



The singing voice is special: Persistence of superior memory for vocal melodies despite vocal-motor distractions[☆]

Michael W. Weiss^{*}, Anne-Marie Bissonnette, Isabelle Peretz

International Laboratory for Brain, Music, and Sound Research (BRAMS), University of Montreal, Montreal, Quebec, Canada

ARTICLE INFO

Keywords:

Memory
Music
Singing
Articulatory suppression
Timbre

ABSTRACT

Vocal melodies sung without lyrics (*la la*) are remembered better than instrumental melodies. What causes the advantage? One possibility is that vocal music elicits subvocal imitation, which could promote enhanced motor representations of a melody. If this motor interpretation is correct, distracting the motor system during encoding should reduce the memory advantage for vocal over piano melodies. In Experiment 1, participants carried out movements of the mouth (i.e., chew gum) or hand (i.e., squeeze a beanbag) while listening to 24 unfamiliar folk melodies (half vocal, half piano). In a subsequent memory test, they rated the same melodies and 24 timbre-matched foils from ‘1–Definitely New’ to ‘7–Definitely Old’. There was a memory advantage for vocal over piano melodies with no effect of group and no interaction. In Experiment 2, participants carried out motor activities during encoding more closely related to singing, either silently articulating (*la la*) or vocalizing without articulating (humming continuously). Once again, there was a significant advantage for vocal melodies with no effect or interaction of group. In Experiment 3, participants audibly whispered (*la la*) repeatedly during encoding. Again, the voice advantage was present and did not differ appreciably from prior research with no motor task during encoding. However, we observed that the spontaneous phase-locking of whisper rate and musical beat tended to predict enhanced memory for vocal melodies. Altogether the results challenge the notion that subvocal rehearsal of the melody drives enhanced memory for vocal melodies. Instead, the voice may enhance engagement.

1. Introduction

“In almost every instance mere muscular movement produces less of a disturbance in memory than does activity of the vocal mechanism. And the reason for this is evident: the articulatory innervations involved in memorising are most conspicuously interfered with by conflicting activity of the vocal mechanism.” W. G. Smith, *The relation of attention to memory*, 1895, p. 62.

This pre-modern observation of Smith has been largely supported by decades of research on memory for verbal material. In brief, processes of subvocal rehearsal that facilitate memory can be disrupted by asking participants to perform a vocal task such as repeating a syllable. The articulation impairs short-term memory (e.g., Larsen & Baddeley, 2003; Murray, 1967a, 1968; Salamé & Baddeley, 1982), pointing to subvocal rehearsal as a key mechanism in maintaining and encoding new verbal information (Baddeley, 2012). But why and how might motor

involvement facilitate cognition? These questions are pertinent to many specialties of research because they concern the relation between major systems (perception, cognition, action) and might connect disparate phenomena. For example, covert rehearsal can facilitate short-term memory for words (Murray, 1967b) and long-term memory for short melodies (Peynircioglu, 1995), and movements of the eye may ‘reinstate’ long-term spatiotemporal memories (Wynn, Shen, & Ryan, 2019). The implication is that motor representations function to support memory across modalities and timescales, and perhaps cognition generally. Recruitment of the motor system is central to controversial debates in cognitive science and neuroscience such as those surrounding ‘embodied cognition’ (Goldinger, Papesh, Barnhart, Hansen, & Hout, 2016) or ‘motor theories’ of speech (Liberman & Mattingly, 1985; Lotto, Hickok, & Holt, 2009). Debates such as these are informed by asking whether a targeted disruption of the motor system such as articulatory suppression will cause a targeted disruption of a cognitive process.

Indeed, articulatory suppression also interferes with memory for

[☆] This paper is a part of special issue “Special Issue in Honour of Jacques Mehler, Cognition’s founding editor”.

^{*} Corresponding author.

E-mail address: michael.weiss@umontreal.ca (M.W. Weiss).

nonverbal auditory material such as music. Singing or speaking a syllable repeatedly during the encoding or retention of a 4-note melody impairs the ability to recognize pitch changes as compared with no suppression (e.g., Schendel & Palmer, 2007). Similarly, singing a children's song covertly during the retention interval impairs participants' ability to detect changes in a 4-note song (Koelsch et al., 2009). Together these results imply that subvocal rehearsal is beneficial in the short-term maintenance of melodic information.

The role of subvocal rehearsal in long-term memory for music remains unexplored. We note, however, that subvocalization may account for the superiority effects observed for vocal material. Weiss and collaborators (Weiss & Peretz, 2019; Weiss, Schellenberg, Peng, & Trehub, 2019; Weiss, Schellenberg, & Trehub, 2017; Weiss, Schellenberg, Trehub, & Dawber, 2015; Weiss, Trehub, & Schellenberg, 2012; Weiss, Trehub, Schellenberg, & Habashi, 2016; Weiss, Vanzella, Schellenberg, & Trehub, 2015) have repeatedly shown that listeners age 7 and older recognize melodies sung by the voice (*la la*) more confidently than the same melodies played on instruments such as the piano. Even expert pianists show a voice advantage relative to piano melodies (Weiss, Schellenberg, et al., 2015). For most listeners—including pianists—the voice is more readily imitable than any instrument, hence, subvocalization may contribute to the 'voice advantage' in long-term memory.

If subvocalization is elicited by vocal rather than piano melodies then the voice advantage observed in memory should be attenuated or eliminated by concurrent articulation.

Two recent studies on earworms provide some support for a role of subvocalization in access to long-term representations for music. Campbell and Margulis (2015) investigated the role of overt and covert motor activity during listening to catchy music on the presence of eventual earworms. Participants were asked to listen without movement, or with movements that either involved the vocal system (humming, whistling, or singing) or the motor system (dancing, tapping, or nodding). An unexpected finding was that participants in the non-movement group found it difficult *not* to move to the music: nearly two thirds reported movement and more than a quarter reported vocalizing despite the instructions, forcing the researchers to analyze the data based on what participants did rather than what they were assigned to do. The results showed a relationship between movement and the incidence of earworms, with the highest incidence in those who moved and vocalized. Beaman, Powell, and Rapley (2015), taking a different approach, asked whether movements that inhibit subvocalization would decrease the incidence of earworms. Their novel and practical manipulation was inspired by an anonymous online commenter who reportedly chewed on cinnamon sticks to rid themselves of earworms. Across three experiments, the researchers found that chewing gum vigorously while listening to a catchy tune reduced the eventual occurrence of earworms. Moreover, their results could not be attributed to general dual-task demands because a tapping task yielded more earworms than the chewing task. Whether a manipulation such as chewing gum affects the encoding of *novel* melodies—vocal melodies in particular—is the current research question.

Here we report three experiments testing whether motor interference disrupts the voice advantage in memory for melodies. In all experiments, participants listened to a set of vocal and piano melodies during an encoding phase, followed by a short break and a recognition memory test (Weiss et al., 2012; Weiss et al., 2016; Weiss et al., 2017; Weiss et al., 2019; Weiss & Peretz, 2019; Weiss, Schellenberg, et al., 2015; Weiss, Vanzella, et al., 2015). During encoding, participants performed a motor activity while listening to the melodies. In the first experiment, we considered a manipulation from Beaman et al. (2015): chewing gum, with a control condition of squeezing a beanbag. In the second experiment we considered two manipulations that focused specifically on the muscles involved in singing: articulating '*la la*' silently or vocalizing '*omm*' continuously. In a final experiment, participants whispered '*la la*' out loud. If the voice advantage depends on covert singing, then covert and overt vocalization should interfere with the memory traces and

eliminate the voice advantage. Alternatively, if inner singing is not the driver of the voice advantage, then vocal-motor suppression should have no appreciable effect on memory.

2. Experiment 1

Previous research suggests that chewing gum can reduce the occurrence of earworms, or involuntary musical imagery, suggesting a connection between recollection and the motor system (Beaman et al., 2015). While chewing does not recruit the vocal system, it is a natural and fairly undemanding motor task that occupies muscles involved in articulation. The current experiment explored whether a similar manipulation would impair memory for vocal music more than for instrumental music, and hence reduce or eliminate the memory advantage for vocal melodies observed in previous research. Memory for vocal and instrumental (piano) melodies was tested after one of two encoding conditions: Listening to melodies while chewing gum (*Chewing group*) or squeezing a beanbag (*Squeezing Hand group*). If chewing gum mobilizes the articulatory system, then the voice advantage should be reduced or eliminated in the Chewing group only.

2.1. Material and methods

2.1.1. Participants

Thirty-eight participants were assigned at random to two groups with different motor tasks during the experiment involving muscles of the mouth ($n = 20$; 15 female; $M = 21.8 \pm 3.2$, $range = 19\text{--}31$ years) or the hand ($n = 18$; 13 female; $M = 22.2 \pm 2.8$, $range = 18\text{--}28$ years). Amount of formal musical training was similar between the groups (*Chewing*: $M = 6.1 \pm 4.2$, $median = 5.5$, $range = 0\text{--}16$ years; *Squeezing Hand*: $M = 5.3 \pm 6.0$, $median = 2.8$, $range = 0\text{--}21$ years). We considered musicianship as an additional between-participants factor (5+ years = musician; <5 years = nonmusician) in post-hoc analyses here and in Experiments 2 and 3, but in all instances there were no main effects or interactions involving musicianship and all other effects were unchanged. For simplicity musicianship is not considered further. All participants were healthy and had normal hearing by self-report. One additional participant was tested but excluded due to technical issues, and two additional participants were excluded¹ for ages more than three standard deviations beyond all participants tested. All participants received compensation for their time, in accordance with methods approved by the ethics committee at the University of Montreal.

2.1.2. Stimuli

The stimuli were 48 melody excerpts based on unfamiliar Irish and British folk tunes. The same melodies have been used in previous research (Weiss, Schellenberg, et al., 2015). Melodies differed in length ($range = 13.8\text{--}21.5$ s), number of notes ($range = 20\text{--}57$), tempo ($range = 70\text{--}130$ beats per minute), and time signature (3/4, 4/4, or 6/8). Two renditions of each melody were performed by amateur musicians in vocal (female, "*la la*") or instrumental (piano) timbres. Performances were recorded to a backing track that was later discarded to ensure consistency. Vocal recordings were pitch corrected note-by-note using the pitch center and pitch drift functions in Melodyne (Celemony, Inc.), which ensures overall accuracy without distorting natural pitch fluctuations or forsaking naturalness (Weiss et al., 2012). Recordings were RMS normalized to a common level using Sample Manager (Audiofile Engineering) and exported as high-quality audio files (16 bit / 44.1 kHz).

To ensure that melody memorability did not influence the manipulations of interest, melodies were randomly assigned to timbre and exposure level separately for each individual. Specifically, half of the

¹ Inclusion of participants excluded for age, here and in Experiment 2, did not alter the results.

melodies were assigned randomly as vocal ($n = 24$) and half as piano ($n = 24$). Then, half of the melodies in each timbre were assigned randomly to be learned during the first phase of the experiment ($n = 12$ per timbre; $n = 24$ total), with the remaining melodies serving as foils during the memory test ($n = 12$ per timbre; $n = 24$ total).

2.1.3. Apparatus

Participants were tested individually, in a double-walled sound-attenuating booth (Industrial Acoustics, Inc.). The experiment was programmed in Matlab (Mathworks) on a Windows PC. Stimuli were presented over high-quality speakers (Genelec) at a comfortable volume (~ 65 dB).

Compliance with the motor task was monitored using surface electromyography (EMG; Biopac, Inc.). For the Chewing group, electrodes were placed on the left masseter muscle (Criswell, 2011). For the Squeezing Hand group, electrodes were placed on the first dorsal interosseus muscle of the left hand (Criswell, 2011). EMG signals were excessively noisy or unavailable for 8 participants, but when available, the EMG data confirmed compliance, and all participants reported compliance verbally during debriefing.

2.1.4. Procedure

In the first phase of the experiment, participants listened to half of the melody set ($n = 24$ melodies; $n = 12$ voice and $n = 12$ piano) while completing a motor task with their mouth (chewing gum continuously; $n = 20$ participants) or their left hand (squeezing a beanbag repeatedly; $n = 18$ participants). To maintain attention, participants rated their liking of each melody on a scale from '1 – Dislike' to '5 – Like'. Trials were self-paced, with a short countdown preceding the onset of sound to alert the participant to begin the motor task. Participants were instructed to perform the motor task (i.e., chew/squeeze vigorously) for the entire duration of the melody and to relax between trials. No instructions were given regarding the rate of movement. The experiment began with a practice trial to demonstrate the procedure. The order of melodies was randomized separately for each individual.

Following the first phase, participants filled out a background questionnaire for 5–10 min. Following the questionnaire, participants completed a surprise recognition test by rating the same 24 melodies heard in the first phase ('old' melodies), as well as the remaining 24 foil melodies in the set ('new' melodies; $n = 12$ voice and $n = 12$ piano) on a scale from '1 – Definitely New' to '7 – Definitely Old'. There was no motor task during the recognition test. Here again, the order of melodies was randomized separately for each individual.

2.2. Results

2.2.1. Liking

Two scores were calculated for each participant by averaging liking ratings from the first phase of the experiment, separately for each timbre. Each score was calculated from 12 individual ratings with a possible range from 1 to 5. A mixed-model ANOVA with a between-participant factor of motor task group (Chewing, Squeezing Hand) and a within-participant factor of timbre (voice, piano) revealed a significant main effect for timbre, $F(1, 36) = 42.62$, $p < .001$, $\eta_p^2 = 0.54$, but no main effect for group, $F < 1$, and no interaction, $F < 1$. Across all participants, vocal melodies ($M = 3.03 \pm 0.77$) received lower ratings than piano melodies ($M = 3.95 \pm 0.59$), as they have in previous research (Weiss et al., 2012). We note, however, that different singers can yield different liking ratings, without differences in eventual recognition of their performances (Weiss et al., 2017).

2.2.2. Recognition

Two scores were calculated for each participant by calculating the area under the receiver operating characteristic curve (AUC; chance = 0.5, perfect recognition = 1.0), separately for each timbre. Each score was calculated using 24 individual ratings (12 old, 12 new). Descriptive statistics for AUC scores are visualized separately by group in Fig. 1A. A mixed-model ANOVA with a between-participant factor of motor task group (Chewing, Squeezing Hand) and a within-participant factor of timbre (voice, piano) revealed a significant main effect for timbre, $F(1, 36) = 8.30$, $p = .007$, $\eta_p^2 = 0.19$, which was the result of better recognition of vocal melodies ($M = 0.80$, $SD = 0.12$) than piano melodies ($M = 0.75$, $SD = 0.11$) across the sample. There was no main effect of group and no interaction between group and timbre, $F_s < 1$. An analysis using differences in ratings rather than AUC scores showed a similarly significant memory advantage for vocal melodies and no main effects or interactions involving group.

2.3. Discussion

Repetitive motor tasks of chewing and hand squeezing did not eliminate the voice advantage in Experiment 1, and there was no indication that chewing in particular reduced the magnitude of the voice advantage. These results could suggest that the voice memory advantage is not related to the phenomenon of earworms (Beaman et al., 2015), but such a conclusion would depend on whether explicit recognition memory is related to the occurrence of earworms to begin with. Chewing gum may act as a helpful distraction or suppression tool while trying to actively ignore a catchy, recently-heard song, and nevertheless have little to no impact on the processing of concurrent music, let alone vocal

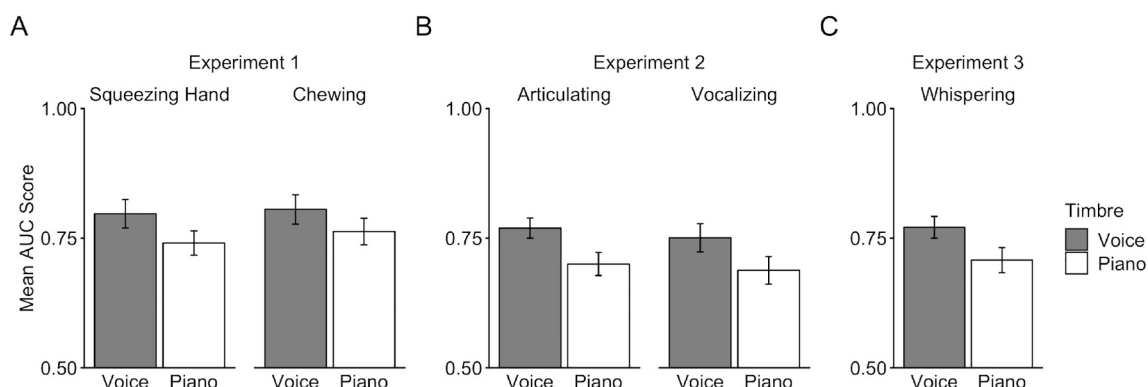


Fig. 1. Mean area under the receiver operating characteristic curve (AUC) by timbre, with separate panels for experiments and separate bars for between-participant manipulation of repetitive motor tasks during melody encoding. Memory for vocal melodies was consistently superior to memory for piano melodies, a replication of the voice advantage observed in previous research. Neither overall memory (i.e., AUC average across timbre), nor the magnitude of the voice advantage (i.e., $AUC_{\text{Voice}} - AUC_{\text{Piano}}$) differed by motor task, whether the comparison was made within experiment or across all experiments. Error bars are S.E.M.

music. Manipulations that overlap more substantially with singing could have more of an adverse effect on memory for vocal materials.

3. Experiment 2

In the current experiment, we repeated the method of Experiment 1, but included motor tasks more congruent with the actions of singing: articulating 'la la' repeatedly without vocalizing, or vocalizing 'omm' in a steady and sustained hum without articulating (i.e., similar to vocal toning in meditation).

3.1. Material and methods

3.1.1. Participants

Fifty-seven participants were assigned at random to two groups with different motor tasks during encoding, either articulating without vocalizing (hereafter, *Articulating* group; $n = 29$; 24 female; $M = 24.0 \pm 4.4$, range = 18–34 years) or vocalizing without articulating (hereafter *Vocalizing* group; $n = 28$; 21 female; $M = 23.1 \pm 3.9$, range = 18–32 years). Amount of formal musical training was similar between the groups (*Articulating*: $M = 4.9 \pm 4.6$, median = 4.0, range = 0–17 years; *Vocalizing*: $M = 6.3 \pm 4.6$, median = 5.3, range = 0–16 years). All participants were healthy and had normal hearing by self-report. Four additional participants were tested but excluded due to an age more than three standard deviations beyond all participants tested ($n = 1$), technical error ($n = 1$), or overall memory performance below chance ($n = 2$). All participants received compensation for their time, in accordance with methods approved by the ethics committee at the University of Montreal.

3.1.2. Stimuli

The stimuli were the same as in Experiment 1.

3.1.3. Apparatus

The apparatus was the same as in Experiment 1. For most of the *Articulating* group ($n = 24$), we placed EMG electrodes on the left masseter muscle to monitor movement of the mouth, but this approach was abandoned because the electrodes failed to consistently provide adequate signal in almost all cases, even as participants were visually observed to be complying through a window into the testing booth. We relied instead on participant self-reports of continued compliance during debriefing. All participants reported compliance.

For participants in the *Vocalizing* group, the apparatus was the same as in Experiment 1 except for the use of headphones (Beyerdynamic DT 770 Pro) rather than speakers to minimize auditory interference. Because participants were asked to vocalize audibly, compliance was confirmed via a microphone in the room (Neumann TLM-103). All participants complied, and reported compliance during debriefing.

3.1.4. Procedure

The procedure was identical to Experiment 1 except for the motor task performed while listening to melodies in the encoding phase. Here, participants were instructed to articulate the syllable 'la' repeatedly during the entirety of the melody, in silence (*Articulating*), or to vocalize with a sustained hum during the entirety of the melody (i.e., 'omm'), only breaking to inhale (*Vocalizing*). Participants in the *Articulating* group were instructed to articulate in a slightly exaggerated manner, as if singing with enthusiasm. No instructions were given regarding the rate of movement. Participants in the *Vocalizing* group were instructed to hum audibly, but at a low enough volume to remain able to attend to the melodies. No instructions were given regarding the pitch to hum or the rate of breathing.

3.2. Results

3.2.1. Liking

As in Experiment 1, two scores were calculated for each participant by averaging liking ratings from the first phase of the experiment, separately for each timbre. Each score was calculated from 12 individual ratings with a possible range from 1 to 5. A mixed-model ANOVA with a between-participant factor of motor task group (*Articulating*, *Vocalizing*) and a within-participant factor of timbre (voice, piano) revealed a significant main effect for timbre, $F(1, 55) = 23.01$, $p < .001$, $\eta_p^2 = 0.30$, but no main effect for group, $F < 1$, and no interaction, $F < 1$. As in Experiment 1, across all participants, vocal melodies ($M = 3.18 \pm 0.86$) received lower ratings than piano melodies ($M = 3.79 \pm 0.64$).

3.2.2. Recognition

Two scores were calculated for each participant by calculating the area under the receiver operating characteristic curve (AUC; chance = 0.5, perfect recognition = 1.0), separately for each timbre. Each score was calculated using 24 individual ratings (12 old, 12 new). Descriptive statistics for AUC scores by group are visualized in Fig. 1B. A mixed-model ANOVA with a between-participant factor of motor task group (*Articulating*, *Vocalizing*) and a within-participant factor of timbre (voice, piano) revealed a significant main effect for timbre, $F(1, 55) = 7.03$, $p = .010$, $\eta_p^2 = 0.11$, which was the result of better recognition of vocal melodies ($M = 0.76$, $SD = 0.13$) than piano melodies ($M = 0.69$, $SD = 0.13$) across all participants. There was no main effect of group and no interaction between group and timbre, $F_s < 1$. An analysis using ratings rather than AUC scores showed a similarly significant memory advantage for vocal melodies and no effects or interactions involving group.

3.3. Discussion

The results of Experiment 2 showed a voice advantage in memory for melodies, despite the inclusion of articulating and vocalizing tasks during encoding. The overall AUC scores collapsed across timbre and motor task were slightly lower here than in Experiment 1 (see Fig. 1A vs. B, and Section 4.3.4), which could be a result of the tasks in Experiment 2 being more difficult or unnatural compared to chewing gum.

Technical issues prevented consistent recording of mouth movements by EMG in the *Articulating* group, though participants reported compliance during debriefing, and could be observed complying from a window into the testing booth. Video monitoring could provide a more reliable measure of compliance in the future.

Articulating and vocalizing may have failed to sufficiently engage or disrupt subvocalization. Silent articulation and toning may become a relatively automatic movement, much in the way that music can elicit the absentminded tapping of fingers, feet, or legs. A vocal-motor task combining both continuous coordination of articulation and sound production might succeed in disrupting the preferential processing of vocal music where these motor tasks failed.

4. Experiment 3

In Experiment 3, participants performed a motor task during encoding that involved both articulation and the generation of sound: Whispering 'la la' repeatedly while listening to melodies at encoding. This task bears resemblance to the methodology reported in Assaneo et al. (2019), which found that participants who spontaneously synchronized their whisper rate to steady-streams of syllables had superior word-learning performance.

4.1. Method

4.1.1. Participants

Twenty-nine participants were included in the analysis (18 female;

$M = 21.1 \pm 2.0$, range = 18–27 years). Participants were recruited without regard to musical training ($M = 7.0 \pm 5.1$, median = 7.0, range = 0–18 years). All participants were healthy and had normal hearing by self-report. Four additional participants were tested but excluded due to technical errors ($n = 3$), or overall memory performance below chance ($n = 1$). All participants received compensation for their time, in accordance with methods approved by the ethics committee at the University of Montreal.

4.1.2. Stimuli

The stimuli were the same as in Experiments 1 and 2.

4.1.3. Apparatus

The apparatus was the same as used with the Vocalizing group of Experiment 2. Because participants were asked to whisper audibly, compliance was confirmed via a microphone in the room (Neumann TLM-103). All participants complied, and reported compliance during debriefing.

4.1.4. Procedure

The procedure was similar to the Articulating group in Experiment 2, except for the motor task performed while listening to melodies in the encoding phase. Here, participants were instructed to articulate and audibly whisper the syllable 'la' repeatedly during the entirety of the melody. Participants were instructed to whisper continuously, even when inhaling. No instructions were given regarding the rate or pattern of whispering.

4.2. Results

4.2.1. Liking

As in Experiments 1 and 2, two scores were calculated for each participant by averaging liking ratings from the first phase of the experiment, separately for each timbre. Each score was calculated from 12 individual ratings with a possible range from 1 to 5. A paired-samples *t*-test showed no significant difference between liking ratings for vocal melodies ($M = 3.59 \pm 0.72$) and piano melodies ($M = 3.82 \pm 0.53$), $t(28) = 1.61$, $p = .118$, although in absolute terms the piano was liked more than the voice, as observed in Experiments 1 and 2.

4.2.2. Recognition

Two scores were calculated for each participant by calculating the area under the receiver operating characteristic curve (AUC; chance = 0.5, perfect recognition = 1.0), separately for each timbre. Each score was calculated using 24 individual ratings (12 old, 12 new). Descriptive statistics for AUC scores are visualized in Fig. 1C. A paired-samples *t*-test revealed a significant main effect for timbre (voice, piano) on ratings, $t(28) = 2.72$, $p = .011$, Cohen's $d = 0.52$, which was the result of better recognition of vocal melodies ($M = 0.77$, $SD = 0.11$) than piano melodies ($M = 0.71$, $SD = 0.13$). An analysis using ratings rather than AUC scores showed a similarly significant memory advantage for vocal melodies.

4.2.3. Phase-locking of whisper and melodic beat

We adapted the phase-locking value (PLV) analysis of Assaneo et al. (2019) to examine the degree to which whispering rate synchronized to the beat of the melodies. Details of the PLV analysis are described in the Appendix. Briefly, we measured of the amplitude envelope of whisper audio against the amplitude envelope of a short tone pulsing in amplitude at the frequency of the nearest underlying beat hierarchy in the melody (Fig. 2). The analysis yielded 24 PLV scores per participant, one for each of the voice ($n = 12$) and piano ($n = 12$) trials during the encoding block of the experiment. PLV values can range from 0 (no phase-locking) to 1 (fully phase-locked).

We first considered whether PLV correlated with memory performance overall. Because of the small sample size and not having a priori

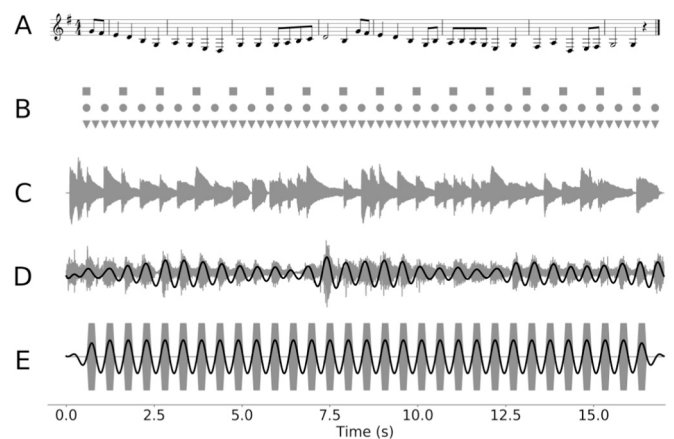


Fig. 2. Example of data for PLV analysis of a single trial. Panel A shows musical notation of the melody for reference. Panel B shows three different potential metrical levels of synchronization aligned to the melody waveform visualized in Panel C, which in this instance was played on piano. Squares represent a slower rate (half-note), circles a moderate rate (quarter-note), and triangles a faster rate (eighth-note), with all markers starting on the first down-beat. Panel D shows the waveform of the recording of participant whispering (grey), which was closest in rate to the moderate (circle) markers in Panel B. Thus, an arbitrary tone pulsing on and off at the rate of the moderate (circle) markers was generated and its waveform (grey) is visualized in Panel E. The sinusoidal lines (black) in Panels D and E represent the bandpass-filtered amplitude envelope of each signal, respectively. Differences in the phases of the sine waves over time were input to the PLV function in segments, as described in the Appendix, and the output was averaged to yield a single PLV value for the trial.

knowledge of the data distribution for the PLV measure, we used non-parametric tests (Spearman), however we note that parametric tests yielded similar results. Because the predictions are directional (i.e., a negative correlation between PLV and recognition would be uninterpretable), we report one-tailed results. A correlation between mean PLV (i.e., average of all 24 PLV scores) and mean recognition rating for the same melodies (i.e., from 1 to 7) was not significant, $r_s(27) = 0.11$, $p = .286$.

We next considered whether PLV differed according to timbre (i.e., average of $n = 12$ trials per timbre). Across participants, the average PLV score for vocal trials ($M = 0.529$, $SD = 0.128$) was nearly identical to the average PLV score for piano trials ($M = 0.523$, $SD = 0.120$), $t(28) = 0.33$, $p = .746$.

Finally, we considered whether PLV correlated with memory, separately by timbre. When limiting the data to vocal target melodies (i.e., averages of $n = 12$ PLV scores at exposure, $n = 12$ recognition ratings at test), the correlation between PLV and recognition rating approached but did not reach significance, $r_s(27) = 0.27$, $p = .077$ (see Fig. 3A, left panel). Considering only piano target melodies, PLV and recognition rating showed no relationship whatsoever, $r_s(27) = -0.06$, $p = .612$ (see Fig. 3A, right panel). A correlation test between the average PLV score at exposure (i.e., average of $n = 24$ PLV scores) and the magnitude of the voice advantage at test (i.e., $AUC_{\text{Voice}} - AUC_{\text{Piano}}$, incorporating both targets and foils) was significant, $r_s(27) = 0.37$, $p = .023$ (see Fig. 3B). Collectively these results provide evidence of a relationship between spontaneous synchronization and eventual memory advantages for the sung melodies, driven by the sung melodies in particular.

4.2.4. Recognition across experiments 1–3

To give context to the current recognition data, we compared performance in Experiment 3 to ungrouped data from Experiments 1 and 2 (i.e., regardless of motor task condition within experiment). First, we compared experiments based on overall memory, by calculating AUC scores using all melodies, regardless of timbre, with each score comprising 24 old and 24 new ratings. A one-way ANOVA approached

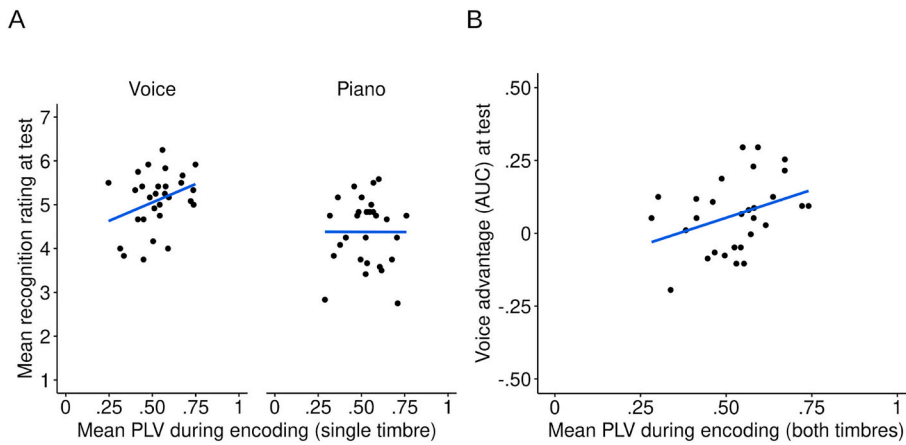


Fig. 3. Panel A shows the correlation of spontaneous synchronization of whisper rate at encoding (PLV) and recognition ratings at test for target melodies by timbre in Experiment 3. Point coordinates are an average of 12 trials at encoding (phase-locking value or PLV score; x-axis) and an average of recognition ratings for the same 12 melodies at test (from 1 – *Definitely New* to 7 – *Definitely Old*; y-axis), with each point representing a participant. For voice melodies alone the correlation approached but did not reach significance, $r_s(27) = 0.27$, $p = .077$ (one-tailed). For piano melodies alone there was virtually no relationship between PLV and recognition ratings, $p > .6$. Note that PLV scores did not differ by timbre in group average, $p > .7$. Panel B shows PLV across all trials at encoding ($n = 24$) plotted against the magnitude of the voice advantage at test (i.e., calculated using all test trials; $AUC_{\text{Voice}} - AUC_{\text{Piano}}$) for each participant. The correlation was significant, $r_s(27) = 0.37$, $p = .023$ (one-tailed).

significance, $F(2, 121) = 3.05$, $p = .051$, $\eta^2 = 0.05$, primarily due to the differences in overall AUC between Experiment 1 ($M = 0.78$, $SD = 0.10$) and Experiment 2 ($M = 0.73$, $SD = 0.08$), rather than Experiment 3 ($M = 0.74$, $SD = 0.10$). In other words, there is no evidence that the manipulation in Experiment 3 was more detrimental to overall memory than the other manipulations. A comparison across motor task groups (Squeezing, Chewing, Articulating, Vocalizing, Whispering) rather than experiment groups showed the same outcome, $F(4, 119) = 1.60$, $p = .178$.

Next, we compared the magnitude of the voice advantage (i.e., $AUC_{\text{Voice}} - AUC_{\text{Piano}}$, calculated for each individual) across experiments. Descriptive statistics are visualized in Fig. 4. A one-way ANOVA showed no hint of a difference, $F(2, 121) = 0.15$, $p = .862$. A comparison across

motor task groups (Squeezing, Chewing, Articulating, Vocalizing, Whispering) rather than experiment groups showed the same outcome, $F(4, 119) = 0.10$, $p = .983$. In short, there was no evidence that the variety of manipulations had any effect on the magnitude of the voice advantage.

4.2.5. Comparison of recognition in current and previous research

Pooling all participants across Experiments 1–3 ($n = 124$), the voice advantage was an average benefit of $M = 0.060$, $SD = 0.15$ (AUC), with an effect size of 0.49 (Cohen's d), and overall memory regardless of timbre was $M = 0.75$, $SD = 0.10$ (AUC). The nearest point of comparison to the current study, but without any motor task, is the first experiment of Weiss et al. (2019), specifically the condition in which half of the

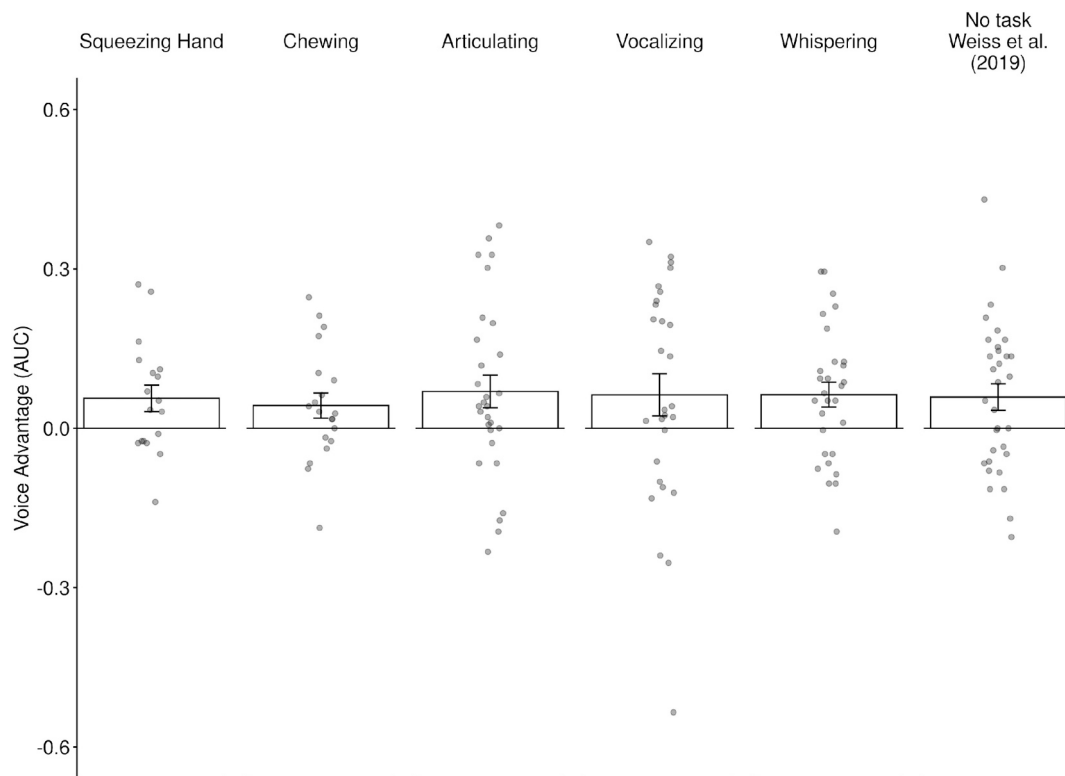


Fig. 4. Bars display mean magnitude of voice advantage (i.e., $AUC_{\text{Voice}} - AUC_{\text{Piano}}$) by motor manipulation task. Scores above zero represent better memory for vocal than piano melodies. There was no difference in the magnitude of the voice advantage across conditions in the current study, $p > .9$, and results reported here (first five bars) were comparable to a previously-published study with no motor condition (rightmost bar). Horizontally jittered points are individual data. Error bars are S.E.M.

melodies were vocal and half were piano (i.e., ‘50% vocal’ condition). Fig. 4 plots the voice advantage from that ‘no task’ condition alongside the current conditions. A reanalysis of those Weiss et al. (2019) data ($n = 32$) using AUC revealed a memory advantage for vocal melodies, $M = 0.059$, $SD = 0.14$ (AUC), with an effect size of $d = 0.46$, and overall memory regardless of timbre was $M = 0.74$, $SD = 0.10$ (AUC). The comparison data are, on average, remarkably consistent with the pooled data in the current experiments.

4.3. Discussion

The whispering task did not eliminate the presence of the voice advantage, nor did it reduce its magnitude compared to the motor tasks of squeezing, chewing, articulating, or vocalizing in Experiments 1 and 2. Whispering was, in principle, a more drastic manipulation because it involved repetitive movement for the duration of the stimulus (as with squeezing, chewing, and articulating) and the generation of sound (as with vocalizing), which none of the previous conditions had combined. Unlike the vocalizing condition in Experiment 2, the Whispering group was not voicing per se because whispers are unvoiced sounds. Their whispers were probably less audible than the toning sounds in the Vocalization group (i.e., when plugging one’s ears, toning is more audible to oneself than is whispering). This begs the question of whether any of these manipulations were drastic enough. It is possible that vocalizing or even singing the syllables ‘la la’ while listening to a novel melody might have more of an impact on the listener, and could be attempted in future research. However, such a dynamic, audible manipulation could also cause a general distraction from the melodies, which could depress overall memory to a point precluding inter-timbre comparisons.

Taken together, there was little evidence that memory was depressed by any of the motor conditions. A sample from Weiss et al. (2019) showed results consistent with those observed in the current experiments. Those participants heard the same number of vocal and piano melodies, nearly identical stimuli (i.e., same melodies and timbres but a different female singer), and they completed the same general procedure. Because the participants in that study had no concurrent motor task (i.e., listening only) but a similar pattern of results, we conclude that the current motor manipulations had little impact on the magnitude of the voice advantage or on overall memory. This is surprising because one would expect an added burden of performing various repetitive motor actions concurrently with a listening task, as when performing dual-versus single-tasks (e.g., Cocchini, Filardi, Crhonkova, & Halpern, 2017).

Spontaneous synchronization of whispers to the underlying beat of the melody did not differ by timbre, and did not significantly correlate with recognition ratings overall. However, it did correlate positively with the magnitude of the voice advantage. Moreover, the direction of the relationship is in line with the hypothesized link between spontaneous synchronization and auditory learning (Assaneo et al., 2019). Given the trends observed here, collection of data with a more powerful design seems warranted to further explore the relationship between auditory learning and synchronization. If degree of synchronization truly relates to enhanced recognition, then directing participants to whisper at the same rate as the beat of the melody might boost memory performance later, either through facilitated recognition of target melodies, facilitated rejection of foil melodies, or both.

The pattern of PLV and recognition results displayed in Fig. 3 raise the possibility that the free-form repetitive whispering task, rather than suppressing subvocalization, entrained subvocalization, which could in turn influence memory. In other words, by giving no instruction to participants about the rate of whispering, some individuals may have spontaneously adapted their rate to vocal melodies more than other individuals, perhaps because those melodies were more engaging and memorable, or vice versa. Individual differences in synchronization to speech tokens in Assaneo et al. (2019) revealed two naturally occurring

subgroups of synchronizers and non-synchronizers, with a boundary near PLV scores of 0.5. Across all trials, no clear bimodal distribution of synchronization was observed in our sample. It should be emphasized that average PLV scores were nearly identical for voice and piano melodies despite the differences in eventual recognition by timbre.

5. General discussion

Occupying different aspects of the vocal system with repetitive (chewing, articulating, whispering) or sustained (vocalizing) actions during encoding of melodies did not eliminate the memory advantage for vocal over instrumental melodies, and the magnitude of the voice advantage did not differ appreciably from previous research with a similar design but no motor distraction task. From these results we conclude that subvocalization is unlikely to be the main driver of the memory advantage for vocal music. This outcome extends previous research showing little impact of motor representations of melodies on memory, since pianists had a voice advantage over piano melodies, and no ‘piano advantage’ over other instruments (Weiss, Schellenberg, et al., 2015). In other words, there was no evidence that the ability to produce the melody facilitates encoding. Moreover, in that experiment, the magnitude of the voice advantage did not co-vary with musical expertise, even though memory was superior in musicians.

Our results provide little support for the idea that perception of singing involves a motor simulation of the performance (Russo, 2020), and align with research showing that listening to vocal music does not elicit vocal motor activity (Bruder & Wöllner, 2019). Nevertheless, the notion that covert simulation is necessarily interrupted by overt movement may not hold true in all scenarios, for example, if the purported simulation recruits different brain networks than those required for action planning and execution of a concurrent task, or if the extra task sufficiently overwhelms cognitive load. In particular, the phase-locking analysis from Experiment 3 provides novel evidence that the degree of spontaneous synchronization between rate of overt movement and musical beat relates to memory advantages for the voice. Perhaps for this reason, requiring participants to tap or articulate rapidly and out of sync with a melody can attenuate the voice advantage and melodic memory in general (Wood, Rovetti, & Russo, 2020). In contrast, repetitive motor actions that become ‘automatic’ or spontaneously align with musical features (e.g., the beat) could be neutral or even beneficial to encoding, even if those actions simultaneously interrupted the simulation of other musical features (e.g., simulating a pitch sequence). Future research could investigate the role of directed synchronization versus directed non-synchronization, or alternatively, directed versus unconstrained or spontaneous synchronization, on memory for music. Another approach could consider melodies that vary in their tendency to elicit movement (i.e., ‘groove’, Matthews, Witek, Heggli, Penhune, & Vuust, 2019), or ask whether the same melody is perceived as ‘groovier’ when sung versus played instrumentally.

The role of the vocal or motor system in other areas of music perception and cognition is an important area of research. In particular, manipulation of auditory imagery seems likely to engage motor planning. Singing accuracy relates to the ability to manipulate auditory information in imagery and singing production tasks, such that inaccurate singers appear to rely more heavily on motor information than accurate singers when asked to produce manipulated melodies, for example, in reversed note order (Greenspon, Pfordresher, & Halpern, 2017; Pfordresher, Halpern, & Greenspon, 2015). Whether singing ability relates to memory for vocal melodies has yet to be tested empirically, though we note that participants with congenital amusia, who tend to be poor singers, showed no difference in the magnitude of the voice advantage relative to control participants (Weiss & Peretz, 2019). Whether there are differences in the ability to manipulate auditory imagery for sung versus instrumental melodies remains to be tested.

Returning to the opening quote from Smith (1895), we are tempted to speculate that the nature of ‘conflicting’—or alternatively,

congruent—activity of the vocal-motor system during perception of novel music could have more to do with networks that mediate beat or timing perception than relative pitch perception. Indeed, listeners can synchronize their body movements to a singer's movements through audio alone (Pouw, Paxton, Harrison, & Dixon, 2020). Spontaneous synchronization to vocal melodies might be an indicator of interest or arousal in vocal signals, and future research could consider this possibility.

The best current account of the voice advantage in memory remains that the voice is special, by virtue of being a conspecific, communicative, and biological signal, which is more engaging or distinctive to listeners than instrumental sounds. In other words, arousal and attention related to the distinctiveness of a conspecific signal could enhance memory processes without motor simulation. Cognitive processes stemming from enhanced arousal and attention for a distinctive stimulus could in turn enhance memory (e.g., Talmi & McGarry, 2012). In a broader sense, this interpretation of the voice advantage is consistent with a view of music as an optimal medium for the communication of arousal (Loui, Bachorik, Li, & Schlaug, 2013; Salimpoor, Benovoy, Longo, Cooperstock, & Zatorre, 2009) and affect (van Goethem & Sloboda, 2011), and as a rewarding stimulus (Salimpoor et al., 2013), especially in contexts hypothesized to contribute to the evolution of music, such as infant-parent interactions that involve singing (e.g., Cirelli, Jurewicz, & Trehub, 2020) and chants or songs that promote social cohesion (e.g., Kirschner & Tomasello, 2010).

Acknowledgements

Michael W. Weiss, Anne-Marie Bissonnette, and Isabelle Peretz, Department of Psychology, University of Montreal, Montreal, QC, Canada. The work is funded by the Fonds de Recherche du Québec: Nature et Technologies (MW), the Natural Sciences and Engineering Research Council of Canada (IP), and the Canada Research Chairs program (IP). We thank Madeleine Borgeat for assistance in data collection, Véronique Martel, Chanel Marion-St-Onge, and two anonymous reviewers for feedback on a previous draft, and M. Florencia Assaneo for sharing analysis code. Address correspondence to Michael W. Weiss: michael.weiss@umontreal.ca

Appendix

The phase-locking value (PLV) analysis was based on Assaneo et al. (2019), but the input data and parameters were modified due to task differences. Specifically, while each study compared the recording of a participant whispering to an auditory stimulus, their stimuli were speech tokens presented isochronously (rate of 4.5 Hz), whereas our stimuli were melodies that were not isochronous and varied in tempo. The musical nature of our stimuli necessitated the consideration of the metrical hierarchy.

First, we determined the approximate rate of whispering for each trial. The onset of each whispered syllable was identified from an audiovisual representation of the recorded audio, using custom software (Python). An inter-onset-interval (IOI) was calculated for each adjacent pair of notes, yielding a series of IOI values for each trial, and the whisper rate was calculated as the median value. Across all trials, the rate ranged from 208 to 991 ms (IOI), which translates to 60.5–288.5 events per minute, or beats per minute. We note that this estimate (median) does not necessarily indicate continuous, isochronous whispering at the calculated rate.²

Next, as visualized in Fig. 2, the nearest underlying beat hierarchy

was considered on a trial-by-trial basis according to the tempo of the melody, the time signature of the melody, and the median rate of syllable whispering during the trial. The purpose of this adjustment was to account for potential synchronization to different metrical levels (e.g., every quarter note, every eighth note, etc.). For example, a melody in 4/4 time signature recorded at 120 beats per minute might elicit synchronization at the rate of half notes (IOI = 1000 ms), at the rate of quarter notes (IOI = 500 ms), or at the rate of eighth notes (IOI = 250 ms). Any of those rates would be equally valid, hence we selected the rate closest to their whisper rate. We then generated a tone pulsing at that metrical rate (i.e., aligned with the stimulus) for comparison to the whisper audio. A preliminary analysis confirmed that for 27 of 29 participants, average PLV across trials was higher when comparing whispers to the pulsing tone envelope rather than the melody envelope.

As in Assaneo et al. (2019), amplitude envelopes were resampled to 100 Hz and bandpass-filtered around the stimulus rate. In our case, the rate of the stimulus varied, so envelopes were bandpass filtered relative to the frequency of the tone pulse (± 0.5 Hz) rather than a single fixed value. PLV was calculated using the MATLAB (Mathworks, Inc.) code below, in segments of 5 s with 2 s segment advancement:

$$PLV = \text{abs}(\text{sum}(\exp(1i * \text{segment_phase_diff}))) / \text{length}(\text{segment_phase_diff})$$

where $1i$ is the basic imaginary unit, `segment_phase_diff` is a vector of sample-by-sample differences in phase angle between the bandpass filtered envelopes of two signals, and `length` returns the number of samples in the segment. Finally, an average PLV score was calculated across the duration of the trial.

References

- Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience*, 22(4), 627–632. <https://doi.org/10.1038/s41593-019-0353-z>.
- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63(1), 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>.
- Beamman, C. P., Powell, K., & Rapley, E. (2015). Rapid communication: Want to block earworms from conscious awareness? B(u)y gum! *Quarterly Journal of Experimental Psychology*, 68(6), 1049–1057. <https://doi.org/10.1080/17470218.2015.1034142>.
- Bruder, C., & Wöllner, C. (2019). Subvocalization in singers: Laryngoscopy and surface EMG effects when imagining and listening to song and text. *Psychology of Music*. <https://doi.org/10.1177/0305735619883681>, 0305735619883681.
- Campbell, S. M., & Margulis, E. H. (2015). Catching an earworm through movement. *Journal of New Music Research*, 44(4), 347–358. <https://doi.org/10.1080/09298215.2015.1084331>.
- Cirelli, L. K., Jurewicz, Z. B., & Trehub, S. E. (2020). Effects of maternal singing style on mother–infant arousal and behavior. *Journal of Cognitive Neuroscience*, 32(7), 1213–1220. https://doi.org/10.1162/jocn_a.01402.
- Cocchini, G., Filardi, M. S., Crhonkova, M., & Halpern, A. R. (2017). Musical expertise has minimal impact on dual task performance. *Memory*, 25(5), 677–685. <https://doi.org/10.1080/09658211.2016.1205628>.
- Criswell, E. (2011). *Cram's introduction to surface electromyography*. Jones & Bartlett Publishers.
- van Goethem, A., & Sloboda, J. (2011). The functions of music for affect regulation. *Musicae Scientiae*, 15(2), 208–228. <https://doi.org/10.1177/1029864911401174>.
- Goldinger, S. D., Papeš, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, 23(4), 959–978. <https://doi.org/10.3758/s13423-015-0860-1>.
- Greenspon, E. B., Pfordresher, P. Q., & Halpern, A. R. (2017). Pitch imitation ability in mental transformations of melodies. *Music Perception: An Interdisciplinary Journal*, 34(5), 585–604. <https://doi.org/10.1525/mp.2017.34.5.585>.
- Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, 31(5), 354–364. <https://doi.org/10.1016/j.evolhumbehav.2010.04.004>.
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: An fMRI study. *Human Brain Mapping*, 30(3), 859–873. <https://doi.org/10.1002/hbm.20550>.
- Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal stm by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *The Quarterly Journal of Experimental Psychology Section A*, 56(8), 1249–1268. <https://doi.org/10.1080/02724980244000765>.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36. [https://doi.org/10.1016/0010-0277\(85\)90021-6](https://doi.org/10.1016/0010-0277(85)90021-6).
- Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 13(3), 110–114. <https://doi.org/10.1016/j.tics.2008.11.008>.

² Four participants had frequent occurrences of pauses in their whispering (IOIs >1500 ms), which upon inspection reflected inhales that were not voiced audibly as directed. Excluding these participants had no effect on the recognition results, and their data were included in all analyses.

- Loui, P., Bachorik, J., Li, H. C., Schlaug, et al. (2013). Effects of voice on emotional arousal. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00675>. Article 675.
- Matthews, T. E., Witek, M. A. G., Heggeli, O. A., Penhune, V. B., & Vuust, P. (2019). The sensation of groove is affected by the interaction of rhythmic and harmonic complexity. *PLoS One*, 14(1). <https://doi.org/10.1371/journal.pone.0204539>.
- Murray, D. J. (1967a). The role of speech responses in short-term memory. *Canadian Journal of Psychology*, 21(3), 263. <https://doi.org/10.1037/h0082978>.
- Murray, D. J. (1967b). Overt versus covert rehearsal in short-term memory. *Psychonomic Science*, 7(10), 363–364. <https://doi.org/10.3758/BF03331125>.
- Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, 78(4 Pt.1), 679. <https://doi.org/10.1037/h0026641>.
- Peynircioglu, Z. F. (1995). Covert rehearsal of tones. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1), 185–192. <https://doi.org/10.1037/0278-7393.21.1.185>.
- 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(3), 242–253. <https://doi.org/10.1525/mp.2015.32.3.242>.
- Pouw, W., Paxton, A., Harrison, S. J., & Dixon, J. A. (2020). Acoustic information about upper limb movement in voicing. *Proceedings of the National Academy of Sciences*, 117(21), 11364–11367. <https://doi.org/10.1073/pnas.2004163117>.
- Russo, F. (2020). Motor system involvement in the perception of singing. In F. Russo, B. Ilari, & A. J. Cohen (Eds.), *The Routledge companion to interdisciplinary studies in singing, volume I: Development*. Routledge.
- Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, 21(2), 150–164. [https://doi.org/10.1016/S0022-5371\(82\)90521-7](https://doi.org/10.1016/S0022-5371(82)90521-7).
- Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., & Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science*, 340(6129), 216–219. <https://doi.org/10.1126/science.1231059>.
- Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PLoS One*, 4(10), Article e7487. <https://doi.org/10.1371/journal.pone.0007487>.
- Schendel, Z. A., & Palmer, C. (2007). Suppression effects on musical and verbal memory. *Memory & Cognition*, 35(4), 640–650. <https://doi.org/10.3758/BF03193302>.
- Smith, W. G. (1895). The relation of attention to memory. *Mind*, 4, 47–73.
- Talmi, D., & McGarry, L. M. (2012). Accounting for immediate emotional memory enhancement. *Journal of Memory and Language*, 66(1), 93–108. <https://doi.org/10.1016/j.jml.2011.07.009>.
- Weiss, M. W., & Peretz, I. (2019). Ability to process musical pitch is unrelated to the memory advantage for vocal music. *Brain and Cognition*, 129, 35–39. <https://doi.org/10.1016/j.bandc.2018.11.011>.
- Weiss, M. W., Schellenberg, E. G., Peng, C., & Trehub, S. E. (2019). Contextual distinctiveness affects the memory advantage for vocal melodies. *Auditory Perception & Cognition*, 2, 47–66. <https://doi.org/10.1080/25742442.2019.1642078>.
- Weiss, M. W., Schellenberg, E. G., & Trehub, S. E. (2017). Generality of the memory advantage for vocal melodies. *Music Perception: An Interdisciplinary Journal*, 34(3), 313–318. <https://doi.org/10.1525/mp.2017.34.3.313>.
- Weiss, M. W., Schellenberg, E. G., Trehub, S. E., & Dawber, E. J. (2015). Enhanced processing of vocal melodies in childhood. *Developmental Psychology*, 51(3), 370–377. <https://doi.org/10.1037/a0038784>.
- Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the way she sings: Enhanced memory for vocal melodies. *Psychological Science*, 23(10), 1074–1078. <https://doi.org/10.1177/0956797612442552>.
- Weiss, M. W., Trehub, S. E., Schellenberg, E. G., & Habashi, P. (2016). Pupils dilate for vocal or familiar music. *Journal of Experimental Psychology: Human Perception and Performance*, 42(8), 1061–1065. <https://doi.org/10.1037/xhp0000226>.
- Weiss, M. W., Vanzella, P., Schellenberg, E. G., & Trehub, S. E. (2015). Pianists exhibit enhanced memory for vocal melodies but not piano melodies. *Quarterly Journal of Experimental Psychology*, 68(5), 866–877. <https://doi.org/10.1080/17470218.2015.1020818>.
- Wood, E. A., Rovetti, J., & Russo, F. A. (2020). Vocal-motor interference eliminates the memory advantage for vocal melodies. *Brain and Cognition*, 145, 105622. <https://doi.org/10.1016/j.bandc.2020.105622>.
- Wynn, J. S., Shen, K., & Ryan, J. D. (2019). Eye movements actively reinstate spatiotemporal mnemonic content. *Vision*, 3(2), 21. <https://doi.org/10.3390/vision3020021>.