Song imitation in congenital amusia: performance partially facilitated by melody familiarity but not by lyrics

Article

Accepted Version

Loutrari, A., Jiang, C. and Liu, F. ORCID: https://orcid.org/0000-0002-7776-0222 (2022) Song imitation in congenital amusia: performance partially facilitated by melody familiarity but not by lyrics. Music Perception, 39 (4). pp. 341-360. ISSN 0730-7829 doi: https://doi.org/10.1525/mp.2022.39.4.341 Available at https://centaur.reading.ac.uk/102531/

To link to this article DOI: http://dx.doi.org/10.1525/mp.2022.39.4.341

Publisher: University of California Press

Song imitation in congenital amusia: Performance partially facilitated by melody familiarity but not by lyrics

Ariadne Loutrari

School of Psychology and Clinical Language Sciences, University of Reading, Earley Gate, Reading RG6 6AL, UK

Cunmei Jiang

Music College, Shanghai Normal University, Shanghai, China, 200234

Fang Liu

School of Psychology and Clinical Language Sciences, University of Reading, Earley Gate, Reading RG6 6AL, UK

Abstract

Congenital amusia is a neurogenetic disorder of pitch perception that may also compromise pitch production. Despite amusics' long documented difficulties with pitch, previous evidence suggests that familiar music may have an implicit facilitative effect on their performance. It remains, however, unknown whether vocal imitation of song in amusia is influenced by melody familiarity and the presence of lyrics. To address this issue, thirteen Mandarin-speaking amusics and 13 matched controls imitated novel song segments with lyrics and on the syllable /la/. Eleven out of these participants in each group also imitated segments of a familiar song. Subsequent acoustic analysis was conducted to measure pitch and timing matching accuracy based on eight acoustic measures. While amusics showed worse imitation performance than controls across seven out of the eight pitch and timing measures, melody familiarity was found to have a favourable effect on their performance on three pitch-related acoustic measures. The presence of lyrics did not affect either group's performance substantially. Correlations were observed between amusics' performance on the Montreal Battery of Evaluation of Amusia and imitation of the novel song. We discuss implications in terms of music familiarity, memory demands, the relevance of lexical information and the link between perception and production.

Keywords:

congenital amusia; imitation; melody familiarity; lyrics; song

Introduction

Humans are thought to be genetically programmed to perceive and produce music (Tan, McPherson, Peretz, Berkovic, & Wilson, 2014). Storing melodic information and recognising familiar melodies appear to emerge naturally in the first year of development among neurotypical children (Plantinga & Trainor, 2005). Singing is another example of a natural musical skill acquired effortlessly during development (Dalla Bella, Giguere, & Peretz, 2007) and could be viewed as a typical human attribute (Mantell & Pfordresher, 2013).

Some individuals experience difficulties with music perception and production, which are not accounted for by brain injury or lack of environmental stimulation (Ayotte, Peretz, & Hyde, 2002). These neurogenetic abnormalities collectively known as 'congenital amusia' (Peretz & Hyde, 2003) are currently estimated to affect 1.5% of the population (Peretz & Vuvan, 2017), while previously the estimate was up to 4% (Kalmus & Fry, 1980). In neurological terms, atypical brain organisation has been observed in amusia, with abnormalities found in the secondary auditory cortex of the right hemisphere (Hyde et al., 2007), the right dorsolateral prefrontal cortex (Schaal, Pfeifer, Krause, & Pollok, 2015) and defective connectivity between the frontal and temporal lobes (Albouy et al., 2013; Hyde, Zatorre, & Peretz, 2011; Loui, Alsop, & Schlaug, 2009). Sensorimotor activity, assessed in terms of the ability to associate the direction of melodic sequences with analogous motoric representations, also appears to differ in amusics, as when they are presented with incongruous musical-visual stimuli (e.g., ascending notes and pictures of descending movements), they do not display the brain activation observed in neurotypical individuals (Jiang, Liu, Zhou, & Jiang, 2019).

In the perception domain, amusics' performance differs considerably from that of neurotypical individuals even in the simple task of detecting melodic violations (Peretz, Brattico, Jrvenp, & Tervaniemi, 2009). Amusics tend to have elevated thresholds for detecting pitch changes (Foxton, Dean, Gee, Peretz, & Griffiths, 2004; Jiang, Lim, Wang, & Hamm, 2013). When it comes to recognising the direction of a pitch interval, amusics' performance is also compromised (Foxton et al., 2004; Jiang et al., 2013; Loui, Guenther, Mathys, & Schlaug, 2008), but as reviewed in Tillmann, Lalitte, Albouy, Caclin and Bigand (2016), amusics with normal pitch thresholds can be found in several cohorts across studies. Perceptual deficits in amusics are not specific to music; they also extend to the speech domain (see, for example, Jiang, Hamm, Lim, Kirk, & Yang, 2010; Liu, Patel, Fourcin, & Stewart, 2010; Nan, Sun, & Peretz, 2010; Tillmann et al., 2011a, 2011b; Zhang, Shao, & Chen, 2018). Some amusics also manifest rhythm-related perceptual deficits (Ayotte et al., 2002; Foxton, Nandy, & Griffiths, 2006; Peretz, Champod, & Hyde, 2003). Task demands may affect the detectability of difficulties in the temporal dimension; rhythmically complex tasks reveal severely compromised performance (Lagrois & Peretz, 2019), but assessing simple local time manipulation detection points to intact performance (Hyde & Peretz, 2004).

Neural models of perceptual-motor behaviour have been developed to examine the relationship between perception and production. Older work on vision (e.g., Goodale & Milner, 1992) and speech (e.g., Hickok & Poeppel, 2004; 2007) has put forward accounts of dual-stream systems associated with distinct neural representations for perception and action. Similar approaches have been adopted in other aspects of music and language processing, such as prosody (Sammler, Grosbras, Anwander, Bestelmeyer, & Belin, 2015), linguistic and musical syntax (Musso et al., 2015) and singing (Loui, 2015). Extending the dual-route account (Griffiths, 2008) into a 'Linked Dual Representation' model, Hutchins and Moreno (2013) proposed that vocal information can be encoded through distinct symbolic and motoric mechanisms, but, in addition, motoric representations can be mediated by symbolic representations.

Consistent with Hutchins and Moreno's (2013) model, there is a general trend of coupled perception and production impairments, as well as violations of this pattern, in the amusia literature (Ayotte at al., 2002; Dalla Bella, Giguère, & Peretz, 2009; Hutchins, Zarate, Zatorre, & Peretz, 2010; Liu et al., 2013, 2016; Tremblay-Champoux, Bella, Phillips-Silver, Lebrun, & Peretz, 2010). For example, in Dalla Bella et al. (2009), variability in singing performance of amusics was found to be explained by pitch perception ability, albeit with exceptions of skilful singing in the presence of compromised pitch perception. Similar paradoxical findings emerged in Loui et al. (2008), where amusics were unable to make pure tone direction judgments but were successful in reproducing these pitch contours and in Liu et al. (2010), where better imitation than identification of intonation contours was also observed. Similarly, amusics were able to imitate, but not discriminate, pitch changes in sentences (Hutchins & Peretz, 2012). Apart from action-perception mismatch as proposed by the dual-route framework (Griffiths, 2008), some discrepancy in the above results may be attributed to the effects of singing with lyrics, memory demand, and melody familiarity on singing performance.

Previous research points to a substantial effect of singing with lyrics on pitch production as well as an interplay between lyrics and memory (Dalla Bella et al., 2009; Mantell & Pfordresher, 2013). Some evidence from amusics suggests that singing with lyrics is easier compared to singing on the same syllable. Amusics in Dalla Bella et al. (2009) had severe difficulties when required to sing a familiar tune replacing the lyrics with a single syllable. Using a more complex design, Tremblay-Champoux et al. (2010) asked amusics to sing a familiar melody in three different conditions; from memory, after being exposed to the melody sung by a model and singing it along with the model. The scores of participants on singing from memory (first condition) were compared to imitation (second and third conditions taken together). It was found that imitation was an easier condition, suggesting that poor singing might be of some memory-related origin, which is consistent with amusics' documented short-term memory deficits (e.g., Albouy et al., 2013; Gosselin, Jolicœur, & Peretz, 2009; Tillmann, Schulze, & Foxton, 2009; Williamson & Stewart, 2010; but see Jiang et al., 2013 on the adverse effect of perceptual complexity on memory performance) and their possibly less efficient long-term memory storage in addition to short-term memory difficulties (Graves et al., 2019).

Literature from both non-amusic and amusic individuals points to a facilitative effect of melody familiarity on cognitive processing. Preschool-aged children are known to form better audio-visual associations when it comes to familiar melodies and to display enhanced processing of familiar melodies when tested on same-different discrimination tasks (Creel, 2019). Passive listening in neurotypical adults has been found to yield different electrophysiological responses depending on whether a song segment is familiar or unfamiliar (Jagiello, Pomper, Yoneya, Zhao, & Chait, 2019). Familiar melodies are associated with the formation of auditory mental imagery (Halpern & Zatorre, 1999; Kraemer, Macrae, Green, & Kelley, 2005) and a greater memory advantage over newly presented material (Peretz, Gaudreau, & Bonnel, 1998). Importantly, compromised mental imagery has been found to contribute to poor pitch imitation (Greenspon, Pfordresher, & Halpern, 2017). When implicitly tested, familiarity appears to influence music processing in the amusic population. Ayotte et al. (2002) found that amusics had difficulties in making familiarity decisions for melodies devoid of lyrics, but when they were required to detect out-of-key notes in a different task, familiar melodies were associated with a better performance over unfamiliar melodies. Tillmann, Albouy, Caclin and Bigand (2014) corroborated this implicit familiarity effect. Their participants were presented with segments of gradually augmenting duration and were asked to provide subjective judgements of perceived familiarity on a rating scale. Amusics performed similarly to controls and their judgments were also consistent throughout trials. Recent work on amusia also gives some indication of reduced electrophysiological responses to pitch deviants in novel (versus familiar) melodies relative to neurotypicals (Quiroga-Martinez et al., 2021). Overall, empirical evidence suggests that melody familiarity may have an implicit facilitative effect on perception in the amusic population. It remains, however, unknown whether imitating familiar versus novel material leads to differences in performance within a given amusic sample.

In addition to pitch production deficits, rhythm-related production impairments are also observed in the amusic population. Lagrois and Peretz (2019) demonstrated that their amusic sample, as a whole, could not successfully tap along to the beat, although some exceptions were found. In another study, singing a familiar song from memory was found to display substantially larger temporal variability in amusics compared to controls, although a large proportion of amusics sang in-time and their performance was found to be unimpaired when analysed in terms of tempo, rubato consistency and number of time errors (Dalla Bella et al., 2009). Tremblay-Champoux et al. (2010) did not find any rhythm-related impairments in amusics in either singing or imitation. In contrast, amusics in Liu et al. (2013) were found to display impaired vocal imitation of speech and song when their output was analysed in terms of absolute time matching, relative time matching and number of time errors.

In summary, the amusia studies reviewed thus far have looked at singing familiar songs from memory (e.g., Dalla Bella et al. 2009; Liu et al., 2016; Tremblay-Champoux et al., 2010), imitating relatively lengthy tunes (Tremblay-Champoux et al., 2010) or single tones (Hutchins et al., 2010), imitating speech (Hutchins & Peretz, 2012) or speech versus song (Liu et al., 2013), or imitating pairs of tones by humming (Loui et al., 2008). Taken together, the literature reveals a general trend between impaired pitch perception and production, but also intact or less impaired imitation compared to perception, less prevalent rhythm-related deficits, less pronounced difficulties in singing on a single syllable when memory demands are relatively low and an implicit effect of familiarity on the performance of amusic individuals. It remains to be explored whether amusics' imitation performance would differ depending on the familiarity of the song to be imitated, and singing with lyrics or on a syllable. In this study, we directly compared imitation of a novel song and a familiar song, with lyrics or on /la/, in Mandarin-speaking amusics and matched controls. To control for memory demands, participants were not required to sing from memory or imitate any segments longer than eight syllables. Following Liu et al. (2013), we based the evaluation of participants' performance on acoustic measures of absolute and relative pitch and timing matching. In light of the existing evidence, we predicted that impaired pitch perception would more often than not lead to impaired production in amusia. Also, given the limited memory demands and the short length of the target stimuli employed here, we aimed to determine whether singing through imitation on the syllable /la/ would be as difficult for amusics as shown in previous work on singing from memory (Dalla Bella et al., 2009). Finally, and most importantly, considering previously reported indirect effects of familiarity on amusic performance and evidence from the general population, we set out to explore whether imitating short segments of a familiar song (the Happy Birthday song) would be as challenging for amusics as imitating segments of an unfamiliar song.

Method

Participants

A total of 13 native speakers of Mandarin Chinese with congenital amusia and an equal number of matched control participants took part in the novel song imitation experiment, which was conducted together with the experiments reported in Liu et al. (2012; 2013). 22 out of the 26 participants (11 amusics and 11 controls) also returned a few months later to participate in the familiar song imitation experiment. The sample size of the present study is comparable to that typically employed in studies on congenital amusia (e.g., Ayotte et al., 2002; Jiang et al., 2010; Lagrois & Peretz, 2019; Peretz et al., 2009; Pfeuty & Peretz, 2010; Tillmann et al., 2009, 2014; Tremblay-Champoux et al., 2010). All participants were recruited through bulletin board advertisements in Beijing, China and were undergraduate or Master's students. Those with a total score of 65 or lower out of 90 on the scale, contour and interval subtests of the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003) were classified as amusics. None of the participants reported any speech, language or hearing impairment, learning difficulty, memory problem or history of neurological or psychiatric disorders. Also, none of the participants had undergone any formal music training. Amusics and controls were matched on sex, handedness, age and education, with their only differing characteristic being performance on the MBEA and pitch threshold tasks (Liu et al., 2012, 2013). The performance difference between amusics and controls was larger on the pitch subtests and was especially pronounced on the scale subtest of the MBEA. Table 1 shows the demographic information as well as MBEA scores of these participants.

[**Insert Table 1 here**]

Stimuli

Target song stimuli were recorded in a soundproof booth at Goldsmiths, University of London by a female amateur singer, who composed the novel song used in this study. The singer was a native speaker of Mandarin Chinese and had received 16 years of formal musical training.

The score of the novel song, entitled "sea turtles", is shown in Fig. 1. It consisted of a total of 17 Mandarin phrases, ranging from two to eight syllables (Table 2). Note that the songwriter did not follow the convention of ending on the tonic, as shown in Fig. 1. However, as singing phrases were presented in isolation, listeners were not exposed to the song as a whole. Hence, the fact that the novel song does not end on the tonic is unlikely to have had an effect on performance. The song was recorded in two different conditions by the singer: one with lyrics and the other with a repeated /la/ syllable replacing all words. The mean pitch of the notes in the song was 290 Hz (S*D* = 59.68 Hz) when it was sung with lyrics and 288 Hz (S*D* = 59.22 Hz) when it was sung on the syllable /la/. The mean duration of the note rhymes was 375 msec (S*D* = 300 msec) when it was sung with lyrics and 376 msec (S*D* = 291) when it was sung on the syllable /la/. The output of the singer is referred to as "the model" in subsequent sections.

[**Insert Figure 1 here**]

[**Insert Table 2 here**]

The familiar song was "Happy Birthday" in Mandarin Chinese, which has the same melody as the English version. Similar to the novel song, it was recorded with lyrics and on the syllable /la/ by the same singer. The song was split into four separate musical phrases, allowing for a similar imitation task as the one carried out for the novel song. All phrases but the third one contained six syllables. The last vowel of the third phrase was sung twice, and so it included seven notes. The mean pitch of the notes in the song was 292 Hz (S*D* = 65.44 Hz) when it was sung with lyrics and 290 Hz (S*D* = 65.30 Hz) when it was sung on the syllable /la/. The mean duration of the note rhymes was 415 msec (S*D* = 176 msec) when it was sung with lyrics and 466 msec (S*D* = 221 msec) when it was sung on the syllable /la/. A detailed description of both songs is presented in Table 3.

[Insert Table 3 here]

Fig. 2 shows the pitch contours of a set of target stimuli (black dots) across conditions plotted against imitation performance by an amusic (red squares) and a control (green diamonds) participant. As can be seen, the model singer produced consistent pitch patterns across the lyrics and /la/ conditions of the songs, and the control participant imitated the songs better than the amusic participant across all conditions.

[**Insert Figure 2 here**]

In order for participants of different gender to imitate target stimuli of the same gender, the recorded novel/familiar song stimuli were synthesised into natural-sounding female (preserving the absolute pitches and formant frequencies of the original recordings) and male voices (changing the original pitches to one octave lower and shifting the frequencies of the original formants by .78 so as to achieve male voice characteristics) using the command “change gender” in Praat (Boersma and Weenink, 2019). None of the participants commented that either the female or male voice sounded unnatural, and no significant differences were found in imitation performance between participants of different gender for either the amusic or control group. Indicatively, a Welch’s *t*-test for unequal sample sizes showed no significant differences between female and male participants in either the amusic ($p = 0.33$) or the control group ($p = 0.17$) in terms of absolute pitch deviation and a similar pattern was observed for duration difference ($p = 0.94$ and $p = 0.28$ for amusics and controls respectively). Therefore, the syntheses of the female/male target stimuli were unlikely to have caused any adverse effects on imitation performance.

Following the synthesis process, target melodies were acoustically analysed using Praat in order to enable comparisons with the participants’ output. More specifically, all syllable rhymes were labelled and extracted and measurements of pitch and duration patterns were obtained employing the ProsodyPro script (Xu, 2013).

Procedure

Testing sessions took place in a quiet room at the Institute of Psychology of the Chinese Academy of Sciences in Beijing, China. Ethical approval was obtained from the Chinese Academy of Sciences and Goldsmiths, University of London. All participants gave written informed consent before taking part in the study. To help them familiarise themselves with the task, practice items were given before the beginning of the main portion of the experiment. Participants were presented with one short song segment at a time and were instructed to try and imitate the pitch and timing patterns as closely as possible directly after stimulus presentation while their imitation was recorded using Praat. Participants had the opportunity to listen to a stimulus again or repeat the trial in case of disfluency. That said, stimuli and trials were very rarely repeated. For the novel song, participants were given the lyrics to read both beforehand and during the recording. All participants were familiar with the lyrics of the "Happy Birthday" song.

Data Analysis

Acoustic analysis of the obtained data was performed using the Praat script ProsodyPro (Xu, 2013), by manually correcting any missing, double or wrongly inserted vocal pulses generated by Praat. To examine imitation accuracy in our participants, we followed the procedure in Liu et al. (2013) and extracted the median F_0 and duration of the syllable rhymes and conducted eight measurements, namely absolute pitch deviation, pitch interval deviation, signed interval deviation, number of contour errors, number of pitch interval errors, duration difference, number of time errors and interonset interval difference using the ensemble files generated by the script. The ensemble files obtained from the model were used as a point of reference for all comparisons on pitch and duration.

Absolute pitch deviation was calculated in semitones as the absolute difference between the median F_0 of each imitated sung syllable and the syllable sung by the model. Bigger values corresponded to bigger deviations and, therefore, less accurate imitation. Octave errors were corrected so that pitch deviations above 6 semitones were replaced by their subtractions from 12 semitones.

Pitch interval deviation was also calculated in semitones and corresponded to the absolute difference between the participants' output and the model's output in intervallic relationships in each sung sentence. The difference in median F_0 between consecutive sung syllables was calculated as the pitch interval. A greater value of the pitch interval deviation was the result of a larger deviation between the imitated interval and the target interval. Similar to the previous measure, octave errors were corrected in the obtained data.

Signed interval deviation corresponded to the difference between imitated intervals (in absolute values) and target intervals (in absolute values), with negative deviations relating to compressions and positive deviations to expansions. Results were obtained after octave error correction.

The number of contour errors corresponded to the total number of imitated intervals that differed from the target pitch direction after octave error correction. Pitch direction could be ascending, descending or level; it was counted as ascending/descending when the difference in median F_0 between two consecutive notes was at least half a semitone (50 cents) and flat when the difference was within 50 cents.

The number of pitch interval errors corresponded to the number of produced intervals that differed in magnitude from the model. An error was scored for every interval that differed by at least half a semitone from the target. Pitch direction was not considered here.

The duration difference was calculated in milliseconds and corresponded to the absolute difference in rhyme length between the imitation and the target. Similar to the previous measures, a bigger value indicated a less accurate imitation.

The number of time errors corresponded to the number of syllables that were more than 25% shorter or longer from target syllables.

Interonset intervals (IOIs) were calculated in milliseconds on the basis of the duration between the onsets of two consecutive rhymes, as a measure of relative time matching, with bigger values pointing to a less accurate imitation.

Statistical analysis and figure plotting were conducted using R (R Core Team, 2019; version 4.0.3). The analysis included data from 26 participants for the novel song conditions and 22 participants for the familiar song conditions with lyrics and on /la/. In addition to the presence/absence of lyrics and melody familiarity, the analysis took into account the number of syllables as well as the mean absolute pitch, pitch intervals, signed pitch intervals, durations and IOIs of the target melodies as covariates. Linear mixed-effects models, fitted using REML estimates of parameters, were employed to assess the effects of the above conditions and stimulus features, with participants and items as random effects. The lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) and lme4 (Bates, Mächler, Bolker, & Walker 2015) packages were used, providing significance testing for fixed effects. Satterthwaite's method for approximating degrees of freedom for the t and F tests were implemented in the packages. Note that linear mixed-effects models can handle missing data (Magezi, 2015), which was necessary in the present study given that not all participants completed both parts. Although mixed modelling is robust to assumption violations (Brown, 2021), model assumptions were inspected visually using residual and normal probability plots. Formulas tested each acoustic measure in relation to group information (amusic vs. control), stimulus type (lyrics vs. /la/), song condition (novel vs. familiar), number of syllables and model characteristics where appropriate. Post-hoc comparisons were conducted using the *emmeans* package, with p-values adjusted using the Holm method and degrees-of-freedom derived using the Kenward-Roger method (Lenth, 2019). Correlation analyses were also conducted to examine the relationship between production (imitation) and perception (MBEA scores and pitch change detection and pitch direction discrimination thresholds).

Results

Absolute pitch deviation

Overall, there was a significant main effect of group on absolute pitch matching, with amusics performing significantly worse than controls [F(1, 24.46) = 14.19, $p < 0.001$, $\eta_p^2 = 0.36$]. There was no significant effect of stimulus type [F(1, 1011.58) = 1.35, $p = 0.24$, $\eta_p^2 = 0.001$], song condition [F(1, 17.70) = 0.33, $p = 0.56$, $\eta_p^2 = 0.01$], model absolute pitch [F(1, 39.70) = 3.67, $p = 0.06$, $\eta_p^2 = 0.08$] or number of syllables [F(1, 20.29) = 0.19, $p = 0.66$, $\eta_p^2 = 0.009$]. No significant interactions were found for group × stimulus type [F(1, 1008.98) = 0.01, $p = 0.89$, $\eta_p^2 = 0.001$], group × song condition [F(1, 1011.03) = 1.31, $p = 0.25$, $\eta_p^2 = 0.001$], stimulus type × song condition [F(1, 1008.98) = 0.08, $p = 0.77$, $\eta_p^2 < 0.001$] or group × stimulus type × song condition [F(1, 1008.98) = 0.49, $p = 0.48$, $\eta_p^2 < 0.001$]. Performance on all conditions is shown in Fig. 3.

[**Insert Figure 3 here**]

Pitch interval deviation

Considering scores on the pitch interval deviation measure as a whole revealed a significant main effect of group on performance, with amusics performing significantly worse than controls [F(1, 28.79) = 16.11, $p < 0.001$, $\eta_p^2 = 0.35$]. There was also a significant main effect of song condition [F(1, 18.70) = 13.78, $p = 0.001$, $\eta_p^2 = 0.42$], pointing to a better performance on familiar melody imitation, and model mean pitch interval [F(1, 17.08) = 113.96, $p < 0.001$, $\eta_p^2 = 0.87$], with bigger intervals leading to less accurate imitation. There was no significant main effect of the number of syllables [F(1, 16.88) = 4.15, $p = 0.057$, $\eta_p^2 = 0.19$] or stimulus type [F(1, 1008.95) = 0.39, $p = 0.52$, $\eta_p^2 < 0.001$]. There was a significant group × song condition interaction, albeit with a small effect size, [F(1, 1016) = 8.45, $p = 0.003$, $\eta_p^2 = 0.008$]. The effect sizes stemming from post-hoc pairwise comparisons suggested that the group difference was more pronounced in the novel song condition [$t(25.1) = 5.47$, $p < 0.001$, d = 2.18] than the familiar song condition [$t(59.5) = 2.28$, $p = 0.02$, d=0.59]. There was no significant group × stimulus type interaction [F(1, 1008.92) = 0.42, $p = 0.51$, $\eta_p^2 < 0.001$], stimulus type × song condition interaction [F(1, 1009.02) = 0.23, $p = 0.62$, $\eta_p^2 < 0.001$] or group × stimulus type × song condition interaction [F(1, 1008.92) = 0.01, $p = 0.91$, $\eta_p^2 < 0.001$]. Performance on the above conditions is shown in Fig. 4.

[**Insert Figure 4 here**]

Number of contour errors

As far as contour errors are concerned, the mixed-effects model detected a significant main effect of group [F(1, 44.20) = 4.93, $p = 0.03$, $\eta_p^2 = 0.10$], pointing to impaired amusic performance, song condition [F(1, 18.14) = 32.60, $p < 0.01$, $\eta_p^2 = 0.64$], with the novel condition being associated with more errors and number of syllables [F(1, 17.84) = 13.03, $p = 0.002$, $\eta_p^2 = 0.42$], with greater syllable numbers leading to more errors. No significant effect of stimulus type was observed [F(1, 1009.43) = 0.09, $p = 0.75$, $\eta_p^2 < 0.001$]. There was a significant group × song condition interaction [F(1, 1028.37) = 5.56, $p = 0.01$, $\eta_p^2 = 0.005$]. Post hoc pairwise comparisons confirmed that amusics performed significantly worse than controls in the novel song condition [$t(28.2) = 4.50$, $p < 0.001$, $d = 1.69$], whereas no significant difference between groups was observed in the familiar song condition [$t(207.8) = 0.27$, $p = 0.78$, $d = 0.03$]. No significant interactions were found for group and stimulus type [F(1, 1009.43) = 0.0005, $p = 0.98$, $\eta_p^2 < 0.001$], stimulus type and song condition [F(1, 1009.43) = 2.54, $p = 0.11$, $\eta_p^2 = 0.003$] or group, stimulus type and song condition [F(1, 1009.43) = 0.16, $p = 0.68$, $\eta_p^2 < 0.001$]. Results are shown in Fig. 5.

[**Insert Figure 5 here**]

Number of pitch interval errors

The analysis here showed a significant main effect of group [F(1, 31.15) = 30.94, $p < 0.001$, $\eta_p^2 = 0.49$], with amusics performing significantly worse than controls. There was also a significant main effect of song condition [F(1, 18.39) = 7.01, $p = 0.01$, $\eta_p^2 = 0.27$], with more errors on average in the novel song condition and a main effect of the number of syllables [F(1, 17.92) = 24.26, $p < 0.001$, $\eta_p^2 = 0.57$], with more syllables leading to worse performance. There was no significant effect of stimulus type [F(1, 1009.13) = 0.35, $p = 0.55$, $\eta_p^2 < 0.001$]. No significant interactions were observed for group and stimulus type [F(1, 1009.13) = 2.72, $p = 0.9$, $\eta_p^2 = 0.003$], group and song condition [F(1, 1018.89) = 0.06,

$p = 0.80$, $\eta_p^2 < 0.001$], stimulus type and song condition [F(1, 1009.13) = 1.50, $p = 0.22$, $\eta_p^2 = 0.001$] or group, stimulus type and song condition [F(1, 1009.13) = 1.15, $p = 0.28$, $\eta_p^2 = 0.001$]. Results are shown in Fig. 6.

[**Insert Figure 6 here**]

Duration difference

Taken as a whole, results pointed to a significant main effect of group on performance [F(1, 29.15) = 5.32, $p = 0.02$, $\eta_p^2 = 0.15$], revealing impaired performance in amusics. There was also a main effect of song condition [F(1, 17.73) = 13.80, $p = 0.001$, $\eta_p^2 = 0.43$], with a larger duration difference from the model when participants imitated the novel song segments, model mean duration [F(1, 22.43) = 137.44, $p < 0.001$, $\eta_p^2 = 0.86$], with larger duration patterns leading to larger deviation and number of syllables [F(1, 18.76) = 38.08, $p < 0.001$, $\eta_p^2 = 0.67$], with more syllables pertaining to worse performance. No significant effect of stimulus type was observed [F(1, 1025.31) = 1.65, $p = 0.20$, $\eta_p^2 = 0.002$]. There was no significant group × stimulus type interaction [F(1, 1008.11) = 3.28, $p = 0.07$, $\eta_p^2 = 0.003$], group × song condition interaction [F(1, 1015.47) = 0.26, $p = 0.60$, $\eta_p^2 < 0.001$], stimulus type × song condition interaction [F(1, 1019.13) = 0.33, $p = 0.56$, $\eta_p^2 < 0.001$] and group × stimulus type × song condition interaction [F(1, 1008.11) = 1.06, $p = 0.30$, $\eta_p^2 = 0.001$]. Performance on all conditions is presented in Fig. 7.

[**Insert Figure 7 here**]

IOIs

As far as IOIs are concerned, the mixed-effects model detected a significant main effect of group [F(1, 36.43) = 21.10, $p < 0.001$, $\eta_p^2 = 0.36$], with amusics performing statistically more poorly than controls, song condition [F(1, 18.73) = 27.53, $p < 0.001$, $\eta_p^2 = 0.59$], with familiar melodies leading to a better performance, model mean IOIs [F(1, 19.98) = 74.24, $p < 0.001$, $\eta_p^2 = 0.78$], with larger model interonset intervals being related to poorer performance and number of syllables [F(1, 17.65) = 54.89, $p < 0.001$, $\eta_p^2 = 0.24$], with larger syllable numbers being associated to less accurate performance. No significant effect of stimulus type was observed [F(1, 1009.29) = 0.84, $p = 0.35$, $\eta_p^2 = 0.001$]. There was no significant interaction between group and stimulus type [F(1, 1009.25) = 2.56, $p = 0.10$, $\eta_p^2 = 0.003$], group and song condition [F(1, 1023.94) = 0.01, $p = 0.89$, $\eta_p^2 = 0.001$], stimulus type and song condition [F(1, 1026.09) = 0.58, $p = 0.44$, $\eta_p^2 = 0.001$] or group, stimulus type and song condition [F(1, 1009.25) = 0.39, $p = 0.53$, $\eta_p^2 < 0.001$]. Results are presented in Fig. 8.

[**Insert Figure 8 here**]

Number of time errors

The analysis revealed a significant effect of group [F(1, 30.95) = 7.25, $p = 0.01$, $\eta_p^2 = 0.19$], with amusics performing worse than controls, stimulus type [F(1, 1009.22) = 13.37, $p < 0.001$, $\eta_p^2 = 0.01$], with singing on the syllable /la/ leading to more time errors, song condition [F(1, 18.57) = 12.24, $p = 0.002$, $\eta_p^2 = 0.39$], with novel segments being on average associated with more errors and number of syllables [F(1, 17.87) = 68.67, $p < 0.001$, $\eta_p^2 = 0.79$], with more syllables being associated with poorer performance. There were no significant interactions between group and stimulus type [F(1, 1009.22) = 1.92, $p = 0.16$, $\eta_p^2 = 0.002$], group and song condition [F(1, 1018.58) = 0.77, $p = 0.38$, $\eta_p^2 = 0.001$], or group, stimulus type and song condition [F(1, 1009.22) = 1.29, $p = 0.25$, $\eta_p^2 = 0.001$]. However, a significant interaction between stimulus type and song condition was observed [F(1, 1009.22) = 7.79, $p = 0.005$, $\eta_p^2 = 0.008$. Post hoc pairwise comparisons showed that singing on /la/ as opposed to singing with lyrics did not differentiate performance in the novel song condition [$t(1009) = 1.06$, $p = 0.28$, $d = 0.06$]. Singing on the syllable /la/ led to more mistakes than singing with lyrics in the familiar song condition [$t(1009) = 3.53$, $p < 0.001$, $d = 0.22$]. The results are visually depicted in Fig. 9.

[**Insert Figure 9 here**]

Signed interval deviation

Fig. 10 presents signed interval deviations of amusics and controls from the model intervals in the four experimental conditions. Deviation from target pitches was calculated in semitones and corresponded to the difference between the absolute value of the imitated interval and the absolute value of the target. Negative deviations represent interval compressions while positive deviations point to interval expansions. As illustrated in Fig. 10, the majority of imitation errors pertained to compressions. Results from the mixed-effects model suggested that signed interval deviation was associated with a significant effect of model mean pitch interval [F(1, 17.29) = 130.62, $p < 0.001$, $\eta_p^2 = 0.88$], with larger intervals relating to greater signed pitch interval deviation. There were no significant effects of group, stimulus type, song condition and number of syllables [F(1, 28.83) = 1.92, $p = 0.17$, $\eta_p^2 = 0.06$, F(1, 1008.74) = 1.40, $p = 0.23$, $\eta_p^2 = 0.001$, F(1, 18.35) = 0.74, $p = 0.40$, $\eta_p^2 = 0.03$ and F(1, 17.03) = 0.02, $p = 0.87$, $\eta_p^2 = 0.001$, respectively]. However, there was a significant interaction between group and song condition, [F(1, 1016.30) = 4.38, $p = 0.03$, $\eta_p^2 = 0.004$]. Post hoc pairwise comparisons showed that amusics' performance was statistically lower than that of controls in the novel song condition [$t(25.1) = -2.42$, $p = 0.02$, $d = -0.96$], while no significant difference was observed between amusics and controls in the familiar song condition [$t(62.5) = -0.37$, $p = 0.71$, $d = 0.09$]. There were no significant interactions between group and stimulus type [F(1, 1008.70) = 0.11, $p = 0.72$, $\eta_p^2 < 0.001$], stimulus type and song condition [F(1, 1008.83) = 0.10, $p = 0.74$, $\eta_p^2 < 0.001$] or group, stimulus type and song condition [F(1, 1008.70) = 0.09, $p = 0.75$, $\eta_p^2 < 0.001$].

[**Insert Figure 10 here**]

Perception and production

Correlation analyses were performed to evaluate the relationship between perception (as assessed by the MBEA and pitch change detection and pitch direction discrimination thresholds) and production (as assessed by the imitation measures for novel and familiar melodies) in the current participants. As higher scores on the MBEA correspond to higher numbers of correct trials and higher pitch thresholds suggest poorer pitch sensitivity, while larger values on the production measures used in this study correspond to less accurate imitation, negative correlations between MBEA scores and imitation measures and positive correlations between pitch thresholds and imitation measures below suggest that better performance on music and pitch perception tallies with better imitation performance.

For the familiar song condition in amusics, MBEA meter was negatively correlated with duration difference scores [$r(9) = -0.74$, $p = 0.009$], IOI scores [$r(9) = -0.64$, $p = 0.03$] and number of time errors [$r(9) = -0.63$, $p = 0.03$]. A positive correlation was observed between MBEA rhythm and the number of contour errors [$r(9) = 0.61$, $p = 0.04$]. However, none of these correlations remained significant after Holm's correction for multiple correlations. The uncorrected correlations are presented in more detail in Table 3.

[**Insert Table 3 here**]

As regards amusics' performance on the novel song condition, MBEA interval was found to be negatively correlated with duration difference [$r(11) = -0.76$, $p = 0.006$] and IOIs [$r(11) = -0.80$, $p = 0.003$]. MBEA memory was negatively correlated with the number of contour errors [$r(11) = -0.63$, $p = 0.03$] and the number of pitch interval errors [$r(11) = -0.64$, $p = 0.03$]. MBEA scale was negatively correlated with pitch interval deviation [$r(11) = -0.70$, $p = 0.01$], number of pitch interval errors [$r(11) = -0.73$, $p = 0.001$], number of contour errors [$r(11) = -0.80$, $p = 0.003$], duration difference [$r(11) = -0.82$, $p = 0.001$], IOIs [$r(11) = -0.61$, $p = 0.04$] and number of time errors [$r(11) = -0.96$, $p < 0.001$]. MBEA meter was negatively correlated with duration difference [r(11) = -0.86, $p < 0.001$], number of time errors [r(11) = -0.81, $p = 0.002$] and number of contour errors [r(11) = -0.63, $p = 0.03$]. MBEA rhythm was negatively correlated with IOIs [r(11) = -0.66, $p = 0.02$]. Negative correlations were also observed between pitch composite score and number of pitch interval errors, number of contour errors, duration difference, IOIs and number of time errors [r(11) = -0.62, $p = 0.04$, r(11) = -0.81, $p = 0.002$, r(11) = -0.87, $p < 0.001$, r(11) = -0.83, $p = 0.001$ and r(11) = -0.75, $p = 0.008$ respectively]. Only the correlations between scale and number of time errors and pitch composite and duration difference remained significant after adjustment using Holm's correction. Significant and non-significant correlations before correction are shown in Table 4.

[**Insert Table 4 here**]

For the familiar song condition, controls' performance on MBEA interval was negatively correlated with absolute pitch deviation [r(9) = -0.65, $p = 0.03$], number of time errors [r(9) = -0.68, p = 0.02] and duration difference [r(9) = -0.70, $p = 0.01$]. MBEA contour was negatively correlated with the number of pitch interval errors [r(9) = -0.65, $p = 0.03$] and pitch change detection was positively correlated with the number of contour errors [r(9) = 0.69, $p = 0.02$]. These correlations did not remain significant after correction. Results are shown in Table 5. [**Insert Table 5 here**]

No correlations between the perceptual measures and the production measures of the novel song condition were seen in controls, as shown in Table 6.

[**Insert Table 6 here**]

Discussion

The present study investigated the ability of congenitally amusic individuals and matched controls to imitate segments of a novel and a familiar song, with lyrics or on the syllable /la/, comparing their output with target stimuli in terms of pitch and timing matching.

Results showed that amusics' performance was significantly lower than that of controls across all pitch and timing imitation measures except signed interval deviation, as suggested by a significant main effect of group across these measures when data from all conditions were pooled. However, three significant interactions were observed between group and song condition, pointing to less impaired performance in amusics when imitating segments of the familiar song as compared to the novel song, in terms of number of contour errors, signed interval deviation and pitch interval deviation (although the difference in terms of the last measure is only appreciable when looking at effect size differences between novel and familiar trials). Nevertheless, for measures of absolute pitch deviation, number of pitch interval errors, duration difference, interonset interval difference, and number of time errors, amusics showed impaired performance relative to controls in both familiar and novel song conditions. The presence of lyrics versus /la/ did not affect either group's performance substantially. A stronger relationship between music perception and novel melody imitation as opposed to familiar melody imitation was observed in amusics. These findings provide insight into how music familiarity, memory demands, lexical information and perception may influence music production.

Our finding on melody familiarity suggests that familiar melody alleviates, to a certain extent, production difficulties in amusics, especially in terms of pitch interval and contour matching, whereas this was not observed in the novel song condition. These results expand previous research in amusia showing reproduction of pairs of tones without violation of the assigned contours (Loui et al., 2008). The encoding of familiar melodies may have an advantage over that of novel melodies due to the presence of memory traces supporting the former (Peretz et al., 1998). Also, given the documented role of auditory mental imagery on imitation (Greenspon et al., 2017) and the evidence that listeners create auditory imagery for familiar music (Halpern & Zatorre, 1999; Kraemer et al., 2005), it can be argued that existing memory structures could facilitate imitation performance in otherwise compromised music cognition. Such explanation is consistent with existing evidence on the indirect effect of familiarity on amusic performance in detecting sour notes in familiar melodies (Ayotte et al., 2002) and providing subjective familiarity judgements (Tillmann et al., 2014). In the production domain, it is worth noting that singing a familiar song and imitating it are distinct processes and should be expected to place different demands on memory. Singing a familiar song from memory has been demonstrated to be impaired in amusics (e.g., Dalla Bella et al. 2009; Liu et al., 2016), but the process of singing a familiar song from memory appears to be more demanding than simply imitating it (Tremblay-Champoux et al., 2010). More specifically, amusics' performance in Tremblay-Champoux et al. (2010) displayed reduced contour errors and interval deviations when imitation was tested. Crucially, however, imitation itself was not without problems in their amusic participants, even in the case of singing in unison. The length of the material to be imitated in Tremblay-Champoux et al. (2010) may have posed heavier memory demands than those of the present study. The song chorus used in Tremblay-Champoux et al. (2010) corresponded to 32 pitches. It can be reasonably assumed that imitating this large number of pitches may remain taxing for amusics' memory, given their known impaired pitch memory representations (Tillmann et al., 2009, 2016; Williamson, McDonald, Deutsch, Griffiths, & Stewart, 2010). The familiar song segments in our study did not exceed seven syllables to be imitated at a time, thus rendering familiarity and low memory demands a possible facilitative combination.

The impaired imitation performance in the novel song condition in amusia may be attributed at least in part to perceptual problems, as our correlation analyses showed that perception and production performance in our amusic participants appeared more closely linked when novel material was considered; the only correlations that remained significant after correction were seen in the novel song condition. However, the correlation results are secondary to the scope of this work and remain indicative. Future work delving into the perception-production link across different degrees of familiarity and complexity of stimuli is needed to shed more light on this intricate relationship. The present findings from the novel song condition are consistent with previous work showing a connection between pitch perception and pitch production deficits in congenital amusia (e.g., Ayotte et al., 2002; Dalla Bella et al. 2009; Liu et al., 2013; Pfordresher & Nolan, 2019). It is crucial to acknowledge, however, that the MBEA (Peretz et al., 2003), employed here and in other studies to assess perception, comprises novel melodies. The current findings could be interpreted in the light of the ‘Linked Dual Representation’ model (Hutchins & Moreno, 2013), which accounts for cases where symbolic encoding impairments lead to production difficulties but also cases where motor representations are successfully encoded without symbolic representation mediation. If motoric representations can be preserved despite compromised perceptual representations (Hutchins & Moreno, 2013; Hutchins et al., 2010), it could be further argued that ubiquitous songs, such as the “Happy Birthday” tune employed here, may come with better motoric representations, conducive to enhanced performance. It has been postulated that melodies heard in a vocal form, as it is often the case with “Happy Birthday”, may come with strong sensorimotor integration, which can result in better consolidated vocal-motor memory traces (Wood, Rovetti, & Russo, 2020). Further, familiar music seems to lead to a more pronounced emotional response (Pereira et al., 2011). This emotional advantage, if spared in amusics (e.g., Gosselin, Paquette, & Peretz, 2015; Lévêque et al., 2018), may also provide more reliable long-term memory representations. These points, however, remain speculative and further research is needed to test such hypotheses and disentangle motoric and affective effects from perceptual memory traces of familiar melodies (see discussion of this point below).

In the current sample, singing performance was found to display substantial variation in line with previous findings from amusics (Hutchins et al., 2010; Hutchins & Peretz, 2013) and neurotypical individuals (Pfordresher & Brown, 2017). Closer inspection of our data corroborated a lack of universal singing ability. It is of note that even in the novel song condition, being amusic did not necessarily lead to a poorer performance and being non-amusic did not always equal accurate singing, as shown in Figures 4-10. Defining amusic and non-amusic singing is by itself challenging. A total of 10-15% of the neurotypical sample studied in Pfordresher and Brown (2007) were found to deviate at least by one semitone from target pitches when trying to imitate them and were subsequently classified as "poor singers". This shows that non-amusics can also manifest poor pitch imitation. Similar to non-amusics, considerable pitch production variation is observed in amusics and they should not be viewed as a homogeneous group (Hutchins & Peretz, 2013). That is, pitch production abilities in amusics do not need to be considered in absolute and universal terms (Hutchins et al., 2010), which could be argued to mirror variation in neuronal abnormalities observed in the population (Albouy et al., 2013; Hyde et al., 2007; 2011; Loui et al., 2009; Schaal et al., 2015).

Moving on to imitation with lyrics and on the syllable /la/, the presence of lyrics was not found to substantially affect performance in the present study. Previous findings support a pronounced difficulty with singing on the syllable /la/ (Dalla Bella et al., 2009), but such difficulty appears to be mitigated in some amusics when singing after a model or in unison with a model as opposed to singing from memory (Tremblay-Champoux et al., 2010). In the general adult population, singing from memory and imitating melodies at a slow tempo have been associated with better performance when lyrics are removed (Berkowska & Dalla Bella, 2009), but the opposite pattern has been also shown for song and speech imitation, with linguistic information facilitating performance (Mantell & Pfordresher, 2013). It should be noted that such advantage does not hold when the individual is presented with lyrics but is required to imitate a target by replacing lyrics with a single vowel, which suggests that when a phonetic advantage is seen, it does not necessarily have a perceptual basis (Mantell & Pfordresher, 2013). It is of special note that in the current study, participants imitated a singer who also sang without lyrics. Hence, no mental transformation of lyrics into the syllable /la/ was needed. The evidence on the effect of lexical content on singing earlier in development is also inconclusive, as studies have pointed to largely mixed findings (Hanna, 1999; Levinowitz et al., 1998; Yarbrough, Green, Benson, & Bowers, 1991). Further research is warranted to determine the exact conditions under which lexical information affects singing and imitation in both amusics and neurotypicals.

The current study demonstrates that deviating from target duration and relative time matching patterns as well as making many time errors is more often than not the case in congenital amusia regardless of whether the content to be imitated is more or less familiar. The present findings substantiate the results of the timing analysis in Liu et al. (2013), which also showed compromised amusic performance across different measures during speech and song imitation. On a broader level, previous research in the literature suggests that about at least one in two amusics have difficulties with rhythm perception and production (Ayotte et al., 2002; Lagrois & Peretz, 2019; Peretz et al., 2003, but see Tremblay-Champoux et al., 2010 for some evidence suggesting intact performance in their sample). When singing a familiar song from memory, a few (less than half) amusics also demonstrated various degrees of rhythmic difficulties across different measures (Dalla Bella et al., 2009). Given the possible dissociation as well as association between pitch and timing processing during music production (Berkowska & Dalla Bella, 2009), more research is needed to elucidate the relationship between pitch and rhythm impairments in amusia.

What our study has shown is that amusics can benefit from familiar melody and their performance can be comparable to controls in the relative pitch domain, but not in the absolute pitch or timing domain. This result is consistent with the important role that interval and contour play in memory for melodies (Dowling, 1978; Dowling & Fujitani, 1971). We cannot rule out structural differences or other features across the two songs having an effect on imitation differences. That is, it may not be the familiarity of the target *per se*, but other properties of the Happy Birthday song that make it easier to imitate. What we are in a position to know is that more syllables and larger intervals in a tune affect imitation performance on a number of pitch and rhythm production measures (Liu et al., 2013). The order of presentation (novel song followed by familiar song trials) may have had a practice effect leading to better performance on the familiar song condition. However, given that amusics can better detect out-of-key notes in familiar melodies (Ayotte et al., 2002) and distinguish familiar from unfamiliar melodies (Tillmann et al., 2004), the scenario of long-term memory traces affecting performance is perhaps more likely than that of a practice effect from a single session. Future work employing scrambled and unscrambled versions of the Happy Birthday song or other familiar songs with a varied order of presentation can determine with certainty whether familiarity *per se* drives differences in performance.

In conclusion, our study adds to previous work pointing to pitch and rhythm difficulties in amusics but also reveals that melody familiarity partially facilitates song imitation in amusia, whereas the presence of lyrics may not play a significant role. Given the paucity of research on pitch and rhythm imitation in amusia, future research is warranted to corroborate the evidence on melody familiarity and the presence of lyrics presented in this study. Further experimental investigations are also needed to establish the exact effects of memory and structural characteristics of the to-be-imitated song stimuli, employing target melodies of varying pitch and rhythm parameters.

References

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P. E., ... & Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain*, *136*, 1639-1661.

Albouy, P., Peretz, I., Bermudez, P., Zatorre, R. J., Tillmann, B., & Caclin, A. (2019). Specialized neural dynamics for verbal and tonal memory: fMRI evidence in congenital amusia. *Human Brain Mapping*, *40*, 855-867.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, *125*, 238-251.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Berkowska, M., & Dalla Bella, S. (2009). Reducing linguistic information enhances singing proficiency in occasional singers. *Annals of the New York Academy of Sciences*, 1169, 108-111.

Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer (Version 6.0.50) [Computer program]. Retrieved from www.praat.org

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*, 2515245920960351.

Creel, S. C. (2019). The familiar-melody advantage in auditory perceptual development: Parallels between spoken language acquisition and general auditory perception. *Attention, Perception, & Psychophysics*, *81*, 948-957.

Dalla Bella, S., Giguère, J. F., & Peretz, I. (2007). Singing proficiency in the general population. *The journal of the Acoustical Society of America*, *121*, 1182-1189.

Dalla Bella, S., Giguère, J. F., & Peretz, I. (2009). Singing in congenital amusia. *The Journal of the Acoustical Society of America*, *126*, 414-424.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*, 341–354.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, Interval, and Pitch Recognition in Memory for Melodies. *The Journal of the Acoustical Society of America*, *49*, 524–531.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain*, *127*, 801-810.

Foxton, J. M., Nandy, R. K., & Griffiths, T. D. (2006). Rhythm deficits in 'tone deafness'. *Brain and Cognition*, *62*, 24-29.

Goodale, M.A., & Milner,A.D. (1992).Separate visual pathways for perception and action. *Trends in Neurosciences*, 15, 20-25.

Gosselin, N., Jolicœur, P., & Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*, 270–272. Gosselin, N., Paquette, S., & Peretz, I. (2015). Sensitivity to musical emotions in congenital amusia. *Cortex*, *71*, 171–182.

Graves, J. E., Pralus, A., Fornoni, L., Oxenham, A. J., Caclin, A., & Tillmann, B. (2019). Short- and long-term memory for pitch and non-pitch contours: Insights from congenital amusia. *Brain and Cognition*, *136*, 103614.

Greenspon, E. B., Pfordresher, P. Q., & Halpern, A. R. (2017). Pitch imitation ability in mental transformations of melodies. *Music Perception: An Interdisciplinary Journal*, *34*, 585-604.

Griffiths, T. D. (2008). Sensory systems: Auditory action streams? *Current Biology: CB*, *18*, R387-388.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, *9*, 697-704.

Hanna, N. J. (1999). *The effects of range of notes, gender, and songs with text or without text on the vocal accuracy of first, second and third grade students*. (Unpublished master's thesis). Glassboro, NJ: Rowan University.

Hutchins, S., Zarate, J. M., Zatorre, R. J., & Peretz, I. (2010). An acoustical study of vocal pitch matching in congenital amusia. *The Journal of the Acoustical Society of America*, *127*, 504-512.

Hutchins, S., & Moreno, S. (2013). The Linked Dual Representation model of vocal perception and production. *Frontiers in Psychology*, *4*, 825.

Hutchins, S., & Peretz, I. (2012). Amusics can imitate what they cannot discriminate. *Brain and Language*, *123*, 234–239.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, *15*, 356-360.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital amusia: when less is better than more. *Journal of Neuroscience*, *27*, 13028-13032.

Hyde, K. L., Zatorre, R. J., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex*, *21*, 292-299.

Jagiello, R., Pomper, U., Yoneya, M., Zhao, S., & Chait, M. (2019). Rapid Brain Responses to familiar vs. Unfamiliar Music–an EEG and pupillometry study. *Scientific Reports*, *9*, 1-13.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia*, *48*, 2630-2639.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2012). Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. *Memory & Cognition*, *40*, 1109-1121.

Jiang, C., Lim, V. K., Wang, H., & Hamm, J. P. (2013). Difficulties with pitch discrimination influences pitch memory performance: evidence from congenital amusia. *PLoS One*, 8, e79216.

Jiang, J., Liu, F., Zhou, L., & Jiang, C. (2019). The neural basis for understanding imitation-induced musical meaning: The role of the human mirror system. *Behavioural Brain Research*, *359*, 362-369.

Kalmus, H., & Fry, D. B. (1980). On tune deafness (dysmelodia): frequency, development, genetics and musical background. *Annals of Human Genetics*, *43*, 369-382.

Kraemer, D. J., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Sound of silence activates auditory cortex. *Nature*, *434*, 158-158.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).

Lagrois, M. É., & Peretz, I. (2019). The co-occurrence of pitch and rhythm disorders in congenital amusia. *Cortex*, *113*, 229-238.

Lenth, R. V. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://cran.r-project.org/package=emmeans

Lévêque, Y., Teyssier, P., Bouchet, P., Bigand, E., Caclin, A., & Tillmann, B. (2018). Musical emotions in congenital amusia: Impaired recognition, but preserved emotional intensity. *Neuropsychology*, *32*(7), 880–894.

Levinowitz, L. M., Barnes, P., Guerrini, S., Clement, M., D'April, P., & Morey, M. J. (1998). Measuring singing voice development in the elementary general music classroom. *Journal of Research in Music Education*, *46*, 35-47.

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain*, *133*, 1682-1693.

Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., & Stewart, L. (2012). The mechanism of speech processing in congenital amusia: Evidence from Mandarin speakers. *PLos one*, *7*(2).

Liu, F., Xu, Y., Patel, A. D., Francart, T., & Jiang, C. (2012). Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: Evidence from Mandarin speakers. *Brain and Cognition*, *79*, 209-215.

Liu, F., Jiang, C., Pfordresher, P. Q., Mantell, J. T., Xu, Y., Yang, Y., & Stewart, L. (2013). Individuals with congenital amusia imitate pitches more accurately in singing than in speaking: implications for music and language processing. *Attention, Perception, & Psychophysics*, *75*, 1783-1798.

Liu, F., Chan, A. H., Ciocca, V., Roquet, C., Peretz, I., & Wong, P. C. (2016). Pitch perception and production in congenital amusia: Evidence from Cantonese speakers. *The Journal of the Acoustical Society of America*, *140*, 563-575.

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action–perception mismatch in tone-deafness. *Current Biology*, *18*, 331-332.

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

Loui, P. (2015). A dual-stream neuroanatomy of singing. *Music Perception: An Interdisciplinary Journal*, *32*, 232-241.

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in psychology*, *6*, 2.

Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition*, *127*, 177-202.

Musso, M., Weiller, C., Horn, A., Glauche, V., Umarova, R., Hennig, J., ... & Rijntjes, M. (2015). A single dual-stream framework for syntactic computations in music and language. *Neuroimage*, *117*, 267-283.

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: association with lexical tone agnosia. *Brain*, *133*, 2635-2642.

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and Emotions in the Brain: Familiarity Matters. *PLoS ONE*, *6*(11).

Peretz, I., Gaudreau, D., & Bonnel, A. M. (1998). Exposure effects on music preference and recognition. *Memory & Cognition*, *26*, 884-902.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: the Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58-75.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, *7*, 362-367.

Peretz, I., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain*, *132*, 1277-1286.

Peretz, I., & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics*, *25*, 625-630.

Pfeuty, M., & Peretz, I. (2010). Abnormal pitch—time interference in congenital amusia: Evidence from an implicit test. *Attention, Perception, & Psychophysics*, *72*, 763-774.

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

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics*, *71*, 1385-1398.

Pfordresher, P. Q., & Nolan, N. P. (2019). Testing Convergence between Singing and Music Perception Accuracy Using Two Standardized Measures. *Auditory Perception & Cognition*, 1-15.

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition*, *98*, 1-11.

Quiroga-Martinez, D. R., Tillmann, B., Brattico, E., Cholvy, F., Fornoni, L., Vuust, P., & Caclin, A. (2021). Listeners with congenital amusia are sensitive to context uncertainty in melodic sequences. *Neuropsychologia*, *158*(June), 107911. https://doi.org/10.1016/j.neuropsychologia.2021.107911

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Sammler, D., Grosbras, M. H., Anwander, A., Bestelmeyer, P. E., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, *25*, 3079-3085.

Schaal, N. K., Pfeifer, J., Krause, V., & Pollok, B. (2015). From amusic to musical?—Improving pitch memory in congenital amusia with transcranial alternating current stimulation. *Behavioural Brain Research*, *294*, 141-148.

Tan, Y. T., McPherson, G. E., Peretz, I., Berkovic, S. F., & Wilson, S. J. (2014). The genetic basis of music ability. *Frontiers in Psychology*, *5*, 658.

Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition*, *71*, 259-264.

Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., & Peretz, I. (2011a). Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Frontiers in psychology*, *2*, 120.

Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umilta, C., & Peretz, I. (2011b). Fine-grained pitch processing of music and speech in congenital amusia. *The Journal of the Acoustical Society of America*, *130*(6), 4089-4096.

Tillmann, B., Albouy, P., Caclin, A., & Bigand, E. (2014). Musical familiarity in congenital amusia: evidence from a gating paradigm. *Cortex*, *59*, 84-94.

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Research*, *1640*, 251-263.

Tremblay-Champoux, A., Dalla Bella, S., Phillips-Silver, J., Lebrun, M. A., & Peretz, I. (2010). Singing proficiency in congenital amusia: Imitation helps. *Cognitive Neuropsychology*, *27*, 463-476.

Williamson, V. J., McDonald, C., Deutsch, D., Griffiths, T. D., & Stewart, L. (2010). Faster decline of pitch memory over time in congenital amusia. *Advances in Cognitive Psychology*, *6*, 15.

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory*, 18, 657-669.

Wood, E. A., Rovetti, J., & Russo, F. A. (2020). Vocal-motor interference eliminates the memory advantage for vocal melodies. *Brain and Cognition*, *145*(September), 105622.

Xu, Y. (2013). ProsodyPro — A tool for large-scale systematic prosody analysis. Retrieved from http://www.homepages.ucl.ac.uk/~uclyyix/ProsodyPro/

Yarbrough, C., Green, G., Benson, W., & Bowers, J. (1991). Inaccurate singers: An exploratory study of variables affecting pitch-matching. *Bulletin of the Council for Research in Music Education*, 23-34.

Zhang, C., Shao, J., & Chen, S. (2018). Impaired perceptual normalization of lexical tones in Cantonese-speaking congenital amusics. *The Journal of the Acoustical Society of America*, 144, 634–647.

Author note

This work was supported by an ERC (European Research Council) Starting Grant, 678733, CAASD, to F.L. and C.J., and an ESRC (Economic and Social Research Council) Postdoctoral Fellowship, PTA-026-27-2480-A, to F.L. Correspondence concerning this article should be addressed to Fang Liu.

Electronic mail: f.liu@reading.ac.uk, phone: +44 (0)118 378 8122

Tables

Characteristics	Amusic	Control	χ^2/t	*p*
Sex	8 F, 5 M	9F, 4 M	0	1
Handedness	2 L, 11R	0L, 13R	0.54	.46
Age	24.08 (2.93)	24.69 (1.84)	0.64	.53
Education (in years)	16.62 (2.53)	17.92 (0.95)	1.74	.09
MBEA scale	16.92 (3.33)	27.00 (1.91)	9.46	<.001
MBEA contour	19.31 (2.90)	26.85 (1.72)	8.06	<.001
MBEA interval	18.69 (2.98)	26.38 (1.61)	8.18	<.001
MBEA rhythm	21.92 (4.54)	27.08 (1.71)	3.83	<.001
MBEA meter	19.54 (4.03)	26.31 (2.32)	5.24	<.001
MBEA memory	21.54 (4.48)	28.23 (2.01)	4.91	<.001
MBEA pitch composite	54.92 (6.97)	80.23 (3.59)	11.64	<.001
Pitch change detection threshold	.19 (.07)	.14 (.04)	-2.15	.04
Pitch direction discrimination threshold	.16 (.06)	.11 (.05)	-2.29	.03

Table 1 Participant characteristics for amusics and controls. F=female; M=male; R=right; L=left; MBEA = Montreal Battery of Evaluation of Amusia (Peretz et al., 2003). Scores in the second and third columns are means, with SDs in parentheses; MBEA subtest figures correspond to number of correct responses out of 30; the pitch composite score is the sum of the first three subtests; pitch thresholds are in semitones; the t value is the statistic of the Welch two-sample t test (two-tailed, df=24); chi2 tests were conducted for sex and handedness. The pitch change detection task required participants to identify which tone included a pitch glide among three flat/gliding tones, and the pitch direction discrimination task required them to report which of the three tones had a different direction compared to the other two (Liu et al., 2012; Liu et al., 2010).

Song phrases	Chinese characters, pinyin, and English translation	No of syllables
1	我曾有个梦想 [W**o** c**eng** y**ou** g**e** m**eng** x**iang**] I once had a dream	6
2	去到海边挖海龟 [Q**u** d**ao** h**ai** b**ian** w**a** h**ai** g**ui**] About going to the beach to dig sea turtles	7
3	金色的沙滩上[J**in** s**e** d**e** sh**a** t**an** sh**ang**] On the golden sandy beach	6
4	爬着一群群的海 [P**a** zh**e** y**i** q**un** q**un** d**e** h**ai** g**ui**] Groups of sea turtles crawled	8
5	翻不过来呀[F**an** b**u** g**uo** l**ai** y**a**] They could not turn over	5
6	就去帮帮它[J**iu** q**u** b**ang** b**ang** t**a**] So I went to help them	5
7	不让太阳晒到肚皮[B**u** r**ang** t**ai** y**ang** sh**ai** d**ao** d**u** p**i**] To prevent the sun from shining on their bellies	8
8	光着小脚丫[G**uang** zh**e** x**iao** j**iao** y**a**] We walked barefoot	5
9	踩着小浪花[C**ai** zh**e** x**iao** l**ang** h**ua**] We stepped on little waves	5
10	我们多快乐呀[W**o** m**en** d**uo** k**uai** l**e** **a**] We were so happy	6
11	海龟[H**ai** g**ui**] Sea turtles	2
12	宽厚的脊背[K**uan** h**ou** d**e** j**i** b**ei**] With broad shoulders	5
13	海龟[H**ai** g**ui**]	2

	Sea turtles	
14	多远也不累[D**uo** **yuan** **ye** b**u** l**ei**] They were not tired regardless of how far they crawled	5
15	海龟[H**ai** g**ui**] Sea turtles	2
16	慢也无所谓[M**an** **ye** w**u** s**uo** w**ei**] It’s okay to be slow	5
17	海龟[H**ai** g**ui**] Sea turtles	2

Table 1 Song phrases (segments) with Chinese characters, pinyin (the rhymes are in bold font), English translation, and number of syllables.

Song	Mode	Average interval size (semitones)	Mode interval size (semitones)	Largest interval size (semitones)	Mean notes duration (msec)	Mean Interonset interval (msec)	Tempo (bpm)
Novel	Major	2.61	2	10	375	436	111
Familiar	Major	2.57	2	12	415	500	114

Table 3 Song characteristics for the novel and the familiar songs based on the lyrics condition.

		Absolute Pitch Deviation	Pitch Interval Deviation	Number of Pitch Interval Errors	Number of Contour Errors	Duration Difference	IOI Difference	Number of Time Errors	Signed Intervals
Pitch Composite	Correlation	-.16	-.20	-.16	.45	-.44	-.40	-.31	.07
	Sig.	.63	.55	.63	.15	.17	.11	.34	.83
Scale	Correlation	-.18	-.28	-.20	.26	-.41	-.44	-.34	.12
	Sig.	.58	.40	.55	.46	.21	.17	.29	.71
Contour	Correlation	-.03	-.23	-.22	.37	-.12	-.12	-.006	.09
	Sig.	.92	.49	.51	.25	.71	.72	.98	.78
Interval	Correlation	-.13	.04	.02	.42	-.44	-.53	-.32	-.04
	Sig.	.69	.90	.93	.19	.17	.08	.33	.89
Rhythm	Correlation	.01	.18	.17	**.61***	-.17	-.14	-.07	-.24
	Sig.	.95	.57	.60	**.04**	.60	.66	.82	.47
Meter	Correlation	.02	.13	.22	.25	**-.74****	**-.64***	**-.63***	.21
	Sig.	.93	.69	.50	.45	**.009**	**.03**	**.03**	.51
Memory	Correlation	-.57	-.45	-.40	.12	-.09	-.17	.06	.14
	Sig.	.06	.15	.21	.72	.78	.61	.84	.66
Pitch Change Detection	Correlation	-.02	.34	.51	-.13	.09	.03	.04	.58
	Sig.	.93	.30	.10	.68	.77	.92	.89	.05
Pitch Direction Discrimination	Correlation	-.02	-.3	.51	-.13	.09	.03	.04	.58
	Sig.	.93	.30	.10	.68	.77	.92	.89	.05

Table 4 Significant and non-significant Pearson correlation pairs with 9 degrees of freedom between perceptual scores and imitation measures in amusics for the familiar song condition. Note that none of these correlations remained significant after Holm's correction for multiple correlations.

		Absolute Pitch Deviation	Pitch Interval Deviation	Number of Pitch Interval Errors	Number of Contour Errors	Duration Difference	IOI Difference	Number of Time Errors	Signed Intervals
Pitch Composite	Correlation	-.28	-.49	**-.62***	**-.81****	**-.87****	**-.83****	**-.75****	.19
	Sig.	.40	.12	**.04**	**.002**	**< .001**	**.001**	**.008**	.56
Scale	Correlation	-.36	**-.70***	**-.73***	**-.80****	**-.82****	**-.61***	**-.96****	.30
	Sig.	.27	**.01**	**.01**	**.003**	**.002**	**.04**	**< .001**	.35
Contour	Correlation	-.06	-.20	-.35	-.43	-.35	-.49	-.03	.17
	Sig.	.86	.54	.29	.17	.28	.12	.92	.61
Interval	Correlation	-.18	-.17	-.31	-.60	**-.76****	**-.80****	-.59	-.03
	Sig.	.59	.60	.35	.05	**.006**	**.003**	.05	.90
Rhythm	Correlation	-.01	.04	-.05	-.37	-.47	**-.66***	-.33	-.14
	Sig.	.96	.88	.87	.26	.14	**.02**	.30	.67
Meter	Correlation	-.06	-.40	-.36	**-.63***	**-.86****	-.19	**-.81****	.32
	Sig.	.85	.22	.26	**.09**	**< .001**	.57	**.002**	.32
Memory	Correlation	-.57	-.40	**-.64***	**-.63***	-.49	-.60	-.45	.07
	Sig.	.06	.21	**.03**	**.03**	.12	.05	.21	.82
Pitch Change Detection	Correlation	-.06	-.11	.12	.03	.23	.25	.32	.38
	Sig.	.85	.73	.71	.91	.49	.44	.32	.24
Pitch Direction Discrimination	Correlation	.42	.09	.36	.18	.05	.35	.01	.07
	Sig.	.19	.78	.26	.58	.87	.29	.96	.83

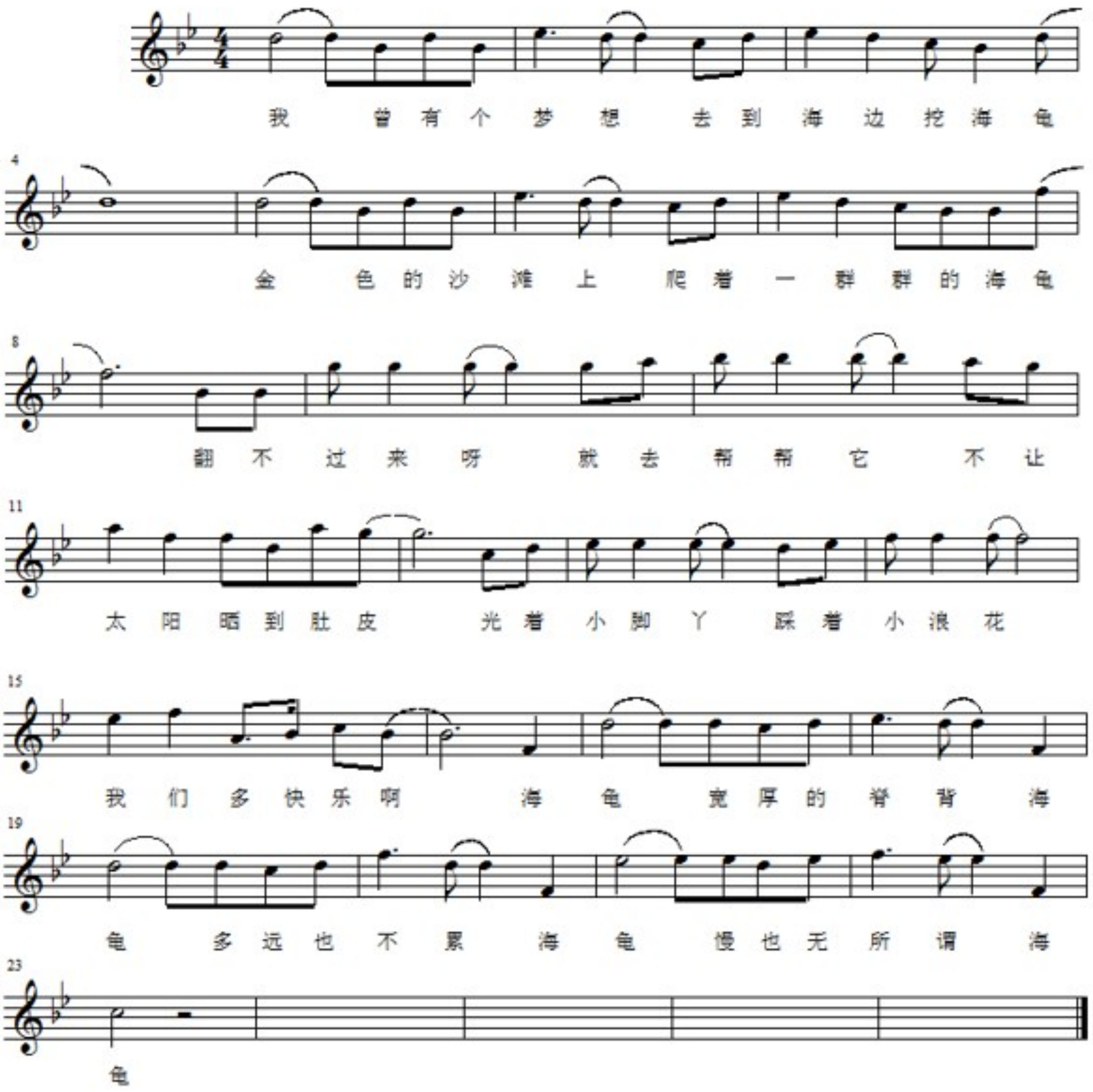
Table 5 Pearson correlation pairs with 11 degrees of freedom between perception and production measures in amusics for the novel song condition. The correlations between scale and number of time errors and pitch composite and duration difference remained significant following Holm's correction for multiple correlations.

		Absolute Pitch Deviation	Pitch Interval Deviation	Number of Pitch Interval Errors	Number of Contour Errors	Duration Difference	IOI Difference	Number of Time Errors	Signed Intervals
Pitch Composite	Correlation	-.53	-.49	-.47	-.21	-.37	.40	-.19	-.01
	Sig.	.08	.12	.14	.53	.26	.21	.56	.95
Scale	Correlation	-.15	-.25	.04	-.18	.25	.54	.29	.24
	Sig.	.66	.44	.90	.58	.44	.08	.37	.47
Contour	Correlation	-.35	-.39	**-.65***	.02	-.41	.05	-.10	-.05
	Sig.	.29	.22	**.03**	.93	.20	.87	.75	.86
Interval	Correlation	**-.65***	-.38	-.42	-.27	**-.70***	.18	**-.68***	-.26
	Sig.	**.03**	.24	.18	.41	**.01**	.58	**.02**	.42
Rhythm	Correlation	.26	.04	-.11	-.27	-.23	.26	-.02	-.08
	Sig.	.43	.90	.74	.41	.49	.43	.93	.81
Meter	Correlation	.38	.18	.00	-.37	-.13	.20	.13	-.22
	Sig.	.24	.59	1.00	.25	.69	.54	.69	.50
Memory	Correlation	-.21	-.17	-.19	-.49	-.52	.39	-.58	-.30
	Sig.	.52	.59	.57	.12	.09	.22	.06	.35
Pitch Change Detection	Correlation	-.32	-.07	-.27	**.69***	-.15	-.26	-.08	.40
	Sig.	.32	.82	.41	**.02**	.64	.42	.81	.22
Pitch Direction Discrimination	Correlation	-.27	.002	-.14	.50	-.03	-.06	-.01	.06
	Sig.	.42	.99	.68	.11	.93	.84	.95	.84

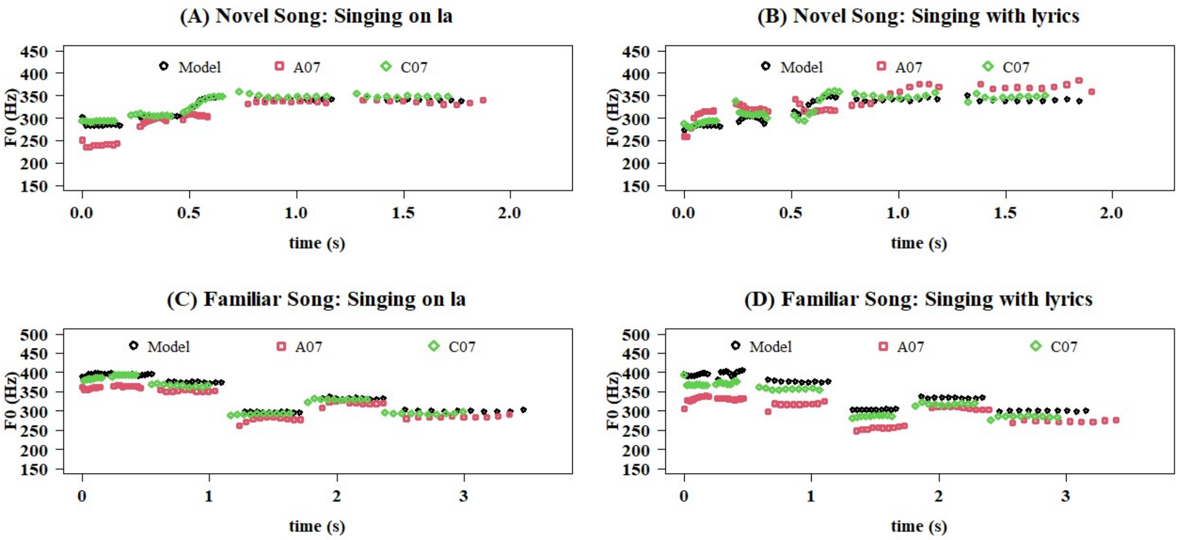
Table 6 All Pearson correlations between perception and production measures in control participants for the familiar song condition with 9 degrees of freedom. None of these correlations remained significant after adjustment using Holm's correction.

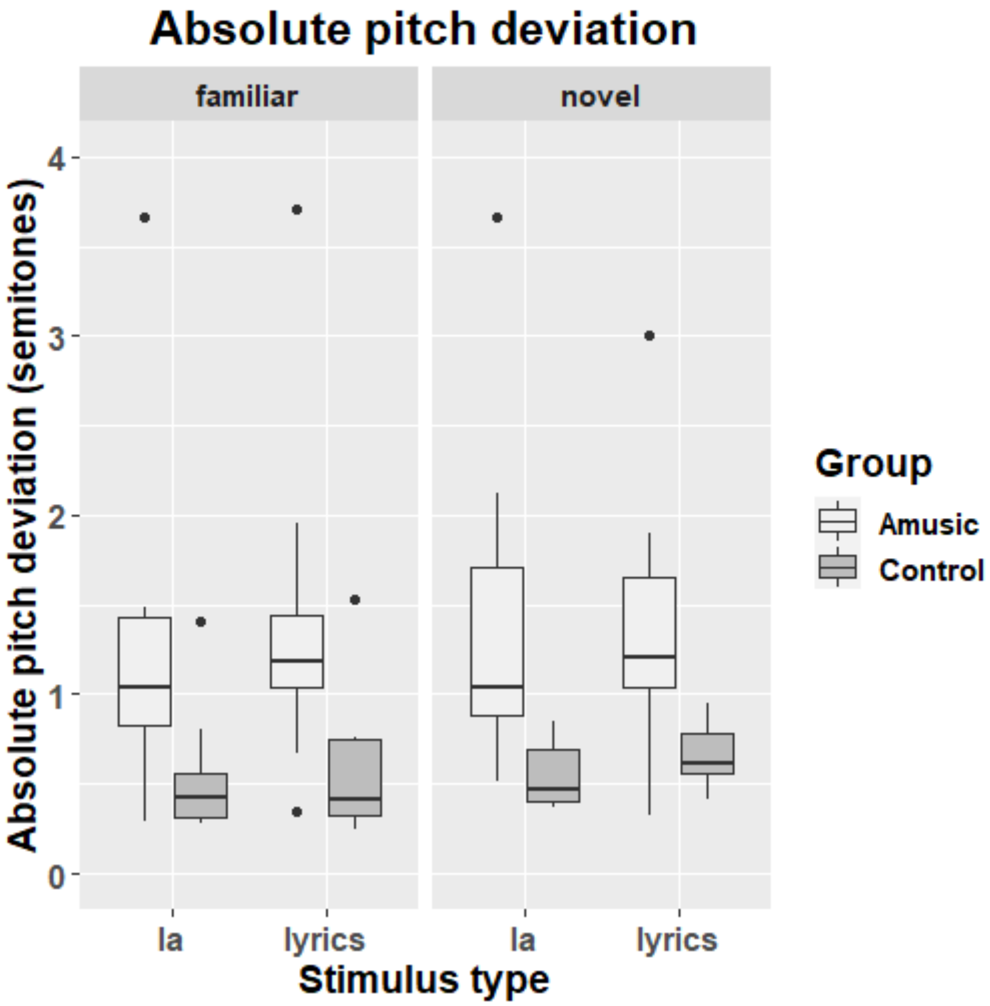
		Absolute Pitch Deviation	Pitch Interval Deviation	Number of Pitch Interval Errors	Number of Contour Errors	Duration Difference	IOI Difference	Number of Time Errors	Signed Intervals
Pitch Composite	Correlation	-.45	-.36	-.24	.35	.10	.12	.08	-.001
	Sig.	.16	.26	.46	.28	.77	.71	.81	.99
Scale	Correlation	-.27	-.18	-.14	.43	.06	.04	-.02	-.08
	Sig.	.42	.59	.67	.18	.84	.90	.94	.80
Contour	Correlation	-.49	-.44	-.43	-.07	-.08	-.01	.03	.23
	Sig.	.11	.16	.18	.83	.81	.97	.92	.49
Interval	Correlation	-.16	-.13	.06	.34	.22	.23	.17	-.14
	Sig.	.62	.69	.84	.30	.50	.48	.61	.67
Rhythm	Correlation	.03	.16	-.33	-.07	-.29	-.21	-.34	-.24
	Sig.	.91	.62	.31	.83	.38	.53	.30	.46
Meter	Correlation	.19	.30	-.24	-.19	-.01	-.12	-.32	-.41
	Sig.	.57	.36	.46	.57	.96	.71	.33	.20
Memory	Correlation	.05	.12	.08	.45	-.14	-.05	-.06	-.22
	Sig.	.87	.71	.80	.16	.66	.86	.85	.51
Pitch Change Detection	Correlation	-.15	-.21	.01	-.40	.22	.38	.51	.55
	Sig.	.65	.52	.97	.22	.50	.23	.10	.07
Pitch Direction Discrimination	Correlation	-.22	-.29	.01	-.001	-.14	-.02	.30	.56
	Sig.	.50	.38	.96	.99	.66	.94	.36	.06

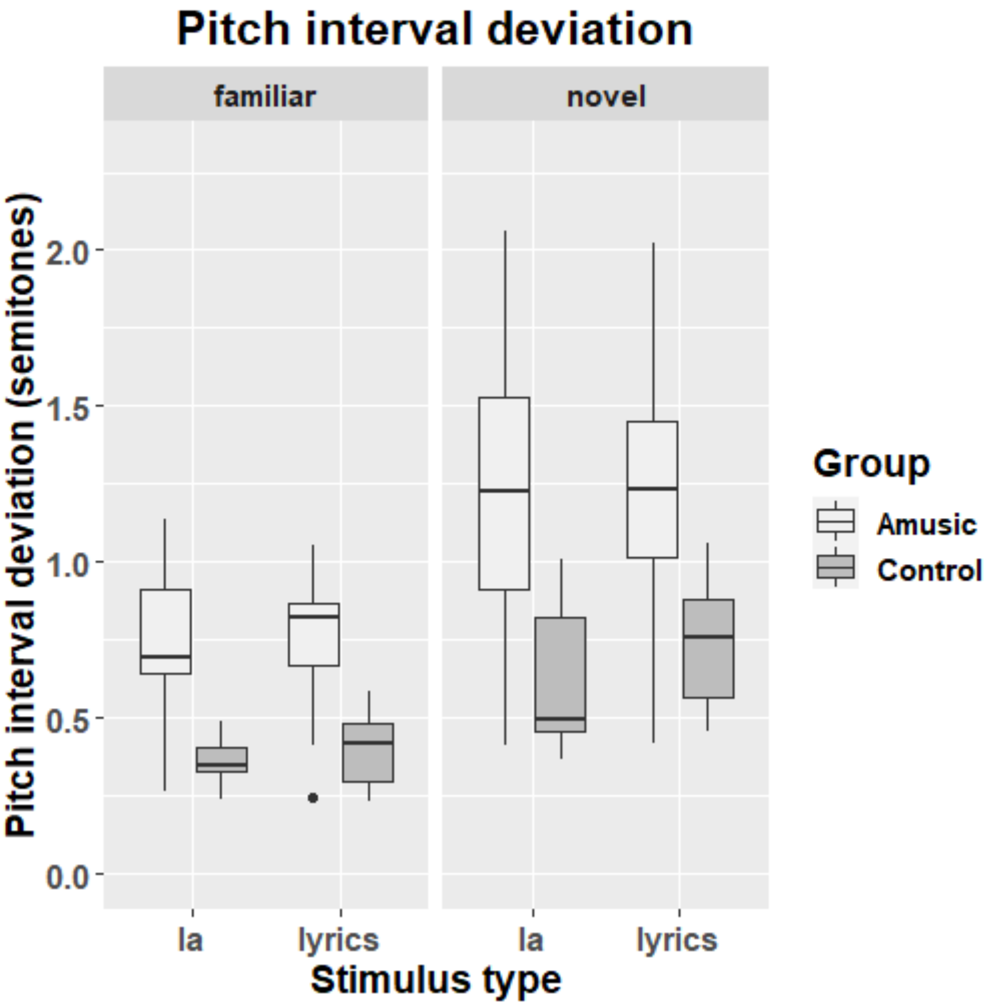
Table 7 Pearson correlation pairs between perception and production measures (df = 11) in controls for the novel song condition.

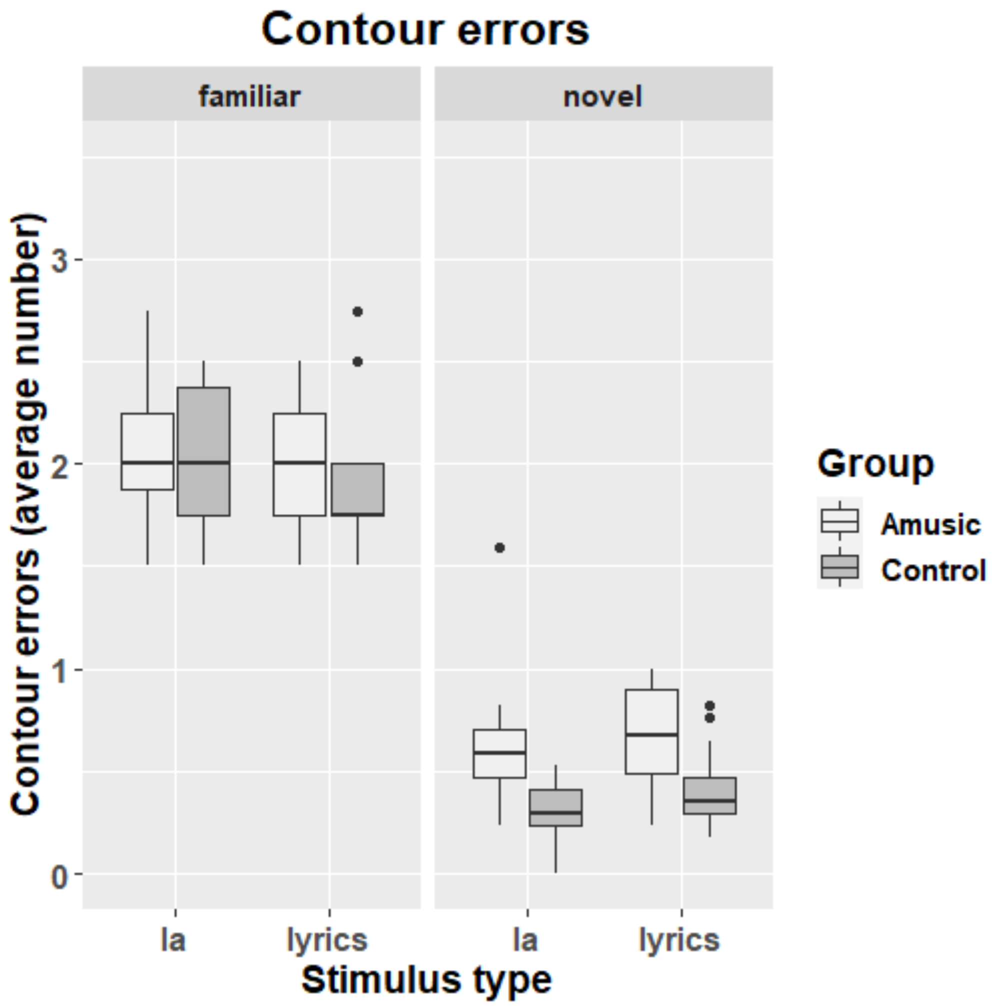
Figures


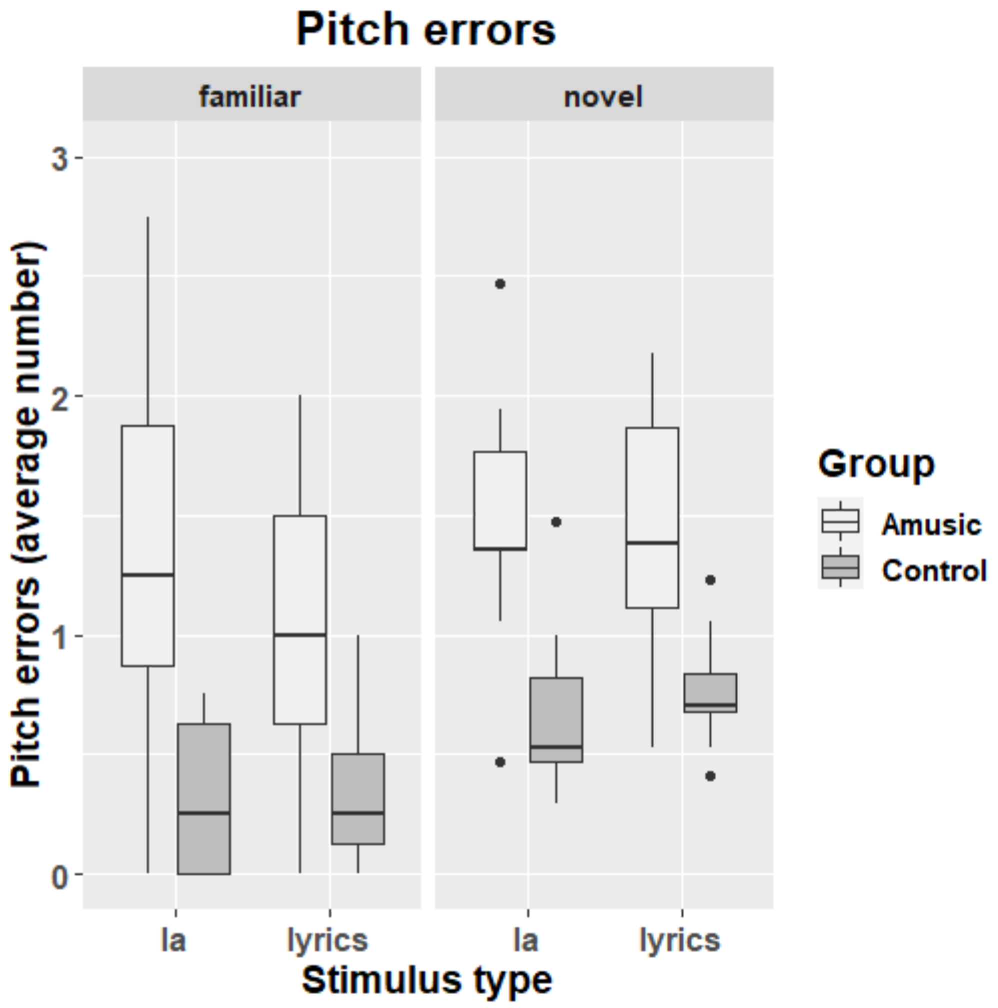
我 曾 有 个 梦 想 去 到 海 边 挖 海 龟
4
金 色 的 沙 滩 上 爬 着 一 群 群 的 海 龟
8
翻 不 过 来 呀 就 去 帮 帮 它 不 让
11
太 阳 晒 到 肚 皮 光 着 小 脚 丫 踩 着 小 浪 花
15
我 们 多 快 乐 啊 海 龟 宽 厚 的 脊 背 海
19
龟 多 远 也 不 累 海 龟 慢 也 无 所 谓 海
23
龟

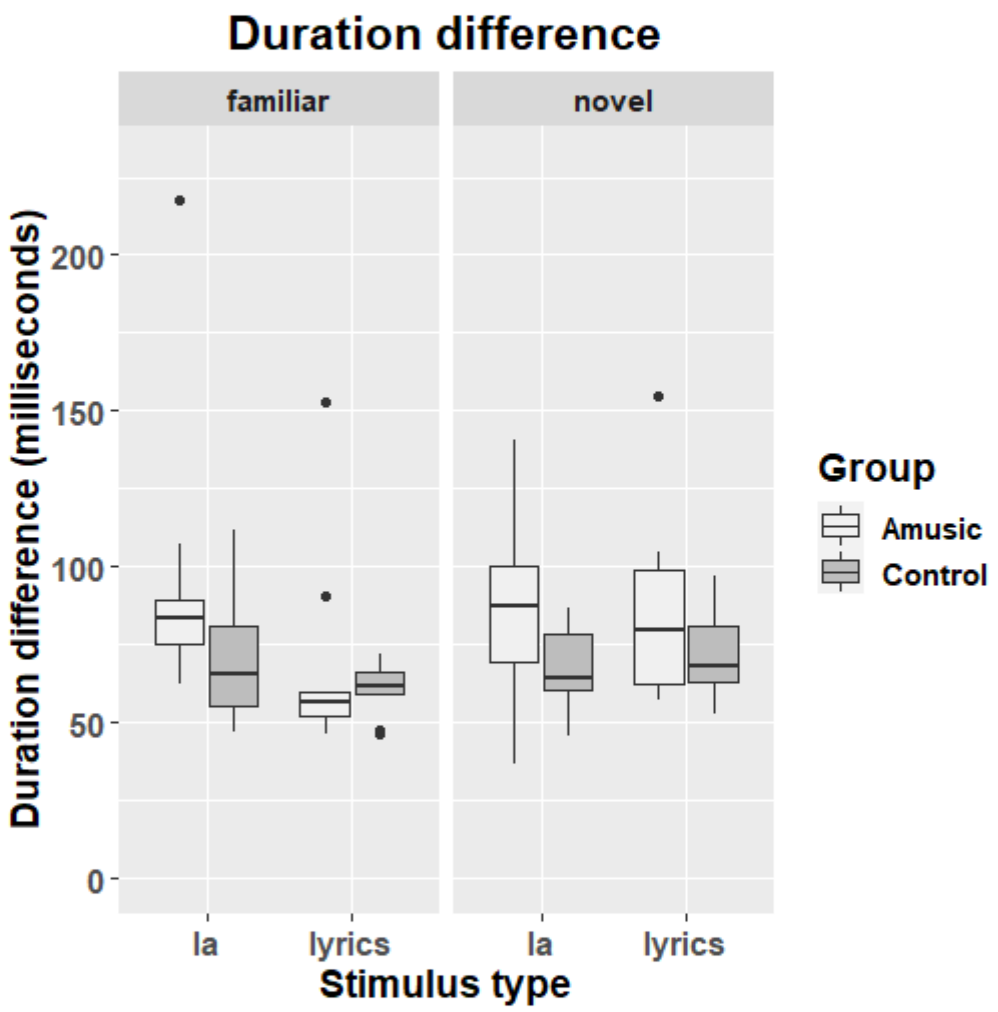

(A) Novel Song: Singing on la
(B) Novel Song: Singing with lyrics
(C) Familiar Song: Singing on la
(D) Familiar Song: Singing with lyrics
Model
A07
C07
F0 (Hz)
time (s)
150
200
250
300
350
400
450
500
0.0
0.5
1.0
1.5
2.0
0
1
2
3

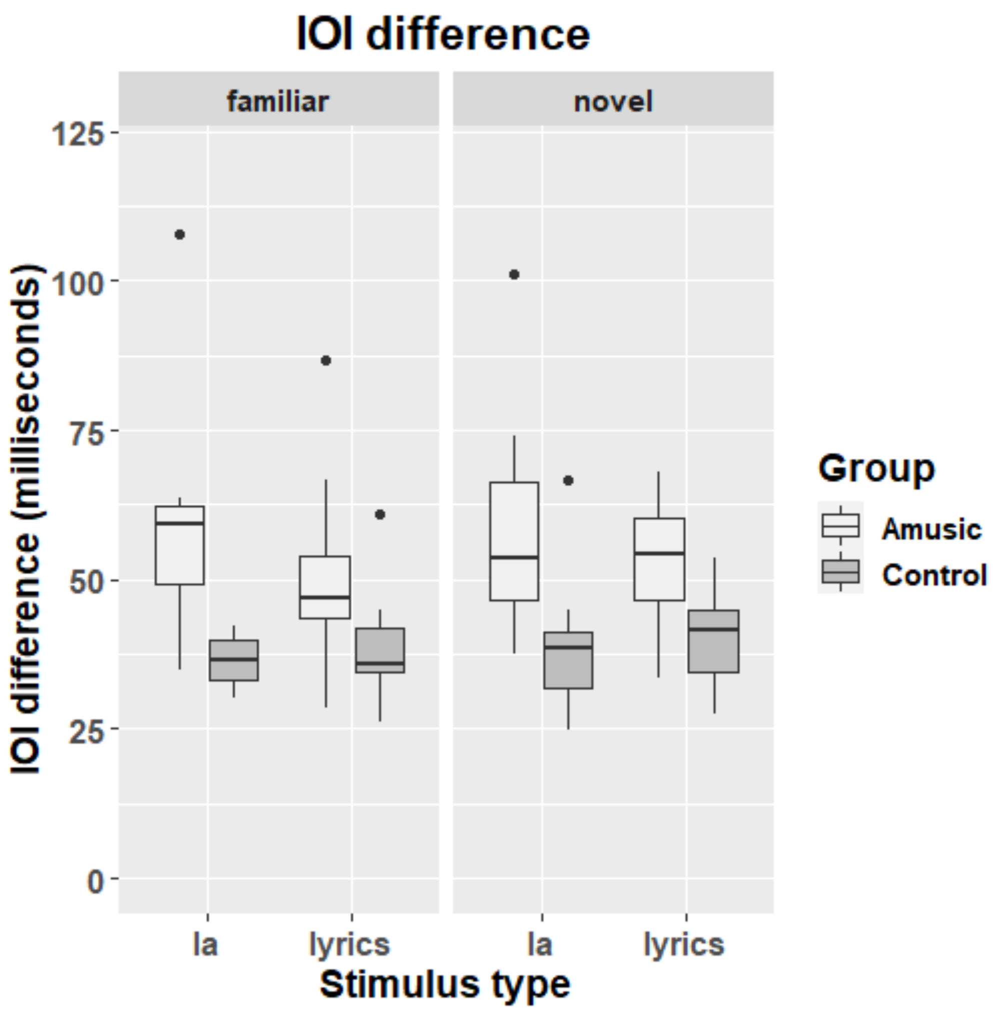

Absolute pitch deviation
familiar
novel
Absolute pitch deviation (semitones)
0
1
2
3
4
la
lyrics
la
lyrics
Stimulus type
Group
Amusic
Control

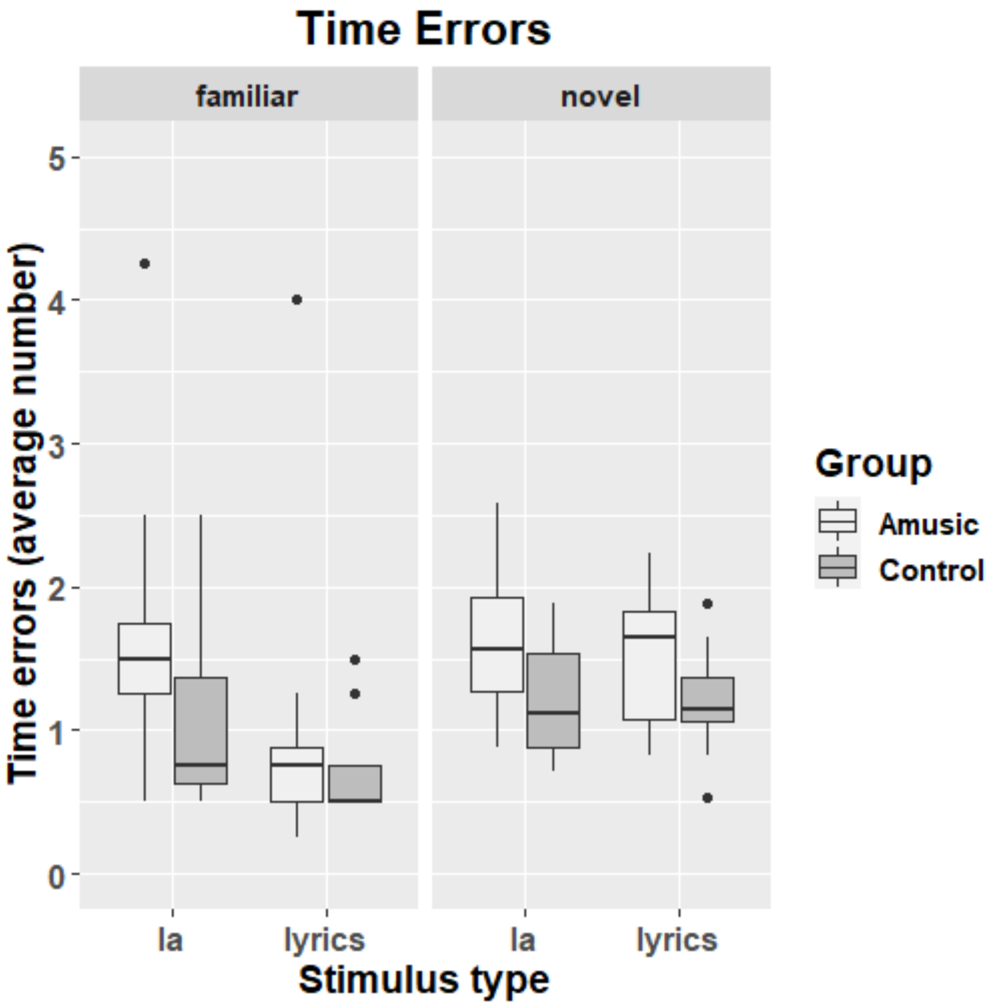

Pitch interval deviation
familiar
novel
Pitch interval deviation (semitones)
2.0
1.5
1.0
0.5
0.0
la
lyrics
la
lyrics
Stimulus type
Group
Amusic
Control


Contour errors
familiar
novel
Contour errors (average number)
0
1
2
3
la
lyrics
la
lyrics
Stimulus type
Group
Amusic
Control


Pitch errors
familiar
novel
Pitch errors (average number)
3
2
1
0
la
lyrics
la
lyrics
Stimulus type
Group
Amusic
Control


Duration difference
familiar
novel
Duration difference (milliseconds)
0
50
100
150
200
la
lyrics
la
lyrics
Stimulus type
Group
Amusic
Control


IOI difference
familiar
novel
IOI difference (milliseconds)
125
100
75
50
25
0
la
lyrics
la
lyrics
Stimulus type
Group
Amusic
Control


Time Errors
familiar
novel
Time errors (average number)
0
1
2
3
4
5
la
lyrics
la
lyrics
Stimulus type
Group
Amusic
Control

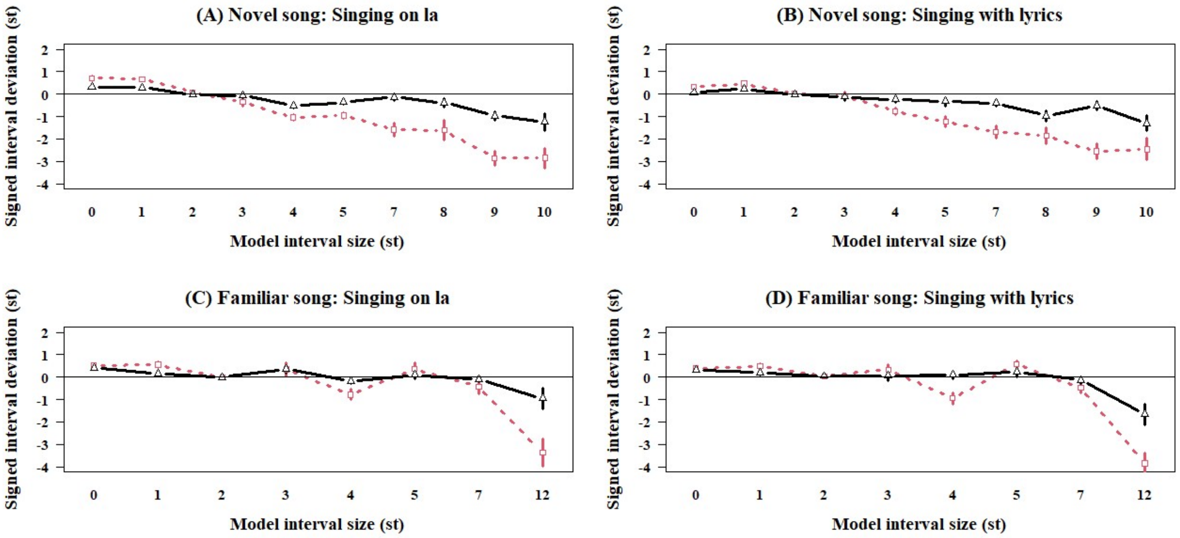

(A) Novel song: Singing on la
(B) Novel song: Singing with lyrics
(C) Familiar song: Singing on la
(D) Familiar song: Singing with lyrics
Signed interval deviation (st)
2
1
0
-1
-2
-3
-4
Model interval size (st)
0
1
2
3
4
5
7
8
9
10
12

Figure Captions

Fig. 1. Musical notation and lyrics for the original song "sea turtles". Note that the model sang one octave lower than the score.

Fig. 2. Pitch contours (in semitones) of target stimuli and imitation performance of one amusic participant (A07) and one control participant (C07). The target appears in black dots. Red squares represent the imitation of the amusic participant and green diamonds that of the control participant.

Fig. 3. Performance of amusic participants and controls on absolute pitch for singing on /la/ and singing with lyrics across the familiar song and the novel song conditions. The boxplot shows the median and the quartiles of the sample. Higher scores indicate a less accurate imitation.

Fig. 4. Performance of amusics and controls on the pitch interval measure for singing on the syllable /la/ and with lyrics across the familiar song and the novel song conditions, with higher scores reflecting a less accurate performance. The whisker boxes show the median and the quartiles.

Fig. 5. Average number of contour errors in amusics and controls for singing on the syllable /la/ and singing with lyrics across the familiar and the novel song conditions. The boxplot shows the median and the quartiles of the sample.

Fig. 6. Average number of pitch interval errors in amusics and controls for singing on the syllable /la/ and singing with lyrics across the familiar and novel song conditions. The boxplot shows the distribution of the data including the median and the quartiles.

Fig. 7. Duration difference (in milliseconds) from the model for amusic and control performance on singing on /la/ and singing with lyrics across the familiar and the novel song conditions. Larger values correspond to a larger deviation from the target. The whisker boxes show the median and the quartiles.

Fig. 8. Interonset interval (IOI) differences (in milliseconds) in amusics and controls for singing on /la/ and with lyrics across the familiar and the novel song conditions. The boxplot shows the median and the quartiles of the sample. Larger values correspond to a less accurate imitation.

Fig. 9. Box plots showing the distribution of time errors (along with the median and the quartiles) in amusics and controls across the familiar and novel song conditions.

Fig. 10. Signed pitch interval deviation from the model for the novel song condition (a and b) and the familiar song condition (c and d). The lines depict mean signed interval deviations and the error bars

represent standard error; controls' deviation from the model is shown in black straight lines and amusics' deviation appears in red dashed lines.

Supplementary Table

Characteristics	Amusic	Control	χ^2/t	*p*
Sex	7 F, 4 M	7F, 4 M	0	1
Handedness	2 L, 9R	0L, 11R	0.55	.45
Age	23.64 (2.20)	24.64 (1.96)	1.12	.27
Education (in years)	16.45 (2.50)	17.73 (0.90)	1.58	.13
MBEA scale	16.45 (3.42)	26.91 (2.07)	8.67	<.001
MBEA contour	18.91 (2.55)	27 (1.79)	8.62	<.001
MBEA interval	18.55 (3.17)	26.45 (1.75)	7.23	<.001
MBEA rhythm	21.36 (4.72)	27.36 (1.57)	12.17	.001
MBEA meter	19.27 (4.36)	26.64 (2.29)	4.95	<.001
MBEA memory	21.09 (4.21)	28.09 (2.12)	4.92	<.001
MBEA pitch composite	53.91 (7.05)	80.36 (3.85)	10.92	<.001
Pitch change detection threshold	0.20 (0.07)	0.14 (0.05)	2.25	.03
Pitch direction discrimination threshold	0.17 (0.07)	0.11 (0.05)	2.14	.04

Supplementary table. Participant characteristics for amusics and controls who participated in the familiar song condition. F=female; M=male; R=right; L=left; MBEA = Montreal Battery of Evaluation of Amusia (Peretz et al., 2003). Scores in the second and third columns are means, with SDs in parentheses; MBEA subtest figures correspond to number of correct responses out of 30; the pitch composite score is the sum of the first three subtests; pitch thresholds are in semitones; the t value is the statistic of the Welch two-sample t test (two-tailed, df=24); chi2 tests were conducted for sex and handedness. The pitch change detection task required participants to identify which tone included a pitch glide among three flat/gliding tones, and the pitch direction discrimination task required them to report which of the three tones had a different direction compared to the other two (Liu et al., 2012; Liu et al., 2010).