribution of sample means (see Data Analysis).

Our analysis of the critical 3-way interaction was motivated by our theoretically based interest in how performance within each group is influenced by the mental transformation of novel or traditional melodies.

We analyzed the performance of each accuracy group separately using a 2 (familiarity) × 2 (transformation) within-participants ANOVA. Each ANOVA for the individual groups yielded a significant main effect of transformation: for accurate, $F(1, 34) = 86.31$, $p < .001$, $\eta^2_p = .72$, for moderate, $F(1, 33) = .78.85$, $p < .001$, $\eta^2_p = .71$, for inaccurate, $F(1, 13) = 7.00$, $p = .02$, $\eta^2_p = .35$. For the extreme accuracy groups (accurate and inaccurate) this was the only significant effect. However, the moderate group yielded an additional main effect of familiarity, $F(1,33) = 17.26$, $p < .001$, $\eta^2_p = .34$, and a familiarity × transformation interaction, $F(1,33) = 16.09$, $p < .001$, $\eta^2_p = .33$. Post-hoc pairwise comparisons on the interaction (Bonferroni corrected) yielded significant contrasts between untransformed and transformed production for each familiarity condition. However, the effect of familiarity was only significant for the production of untransformed melodies and was not significant for the production of transformed melodies. Thus, the advantage of familiarity was highly constrained in production, and borne out primarily among participants of middling accuracy while producing excerpts in their original configuration.

Analyses of production data here focused on accuracy of matching individual pitches. We also analyzed relative pitch using interval errors (differences between sung and target intervals greater than 50 cents). Those results did not yield any interactions with group and were thus considered less informative; nevertheless, means and SD are included in the Supplementary Results. One noteworthy difference in relative pitch accuracy was that interval errors were significantly lower in transposition conditions than other transformation conditions, in contrast to the results presented above.

Recognition Task

The ability to recognize pitch was related to one's ability to recall pitch information; mean percent correct across all trials for each participant was significantly and positively correlated across production and recognition tasks, $r(81) = .65$, $p < .001$. Like the production task, performance in the recognition task varied by group, familiarity, and transformation type as revealed by a 3 (Accurate, Moderate, Inaccurate) x 2 (Traditional, Novel) × 2 (Untransformed, Transformed) mixed factors ANOVA on the proportion of correct responses. There was a main effect of group, $F(2,80) = 11.79$, $p < .01$, $\eta^2_p = .23$. As found in the production task, post-hoc Tukey-Kramer tests indicated that accurate ($M = .76$) and moderate ($M = .73$) singers performed better than inaccurate singers ($M = .59$), however, unlike the production task and in line with Demorest and Clements (2007), accurate and moderate singers did not differ from one another in recognition performance, see Figure 3.

Similar to the production task, familiarity facilitated performance in the recognition task, $F(1,80) = 9.88$, $p < .01$, $\eta^2_p = .11$, with traditional melodies being recognized more accurately ($M = .75$) than novel variations ($M = .69$). In addition, recognition was more accurate for the original patterns ($M = .84$) than when based on mental transformations of excerpts ($M = .60$), $F(1,80) = 152.40$, $p < .001$, $\eta^2_p = .65$. The transformation × familiarity interaction, $F(1,80) = 3.97$, $p < .05$, $\eta^2_p = .05$, also generalized to the recognition task, though with a weaker effect size. Unlike the production data, there were no significant interactions with the factor group, although it is worth noting that performance among moderate and inaccurate groups approximates chance performance for the recognition of mentally transformed excerpts – similar to the production results.

Discussion

The current study assessed the role of familiarity in the ability to mentally transform musical representations, and how this ability may vary in perception and production tasks as well as among individuals who represent different levels of musical production skill. The current design contributes to the existing literature on long-term auditory representations by offering a comparison between previously learned auditory sequences and matched novel sequences. We highlight three main findings. First, we found that familiarity with a melody facilitates performance in both production (recall) and recognition of short tone sequences. However, this advantage was highly qualified by the other two main findings. Second, we replicated the difficulty found in producing and recognizing mental transformations of tone sequences (Greenspon et al., 2017)[5] and found further that the advantage of familiarity vanishes after mental transformation. This second finding is contrary to our hypothesis that familiarity with a melody would facilitate mental transformations. Instead, the disappearance of the familiarity effect after mental transformations suggests that traditional melodies, like novel melodies

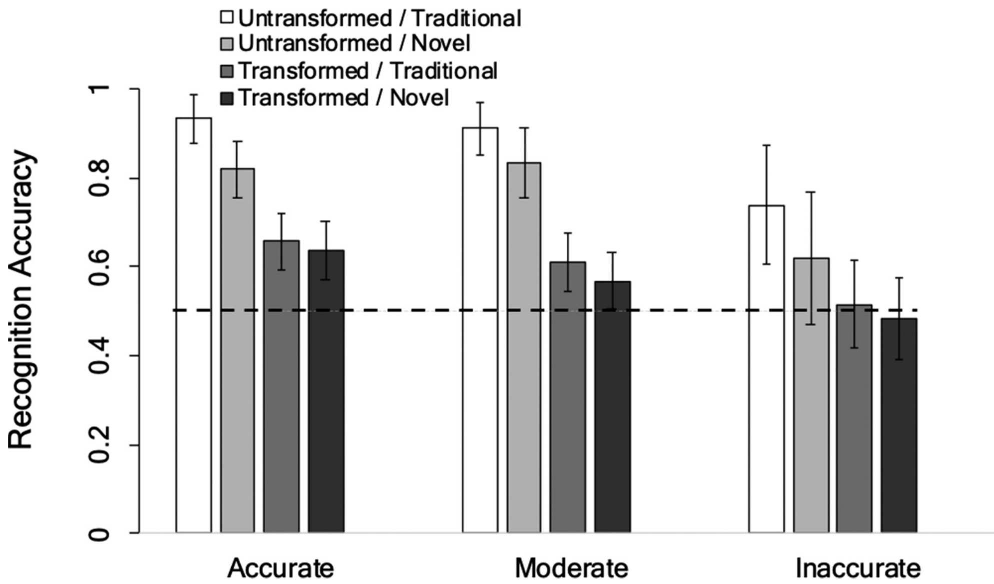

Figure 3. Mean proportion correct in the recognition task for each group for traditional and novel untransformed and transformed melodies. Means are plotted with 95% confidence intervals. Chance performance is denoted by the dashed line.

(Greenspon et al., 2017), reflect crystalized representations that are resistant to change. Third, we found a qualifying effect of singing accuracy ability on production. Although we hypothesized that inaccurate singers would exhibit the largest advantage of familiarity, instead, we observed that the advantage of familiarity appeared only for participants with an intermediate level of singing accuracy. By contrast, singing accuracy did not influence the effects of familiarity and mental transformation for recognition.

Our primary interest in this study was to address how the familiarity associated with excerpted pitch patterns from traditional melodies may influence the flexibility with which one can engage in mental transformations of the pitch pattern. This focus was based in large part on the surprising difficulty participants exhibited in producing and recognizing mental transformations of novel melodies in our previous work (Greenspon et al., 2017). In both the current production and recognition tasks, familiarity yielded an advantage that was limited to excerpts that retained their original configuration. Moreover, as in our previous research, mental transformations were difficult to execute. This was especially true of participants exhibiting low levels of singing accuracy, who performed at chance. For production, this familiarity advantage was further constrained in that it was only present for participants exhibiting intermediate levels of singing accuracy. Although it is possible that the poorest performing participants may not exhibit a familiarity advantage given their low baseline levels of accuracy, it would have been possible for the most accurate singers to exhibit a familiarity advantage, given that their highest levels of accuracy were typically under 80% correct. This finding highlights the utility of analyzing effects related to singing accuracy more broadly than in the common practice of dichotomizing participants into two groups (cf. Demorest, 2001; Demorest & Clements, 2007; Pfordresher & Larrouy-Maestri, 2015).

This study yielded differences between production and recognition tasks that offer hypotheses for future research. A general theme in previous studies of singing accuracy suggests that associations with success on auditory perception tasks are stronger for more complex perceptual tasks that may draw on similar sensorimotor representations as those that are used for singing. Thus, associations with auditory perceptual discrimination may be weak or non-significant (e.g., Greenspon & Pfordresher, 2019; Pfordresher & Brown, 2007), whereas stronger associations are found between singing accuracy and melody discrimination (Pfordresher & Nolan, 2019), measures of auditory imagery, and auditory short-term memory (Greenspon & Pfordresher, 2019; Pfordresher & Halpern, 2013). In both the current study and our previous study of mental transformations for novel melodies we found a relationship between production and recognition (Greenspon et al., 2017). In the present study, we were able to qualify this relationship through the use of three groups, which yielded subtle but compelling perception/production dissociations. First, whereas the moderate group exhibited accuracy levels in production that were similar to inaccurate participants, in recognition the moderate group was more comparable to the accurate participants. This difference is analogous to the result of Demorest (2001) who found that singers exhibiting inconsistent accuracy (effectively falling between an accurate and an inaccurate group) performed relatively inaccurately in production but performed accurately when matching pitch using a dial. Such intermediate levels of performance may indicate that moderate singers have more of a problem with vocal-motor tuning, and vocal precision, as opposed to sensorimotor mapping (cf. Hutchins, Larrouy-Maestri, & Peretz, 2014; Hutchins & Peretz, 2012; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010). This dissociation is thus useful to help us understand the complicated systems involved in what might at first have been thought of as a simple mirroring task of pitch matching.

A second difference between production and recognition tasks has to do with the greater reliability of the familiarity advantage for the latter task. This may reflect different roles for familiarity across tasks. Our recognition task requires high precision for processing of individual notes that may or may not be displaced. As such, the effect of familiarity in the recognition task may rely more strongly on cognitive abilities, such as pitch working memory (Van Hedger et al., 2018), regardless of how effectively one can engage in sensorimotor mapping. By contrast, the advantage for familiarity in production may be contingent on having sensorimotor mapping that is functioning to some degree, but is also amenable to improvement via guidance from long-term memory. This inverted U-shaped function may indicate that the moderate singers may be both able to benefit from LTM scaffolding and have a basic level of skill that is nevertheless capable of improvement. Inaccurate singers, however, are theorized to experience degraded sensorimotor mapping (Greenspon et al., 2017; Pfordresher et al., 2015) which may have precluded them from benefits associated with long-term representations of pitch.

All our participants engaged in assessment measures to establish baseline measures of accuracy on various tasks. There were some critical respects in which performance on the assessment measures differed from performance in the primary experiment. First, whereas the group comparisons did not reach significance with respect to simple pitch discrimination during assessment, as discussed above, there were considerable differences across groups with respect to recognition performance in the primary experiment. This points to the aforementioned role of complexity in perception/production associations. Moreover, the fact that groups were defined in assessment only by singing accuracy suggests that the recognition task which might seem entirely cognitive may rely to some degree on sensorimotor functioning. Second, although all three groups differed from each other in the pitch imitation assessment task (which isn't surprising given that the grouping itself was based on this performance), the only group difference in production accuracy in the primary experiment was between accurate and moderate participants. This likely reflects the difficulty of the primary experiment task relative to the pitch imitation assessment task, as evidenced by the fact that the three groups did all differ from each other even in the primary experiment for productions of untransformed patterns, which was the easiest task as seen by overall performance levels. It is also worth mentioning that accurate singers exhibited larger short-term memory capacity and higher self-reported imagery abilities compared to moderate singers, while moderate and inaccurate singers performed similarly on assessment measures of auditory STM and imagery.

One limitation of the current research was the frequency with which participants recalled the wrong number of notes (see supplementary results for detailed analyses). Although participants' auditory short-term memory capacity based on our assessment measure was well beyond the limit of 5 units, which was the longest melody length in the current study, this capacity was assessed using a verbal measure (i.e., forward digit span). However, other previous work suggests that pitch STM has a lower capacity than verbal STM (Greenspon & Pfordresher, 2019; Williamson & Stewart, 2010). For instance, Williamson and Stewart (2010) found that individuals without a musical deficit had an average digit span of 7.6, but an average tone span of 6.8. Greenspon and Pfordresher (2019) measured pitch and verbal STM capacity for a sample of over 200 undergraduate students and found overall lower average spans for digits (6.3) and tones (4.5) compared to Williamson and Stewart (2010), but a larger difference between the two. The tone spans reported by Greenspon and Pfordresher (2019) suggest that the length of sequences used in the current study may have exceeded some participants' pitch STM capacity. Participants sang the wrong number of notes even when recalling excerpts of traditional tunes, suggesting that LTM did not reliably extend STM capacity for pitches in the current study.

Another possible limitation stems from the fact that we used traditional melodies that have lyrics, even though we did not present the lyrics during melody presentation. If participants used lyrics as a retrieval cue, the difficulty in re-ordering words may have contributed to the transformation effect. We have two reasons to doubt that this potential issue had a strong effect on results. First, performance on transformed conditions was equivalent for traditional and novel melodies. Second, participants only sang or recognized a short excerpt of the melody. Therefore, they would have only access to incomplete lyrics (e.g., only part of a sentence and in some cases the first syllable of a word). Finally, lyrics do not always facilitate recall. Previous studies have shown that recall of traditional melodies can be less accurate when participants have to generate both text and tune than when they produce either source of information alone (Berkowska & Dalla Bella, 2009; Racette & Peretz, 2007). Although explicit lyric recall was not required in our task, these previous studies suggest that a strategy of using lyrics to retrieve pitch patterns may not necessarily be advantageous.

By using a controlled comparison between well-learned and novel melodies we were able to demonstrate that the effect of familiarity varies by skill level, and that this interaction was seen only for information recall, but not recognition. On the other hand, mental transformations remained reliably difficult regardless of one's degree of familiarity with the tune, the type of manipulation that was performed, or whether memory was measured via recall or recognition. However, we did find that the ability to manipulate musical information varied by skill level: Accurate singers consistently demonstrated greater flexibility of musical information compared to the other two groups. The current study extends Greenspon et al. (2017) claim that musical representations are inflexible by demonstrating that singing ability, but not familiarity, predicts one's ability to manipulate musical information.

Notes

1. This sample was drawn from a larger original sample (N = 134). Forty-six participants from the larger sample (34%) sang the wrong number of notes on the majority of trials in the production task, and were dropped from the analyzed sample. The frequency of dropped trials did not vary by group, transformation condition, or level of familiarity, see Supplementary Results. Five more participants (<4%) were dropped due to extreme pitch discrimination thresholds or missing data resulting from experimenter error. The final sample (n = 83) was selected because it provided the least ambiguous scoring of production data.
2. One traditional melody ("Jingle Bells") comprised eighth notes and one traditional melody ("This Old Man") comprised of quarter notes and one-half note. The rhythm reflected in these two melodies were not considered complex (i.e. syncopated) rhythms. All other melodies in the experiment were composed of isochronous rhythms comprising quarter notes. We observed a similar pattern of results for the study when the production and recognition data was analyzed without the traditional/novel melody pairs for "Jingle Bells" and "This Old Man."
3. As shown in Table 1, we did find group differences in assessment measures in addition to differences in singing accuracy. However, follow-up analyses in which participants were grouped based on their performance on cognitive and self-report assessment measures did not replicate the effects of the Group factor found in the current study, see Supplementary Results.
4. This participant was retained in analyses of production and recognition performance because his production and recognition data did not show evidence that the manipulation of melody familiarity affected him differently than the rest of the sample.
5. In addition to the replicated group × condition interaction in the current study, we directly addressed whether the current study replicated the results of Greenspon et al. (2017) by re-analyzing the current data using the criteria for singing accuracy defined in our earlier paper. As shown in the Supplementary Results, the present data replicated those results.

Acknowledgments

We would like to thank the following undergraduate research assistants from the University at Buffalo, SUNY for their assistance with data collection for the current study: Amy O'Leary, Melissa Mercado, Nicholas Nolan, and Cailin Shupbach.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by NSF Grant BCS-1256964.

References

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

Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology*, *5*, 69–83.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 641–653.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.

De Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, *111*, 1917–1930.

Demorest, S. M. (2001). Pitch-matching performance of junior high boys: A comparison of perception and production. *Bulletin of the Council for Research in Music Education*, *151*, 63–70.

Demorest, S. M., & Clements, A. (2007). Factors influencing the pitch-matching of junior high boys. *Journal of Research in Music Education*, *55*, 190–203.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long- term memory for melodies. *Psychomusicology*, *1*, 30–49.

Greenspon, E. B., & Pfordresher, P. Q. (2019). Pitch-specific contributions of auditory imagery and auditory memory in vocal pitch imitation. *Attention, Perception, & Psychophysics*, *81*, 2473–2481.

Greenspon, E. B., Pfordresher, P. Q., & Halpern, A. R. (2017). Pitch Imitation ability in mental transformations of melodies. *Music Perception*, *34*, 585–604.

Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*, *17*, 572–581.

Halpern, A. R. (2015). Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain*, *25*, 37–47.

Halpern, A. R., & Bartlett, J. C. (2010). Memory for melodies. In M. Jones, A. Popper, & R. Fay (Eds.), *Music Perception* (pp. 234–258). New York, NY: Springer.

Hutchins, S., Larrouy-Maestri, P., & Peretz, I. (2014). Singing ability is rooted in vocal- motor control of pitch. *Attention, Perception, & Psychophysics*, *76*, 2522–2530.

Hutchins, S. M., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology: General*, *141*, 76–97.

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology*, *29*(3229–3243), e3212.

Krumhansl, C. L. (2010). Plink: "Thin slices" of music. *Music Perception*, *27*, 337–354.

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *Journal of Neuroscience*, *29*, 10215–10220.

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action–perception mismatch in tone-deafness. *Current Biology*, *18*, R331–R332.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of" tone deafness". *Music Perception*, *25*, 95–115.

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *The Journal of the Acoustical Society of America*, *128*, 2182–2190.

Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review*, *20*, 747–753.

Pfordresher, P. Q., Halpern, A. R., & Greenspon, E. B. (2015). A mechanism for sensorimotor translation in singing: The multi-modal imagery association (MMIA) model. *Music Perception*, *32*, 242–253.
Pfordresher, P. Q., & Larrouy-Maestri, P. (2015). On drawing a line through the spectrogram: How do we understand deficits of vocal pitch imitation? *Frontiers in Human Neuroscience*, *9*, 271.
Pfordresher, P. Q., & Nolan, N. P. (2019). Testing convergence between singing and music perception accuracy using two standardized measures. *Auditory Perception & Cognition*, *2*, 67–81.
Racette, A., & Peretz, I. (2007). Learning lyrics: To sing or not to sing? *Memory & Cognition*, *35*, 242–253.
Schellenberg, E. G., Iverson, P., & McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review*, *6*, 641–646.
Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, *14*, 262–266.
Takahashi, M., Shimizu, H., Saito, S., & Tomoyori, H. (2006). One percent ability and ninety-nine percent perspiration: A study of a Japanese memorist. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1195–1200.
Van Hedger, S. C., Heald, S. L., & Nusbaum, H. C. (2018). Long-term pitch memory for music recordings is related to auditory working memory precision. *Quarterly Journal of Experimental Psychology*, *71*, 879–891.
Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory*, *18*, 657–669.
Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, version 2.00. *Behavioural Research Methods, Instruments and Computers*, *20*, 6–11.
Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., & Evans, A. C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proceedings of the National Academy of Sciences*, *95*, 3172–3177.

Appendix A

Traditional:
Twinkle, Twinkle, Little Star

Novel:
Arise, Arise, Kind Brother

Traditional:
Old McDonald Had a Farm

Novel:
Wild Robinson Got One Boot

Traditional:
Yankee Doodle

Novel:
Chirpy Canary

Traditional:
Jingle Bells

Novel:
Noble Poet

Traditional:
This Old Man

Novel:
Summer of Fun

Traditional:
Mary Had a Little Lamb

Novel:
Ah David, You Got Music

Figure 1A. Notation for the Traditional and Novel melodies used in the experiment.