Attention, Perception, & Psychophysics
https://doi.org/10.3758/s13414-025-03073-y

Effects of vocal-motor interference on vocal pitch imitation

Emma B. Greenspon[1,2] · Tim A. Pruitt[3] · Andrea R. Halpern[4] · Peter Q. Pfordresher[3]

Accepted: 28 March 2025

Abstract

The act of singing involves perceptual processing of pitch as well as vocal motor planning. One mechanism that may underlie this type of sensorimotor processing is auditory imagery. Prior research showed that less accurate singers, who report less vivid auditory imagery, engage in more preparatory subvocal movements than more accurate singers. The current research addressed the degree to which motor processing is causally involved in the preparatory auditory imagery that is associated with accurate vocal pitch matching. On each trial, participants were presented with a novel four-note target melody, then imagined the melody, and finally sang the melody aloud. During the auditory imagery period of the trials, participants either simply imagined the target melody (control) or imagined the melody while simultaneously engaged in one of the following secondary tasks: silently repeating the syllables /bɑtah/ (articulatory interference), continually droning a low quiet hum (phonatory interference), or exhaling continuously (phonatory suppression). Phonatory interference significantly disrupted vocal pitch matching relative to performance in the control task. Furthermore, the degree of phonatory disruption was related to participants' overall singing accuracy, such that less accurate singers were more disrupted by phonatory interference than more accurate singers. This suggests that less accurate singers rely on subvocalization to prepare for imitation, whereas more accurate singers may be able to rely on mental models of the actions needed to imitate pitch when subvocalization is prevented.

Keywords Auditory imagery · Subvocalization · Motor planning · Singing

Vocal imitation is a behavior that supports the development of language (Kuhl & Meltzoff, 1996) and may underlie the development of musical skills. It is well known that individuals vary considerably in the accuracy with which they can vocally imitate pitch. Complementing work on vocal pitch imitation in music, individual differences in vocal pitch imitation are also observed in the domain of speech, and importantly, poor-pitch matchers tend to be inaccurate at replicating the pitch contour of spoken sentences (Mantell & Pfordresher, 2013). Poor-pitch singing may therefore be associated with a broader *vocal pitch imitation deficit* (Pfordresher & Larrouy-Maestri, 2015) that can potentially limit the accuracy of vocal imitation in a variety of contexts. The present study is part of a larger effort by the authors to better understand, and ultimately remediate, problems some individuals experience with vocal pitch imitation.

Vocal pitch imitation relies on a network of processes, including motor planning, auditory feedback, perception, and memory, as summarized in the *sensorimotor vocal loop* model (Berkowska & Dalla Bella, 2009). Recent research has supported the importance of pitch perception in vocal pitch matching through observations that pitch imitation accuracy, not surprisingly, is positively associated with simple pitch discrimination and higher-order measures of pitch perception (Greenspon & Pfordresher, 2019; Pfordresher & Nolan, 2019). However, inaccurate vocal pitch imitation can exist without impaired pitch perception. For this reason, this behavior can be differentiated from congenital amusia—a rare musical disorder characterized by impaired pitch perception (estimated 1.5%; Peretz & Vuvan, 2017), which is far less prevalent than inaccurate singing in the general

✉ Emma B. Greenspon
egreenspon@sjf.edu

1 Department of Psychology, St. John Fisher University, Rochester, NY, USA

2 Department of Psychology, Monmouth University, West Long Branch, NJ, USA

3 Department of Psychology, University at Buffalo, The State University of New York, Buffalo, NY, USA

4 Department of Psychology, Bucknell University, Lewisburg, PA, USA

population (closer to 30%; Berkowska & Dalla Bella, 2013; Hutchins & Peretz, 2012; Pfordresher & Demorest, 2021; Pfordresher & Larrouy-Maestri, 2015). For these reasons, along with evidence that poor singers typically have intact ability to control vocal pitch in nonimitative contexts (Pfordresher & Brown, 2007; Pfordresher & Mantell, 2009), our research focuses primarily on the role of *sensorimotor translation* between auditory representations and vocal motor planning.

The multi-modal imagery association (MMIA) model developed by our group links vocal pitch matching to the accuracy and precision of associations between internal simulations for pitch (auditory imagery) and the vocal motor plans underlying pitch production (Pfordresher et al., 2015). This model, based on multiple sources of evidence, further proposes that auditory imagery plays a critical role in mediating between perceptual and motor representations of a target pitch, such that accurate images may prime the activation of associated motor plans (Pfordresher et al., 2015). First, neuroimaging studies have found evidence that auditory imagery involves sensorimotor processing, as supported by observed activation in motor planning areas—particularly the supplementary motor area—during auditory imagery (Lima et al., 2015, 2016; Zatorre & Halpern, 2005). Second, accuracy in vocal pitch imitation correlates with self-reported vividness of auditory imagery (Pfordresher & Halpern, 2013), accuracy of pitch imagery (Greenspon & Pfordresher, 2019), and performance on auditory mental transformation tasks that are thought to rely heavily on imagery (Greenspon et al., 2017, 2020, 2023). Finally, auditory imagery seems to rely on covert motor processes for the formation of both verbal and pitch imagery. Evidence for this claim comes from previous studies which have shown that auditory imagery is disrupted when tasks interfere with subvocalization during the formation of auditory representations of speech and music (Aleman & van't Wout, 2004; Reisberg et al., 1989; Smith et al., 1995). More recently, studies relying on physiological measures of subvocalization suggest that covert laryngeal muscle activations are recruited during auditory imagery (Bruder & Wöllner, 2021; Pruitt et al., 2019) and silent sight-reading of musical scores (Brodsky et al., 2008). The current research focuses on implications arising from this latter body of evidence and addresses the degree to which motor processes are recruited during auditory imagery of pitch using secondary tasks that were designed to disrupt various forms of muscle activity involved in vocal production.

The present research tests the hypothesis arising from the theory that auditory imagery supports the processes of sensorimotor translation during vocal imitation (Pfordresher et al., 2015). This theory is based on evidence showing that ease of auditory image formation is associated with enhanced neural activity in motor planning areas (Lima et al., 2016), that more gray matter in supplementary motor areas correlates with imagery vividness (Lima et al., 2015), and peripheral muscle activity accompanies forms of auditory imagery (Brodsky et al., 2008). In addition, auditory imagery shows vulnerability to disruption by competing motor tasks (e.g., Smith et al., 1995). Based on this theory, auditory imagery should be associated with peripheral motor activity within muscle groups used for vocal pitch control, and this motor activity may play a significant role in the accuracy of vocal pitch imitation.

A previous study tested the first part of this hypothesis by measuring orofacial activity using surface electromyography (sEMG) during auditory imagery (Pruitt et al., 2019). Participants listened to brief melodies, mentally rehearsed those melodies, and then reproduced those melodies via singing. During mental rehearsal, sEMG measurements were collected from muscles associated with vocal pitch control in the throat (the left and right sternohyoid muscles), facial muscles associated indirectly with pitch (muscles around the lips and eyebrows), and a control site on the bicep. Only sternohyoid muscle activity increased significantly during auditory imagery relative to a visual imagery comparison task. This result supports the presence of muscle movements during auditory imagery that precedes vocal production and complements previous physiological findings that covert laryngeal activity is recruited during auditory imagery and musical sight-reading (Brodsky et al., 2008; Bruder & Wöllner, 2021). However, these findings do not address the extent to which this motor activity influences pitch imitation accuracy.

To test a possible causal role of subvocalization in preparatory auditory imagery, here, we adopted a production task modeled after prior research on auditory imagery (Smith et al., 1995), and more recent studies on the role of covert vocal production in memory for melodies (Wood et al., 2020). In the current study, participants listened to, mentally rehearsed, and then vocally reproduced short melodies. It is worth noting that participants were instructed to refrain from any overt stimulus-specific vocalization during perception and mental rehearsal of the target melodies; a protocol we have used in prior research (Greenspon et al., 2023; Pruitt et al., 2019). In different blocks of trials, participants engaged in secondary tasks during mental rehearsal that were designed to either suppress or actively interfere with motor activity. In doing so, these secondary tasks were used to test different hypothetical roles for motor processes during the formation of auditory imagery preceding vocalization.

First, it is possible that preparatory imagery depends on subvocal motor movements that are specific to the demands associated with the task. In the present study, participants completed singing tasks in which they varied their vocal pitch to reproduce different tones in a melody. Thus, task-specific motor movements would be laryngeal. By contrast, because each tone is articulated the same way (i.e., with the same syllable), articulatory demands are minimal. Two tasks were designed specifically to target laryngeal muscles (left and right sternohyoid) involved in vocal phonation that would play a role in subvocal control of pitch during mental rehearsal. In one task, the *phonatory interference* task, participants hummed a quiet drone tone while mentally rehearsing the target melody. The phonatory interference task was anticipated to interfere with the use of laryngeal muscles, including the sternohyoid, during subvocalization of the imagined melody by reallocating the use of these muscles to the irrelevant act of quietly humming a single tone. In another task, the *phonatory suppression* task, participants silently exhaled without overt vocalization while mentally rehearsing the target melody. Sternohyoid activity in particular has been shown to be involved in pitch control during phonation and is suggested to be more directly involved in inhalation than exhalation during respiration (for reviews, see Jürgens, 2002; Kubin, 2016). For these reasons, silent exhalation was chosen as a task intended to limit laryngeal activity, specifically of the sternohyoid muscle, during subvocalization of the imagined melody. If subvocal phonatory movements are necessary for preparatory imagery, then both phonatory interference and suppression tasks should disrupt imagery, and therefore disrupt accuracy of production.

A second hypothesis is that preparatory imagery does not depend on laryngeal muscle activity during subvocalization, but subvocalization may nevertheless generate an auditory image as a result of activating cortical sensorimotor associations. If so, then suppression of phonation may not disrupt production, but interference with phonation would be disruptive. The logic of this hypothesis is that the interference task would cause participants to generate a competing auditory image, by way of vocal motor control, that disrupts production. However, because imagery may not depend on subvocal movements, simply suppressing their activity should not lead to significant disruption.

A third hypothesis is that motor interference does not target task-specific auditory imagery, but instead disrupts production due to competing demands. This hypothesis is addressed by a third task, *articulatory interference,* which was designed specifically to target lip muscles that are involved in the articulation of speech sounds. In this task, participants silently articulated syllables during the mental rehearsal phase of a trial. Similar types of subvocal articulation have been used in past research of motor processes to disrupt auditory imagery (Smith et al., 1995). Finally, we included a control task based on the auditory imagery task of Pruitt et al. (2019) in which participants mentally rehearsed the target melody without engaging in a secondary task (i.e., suppression or interference).

We hypothesized that phonatory, but not articulatory, interference would influence pitch imitation accuracy. We left open the direction of this influence. One plausible direction would be that interference would disrupt accuracy, leading to more pitch errors in imitation. This prediction follows from the hypothesis that covert vocal muscle activity is necessary for imitation accuracy. However, the previous results of Pruitt et al. (2019) and Greenspon et al. (2023) suggest the possibility of the reverse direction. Specifically, in both studies increased subvocal phonatory sEMG activity was associated with *reduced* pitch imitation accuracy. This result can be interpreted as implying that subvocal activity constitutes a maladaptive strategy that interfered with auditory imagery. We also approached this study without an a priori hypothesis concerning whether both phonatory interference and suppression tasks, or only the interference task, would disrupt production. Analyses of individual differences were then used to address whether effects of the secondary tasks vary across participants with respect to overall vocal imitation accuracy.

Method

Participants

Forty-five Introductory Psychology students at the University at Buffalo, SUNY, received course credit for their participation in this study. Four participants exhibited a pitch discrimination threshold that failed to meet the inclusion criterion of a threshold of 100 cents or lower and were thus excluded from further analysis. The final sample consisted of 41 participants who had a mean age of 19.05 years (range: 18–22), and 21 participants exhibited a female vocal range.[1] Musical experience also varied within the sample, where participants reported an average of 4.38 years of musical experience (range: 0–17).

Materials and equipment

Auditory stimuli

Target stimuli were eight melodies, each comprising four-note sequences performed by a model vocalist singing on the syllable /du/ (previously used in Pruitt et al., 2019). Participants imitated male or female vocal models that matched their vocal range, and vocal range was defined independent of a participants' self-reported gender.[2]

[1] Demographic information for the variables age, vocal range, and musical experience were missing for three participants due to experimenter error.

[2] Self-reported gender was not used to define a participants' vocal range though there were no instances where a participants' vocal range did not align with their self-reported gender for this sample.

Additionally, target melodies were centered around a musical key that approximated the participant's comfort pitch to further minimize possible vocal strain. Participants that exhibited a male vocal range imitated melodies in either the key of A2, D3, or F3[3] sung by a male vocal model, and participants that exhibited a female vocal range imitated melodies in either the key of F3, A3, or D4 sung by a female vocal model.

Audio recordings and experiment interface

Participants' vocalizations were recorded in a sound-attenuated booth (Whisper Room Inc., SE 2000 Series, Morristown, TN) using a Shure PG58 dynamic microphone connected to Focusrite Scarlett 2i2 USB audio interface. Digital recordings were stored as .wav files and captured at a 22050 Hz sampling rate with 16-bit resolution. Stimulus presentation and vocal recordings were carried out using MATLAB (The MathWorks, Natick, MA) on a 3.6 GHz Dell computer. Visual cues appeared on a 15-in. LCD computer screen positioned directly in front of the participant. Auditory stimuli and feedback were played through two sets of speakers. The primary source was a pair of Mackie CR3 series Multimedia Monitors (LOUD Technologies, Woodinville, WA, USA) which flanked the LCD computer screen. The secondary source, which supported accelerometry recordings used for sEMG data synchronization, was a single Yamaha KS10 Keyboard Speaker (Yamaha Pro Audio, Inc., Hamamatsu, Shizuoka, Japan).

Electromyography recording

All sEMG data were collected with the Mini Wireless EMG system (Delsys Trigno Wireless EMG Systems, Boston, MA, USA). Based on our previous work (Pruitt et al., 2019) and others (e.g., Stepp, 2012), we elected to focus on two anatomical sites: the upper lip (*m. orbicularis oris superioris*) and sternohyoid (*m. sternohyoideus*) muscles. Left and right upper lip sensors were placed lateral to the philtrum and superior to the vermilion border. Left and right sternohyoid sensors were placed by first determining the location of the space between the thyroid and cricoid cartilage of the neck and then positioning the sensor 1 cm lateral and 1 cm superior to this referent. The participant's nondominant bicep (*m. biceps brachii*) was used as a control site.

Procedure

Screening task

All participants were screened approximately 1–2 weeks prior to the main experiment. This session served to initially assess the participant's pitch imitation and discrimination abilities, exclude those with hearing deficits, identify individuals for whom facial sEMG could be reliably measured (e.g., possessing minimal-to-no facial hair), and administer a battery of questionnaires.

The session began with the Seattle Singing Accuracy Protocol (SSAP; Demorest et al., 2015; Pfordresher & Demorest, 2020), which includes pitch imitation and pitch discrimination measures. The pitch imitation tasks required participants to complete three kinds of imitation tasks in which pitches fell within 7 semitones of their comfort range (determined by asking participants to produce a single pitch that they felt comfortable singing and singing a familiar song in a comfortable key). First, participants completed 10 single-note pitch imitation trials based on a vocal timbre for the target. Vocal timbres were drawn from recordings of a trained male and female singer, with target gender matching the participant's vocal range defined by their comfort pitch. Participants then completed 10 single-note pitch imitation trials based on a piano timbre for the target. Finally, participants completed six trials in which they imitated pitch patterns of varying melodic contour produced with a vocal timbre. After the imitation tasks, participants sung a familiar song from memory twice, once using the song's lyrics and once only using the syllable /du/. Next participants completed an adaptive pitch discrimination task based on one devised by Loui et al. (2008). The SSAP procedure concluded with brief music, language, and demographic questionnaires.

The Bucknell Auditory Imagery Scale (BAIS) was also included in the set of questionnaires. This self-report inventory is composed of 28 items, half of which query imagery vividness and the other half query imagery control. Previous work has shown that BAIS responses correlate with several neural and behavioral measures of auditory imagery (Halpern, 2015) as well as pitch imitation accuracy (Greenspon et al., 2017; Greenspon & Pfordresher, 2019; Pfordresher & Halpern, 2013).

Experimental task

Once informed consent was gained, sEMG sensors were affixed to the anatomical target sites following sEMG protocols outlined in Pruitt et al. (2019). Participants then engaged in a series of vocal warm-up exercises followed by a series of practice trials to introduce the experimental task.

[3] None of the participants with a male vocal range in the final sample exhibited a comfort pitch of F3, so melodies sung by a male vocal model in the key of F3 were not used. All other keys were used for male and female vocal ranges.

On each trial (see Fig. 1), a fixation cross first appeared for 2,000 ms followed by the presentation of the target melody. After the presentation of the melody, participants completed the imagery period of the trial in which an illustration of a thought cloud appeared for 4,000 ms to remind the participant they should be mentally rehearsing the just-heard melody. Trials were organized in four blocks and the first block of trials consisted of the control task in which participants simply imagined the target melody during the imagery period without any secondary task. The remaining trials were blocked by secondary task (phonatory interference, phonatory suppression, and articulatory interference) with the secondary tasks counterbalanced across blocks. Participants engaged in the secondary tasks during this *imagery period.* Low-amplitude masking noise was also played to obscure participants' self-produced auditory feedback that may have resulted from the secondary tasks. After the imagery period, participants were prompted to imitate the melody by the appearance of a green dot on the screen.

The three secondary tasks were selected to specifically affect different components of vocalization. *Phonatory interference* required participants to emit a quiet hum at a constant pitch for 4,000 ms during the imagery period to engage muscles related to pitch production and control. Before the start of the block of trials for the phonatory interference task, participants were first asked to hum out loud at a comfortable level using a pitch of their choice. They were then asked to hum at the softest possible level to anchor the participants' volume at the lowest level for the task. Participants were then given the instructions for the structure of the trials and were explicitly instructed to "hum as quietly as you can" during the imagery period. Whereas phonatory interference involved engagement of pitch-control muscles, a second task, *phonatory suppression,* involved disengagement of vocal muscles by requiring participants to exhale silently without overt vocalization throughout the imagery period. The third task, *articulatory interference,* aimed to engage the tongue and lip articulators by having participants silently repeat the disyllabic nonword /bɑtah/ during the imagery period. Participants were explicitly instructed to " [move] your mouth as if you were speaking, without making any noise."

A total of 96 experimental trials were subdivided into four blocks of 24 trials each, comprising three repetitions for each of the eight melodies, with each repeated melody occurring after a complete cycle of all eight melodies in a random order. All participants first completed a block of control task trials, and the subsequent three blocks were counterbalanced across phonatory interference, phonatory suppression, and articulatory interference tasks. The control task was always presented in the first block of trials to obtain a baseline measure of performance prior to introduction of any interference instructions since we assumed it would be difficult to inhibit a secondary task once learned. At the onset of Blocks 2–4, participants were instructed on how to engage in the secondary task during the imagery period of all trials in that block. They then completed a brief set of practice trials to ensure they

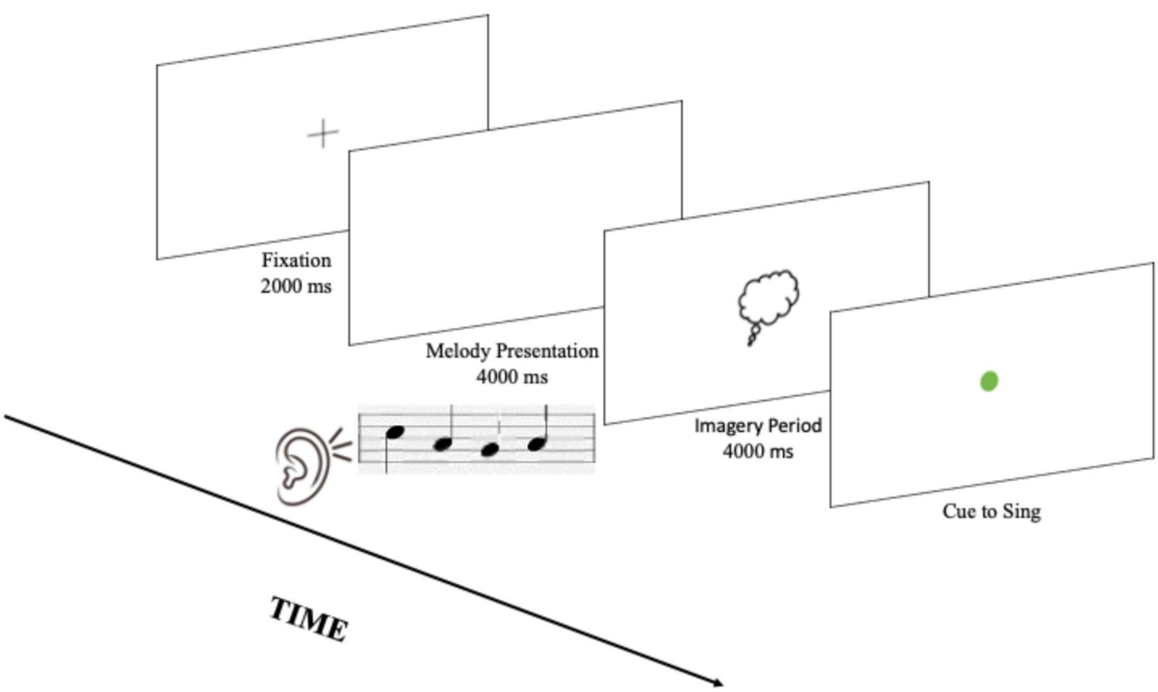

Fig. 1 Illustration of experimental trial structure

understood the task before completing the experimental trials.

Data analysis

Pitch imitation recordings The MATLAB function YIN (de Cheveigné & Kawahara, 2002) was used to extract f_0 from the audio recordings of the participants' imitations. Boundaries of each sung pitch were then identified using a semi-automated procedure in MATLAB based on fluctuations in the amplitude envelope. This same procedure generated estimates of sung pitch based on median f_0 in the central 50% of samples within defined boundaries. The absolute difference between the produced f_0 associated with each note and the corresponding f_0 from the target constituted the primary measure of vocal imitation accuracy.[4] We averaged these absolute deviation scores across all notes in a trial, and then across all trials for a given task before submitting the data to further analyses.

sEMG recordings The sEMG recordings were processed following guidelines by Stepp (2012) and procedures used previously by our laboratory (Greenspon et al., 2023; Pruitt et al., 2019). Linear detrending corrected for DC-offsets and high-pass filtering with a 10 Hz cutoff controlled for movement artifacts. Then, 20–450 Hz band-pass filtering emphasized frequencies commonly associated with muscle contractions. Electrical power contaminant signals were removed using an infinite impulse response (IIR) notch filter centered at 60 Hz. The sEMG signals were then rectified and smoothed. Research assistants then visually inspected the data for movement artifacts not accounted for in the filtering process by identifying large signal deflections, marking the onset and offset, and removing the intervening data points.

Standardized time points in a trial were next identified for normalization purposes to account for individual and morphological differences by subtracting muscle-specific means of a trial's listening period (i.e., baseline) from corresponding imagery period means (Greenspon et al., 2023; Livingstone et al., 2016). Baseline values were determined by passing a 1,000 ms moving average window across the listening period of the trial to identify a segment exhibiting average minimum activity. Secondary task motor activity values were determined by passing a 200 ms moving average window across the entire imagery period and extracting the highest mean to serve as a peak activity estimate. The sEMG data were then filtered using an upper boundary of +100 μV and lower boundary of −20 μV as identified by examining each sensor's interquartile ranges to remove extreme values. Statistical outliers were then defined as peak muscle activity values beyond ±3 standard deviations from the mean and were removed from further processing.[5]

Results

Pitch imitation accuracy

We analyzed the effect of the secondary tasks on mean absolute pitch deviation scores (henceforth, pitch deviations) using a single-factor within-subjects analysis of variance (ANOVA). Means and standard errors for each task are shown in Fig. 2. The effect of task was significant, $F(1, 40) = 3.75$, $p = .013$, $\eta_p^2 = .09$. Pairwise contrasts using Holm–Bonferroni correction revealed significantly worse performance in the phonatory interference task compared with the control task ($p = .02$), but no other contrasts were significant (phonatory suppression versus control, $p = .12$, articulatory interference versus control, $p = .10$).

To further explore the role of subvocal muscle activity within our study, we next evaluated whether there was a relationship between disruption from the secondary tasks in the experiment and singing accuracy as measured by pattern imitation in the SSAP. To provide a measure of phonatory disruption, we subtracted each participant's average performance in the control task from their average performance in the phonatory interference task. We found a significant association such that more accurate singers (i.e., singers with lower pitch deviations) tended to show less phonatory disruption than less accurate singers (i.e., singers with higher pitch deviations), $r(36) = .44$, $p < .05$, as shown in Fig. 3A.[6]

We next evaluated this relationship for suppression disruption (phonatory suppression–control) and articulatory disruption (articulatory interference–control). We found a similar relationship for articulatory disruption as shown in Fig. 3B, $r(36) = .38$, $p < .05$, but not suppression disruption ($p > .5$). No other predictors from the screening task, including self-reported auditory imagery as measured by the BAIS vividness and control subscales, were significantly

[4] Many studies of vocal pitch imitation use the proportion of produced pitches categorized as accurate, with absolute deviations within 50 cents serving as the criterion. We analyzed such scores for the current data, which lead to comparable results but were less statistically robust. The correlation between the two measures of vocal accuracy across all participants and experiment tasks was strong and positive, $r(178) = .88$, $p < .001$.

[5] The outlier removal processes resulted in the loss of less than 12% of all trials for the right lip sensor and the loss of less than 6.5% of all trials within each of the other four sensors.

[6] We defined outliers for the correlational analyses in Fig. 3 as any value beyond ±3 standard deviations from the mean. All data fell within the defined boundary so no values were removed from the analyses.

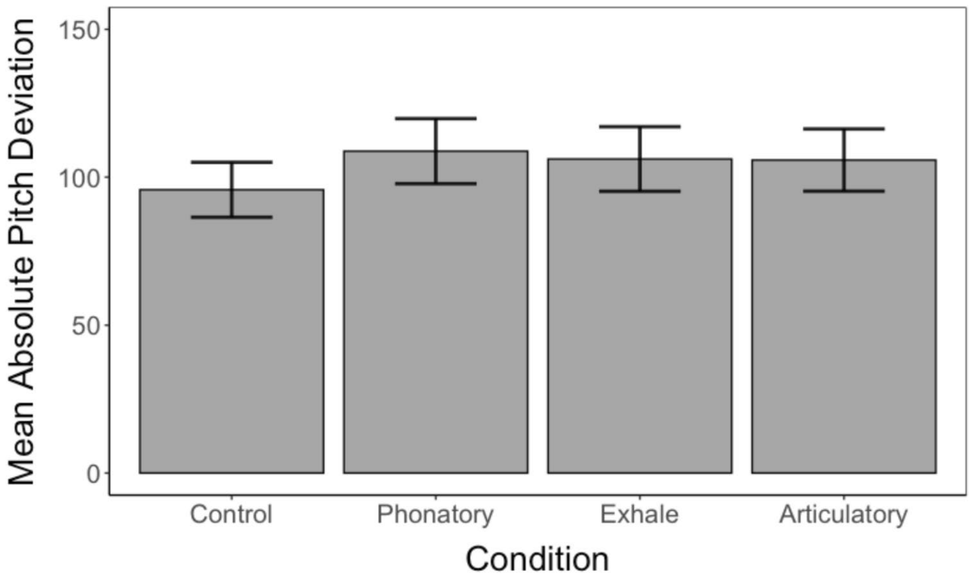

Fig. 2 Mean absolute pitch deviations across secondary tasks. Control = control task, Phonatory = phonatory interference task, Exhale = phonatory suppression task, Articulatory = articulatory interference task. Error bars represent 1 standard error of the mean

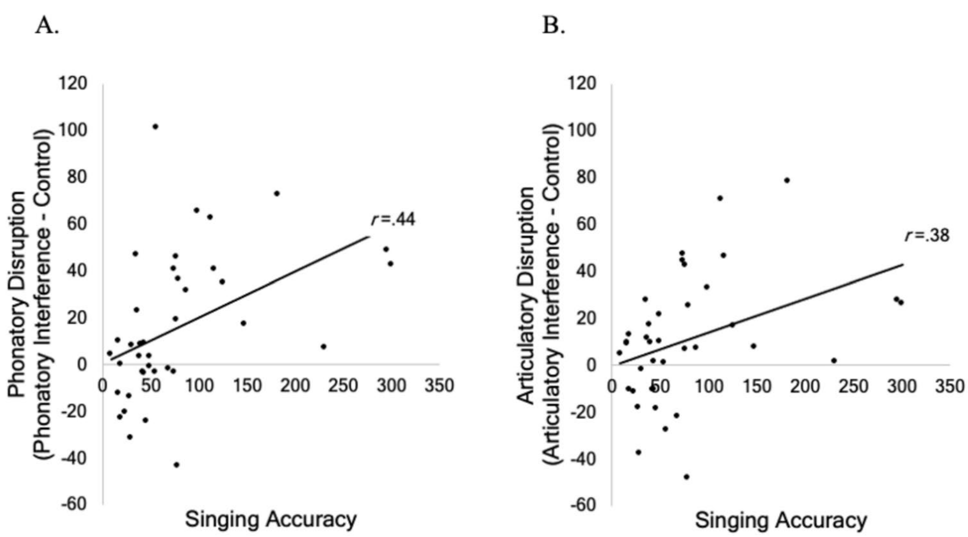

Fig. 3 Scatterplots illustrating the relationship between disruption and singing accuracy. **A.** Association between phonatory disruption and singing accuracy. **B.** Association between articulatory disruption and singing accuracy

correlated with any of the three disruption measures (all p values > .08 after false discovery rate correction).

sEMG activity

All sEMG activity was analyzed using a 4 (secondary task: control, phonatory suppression, phonatory interference, articulatory interference) × 5 (sensor: left lip, right lip, left sternohyoid, right sternohyoid, bicep) within-subjects ANOVA. There was a significant main effect of secondary task, $F(3, 123) = 8.11$, $p < .001$, $\eta_p^2 = .17$, a significant main effect of sensor, $F(4, 164) = 4.35$, $p = .002$, $\eta_p^2 = .10$, and significant Secondary Task × Sensor interaction, $F(12, 492) = 3.87$, $p < .001$, $\eta_p^2 = .09$. Next, a series of one-way ANOVAs were run to test the secondary task's effects within a specific sensor since the main effect of sensor was driven by broad mean differences across sensors. Descriptive statistics for the secondary tasks for each sensor are shown in Table 1.

We anticipated that lip movements would be most prevalent during the articulatory interference task. The

Table 1 sEMG activity in microvolts (μV) by secondary task (row) and muscle sensor (column)

	Left sternohyoid		Right sternohyoid		Left lip		Right lip		Bicep	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control	5.59	4.02	4.38	2.84	5.50	4.20	5.68	3.79	0.58	0.38
Phonation	8.96	7.83	7.71	6.22	7.22	6.48	10.34	9.18	0.65	0.66
Exhalation	7.67	7.01	6.56	5.44	10.68	8.47	15.37	11.52	0.68	0.53
Articulation	8.82	8.24	7.74	5.44	19.08	11.59	23.99	14.45	0.54	0.35

one-way ANOVA for each lip sensor yielded a significant effect of task: left sensor, $F(3, 123) = 30.04$, $p < .001$, $\eta_p^2 = .42$; right sensor, $F(3, 123) = 31.86$, $p < .001$, $\eta_p^2 = .44$. Post hoc pairwise contrasts (Holm–Bonferroni corrected) revealed greater activation during the articulatory interference task than any other task for each sensor (both *p* values < .001). In addition, for both sensors activation was greater during exhalation than the control task (both *p* values < .001), suggesting some lip movement during exhalation. The right, but not left, sensor also showed significant differences between phonatory interference (i.e., humming) and the control task as well as between phonatory interference and phonatory suppression (all *p* values < .01). These sEMG data suggest that participants were following instructions during the articulatory interference task, insofar as there were larger lip muscle contractions during articulation than during other secondary tasks.

We also anticipated that the phonatory interference task would lead to activation of the sternohyoid muscle, which contributes to vocal pitch control. We found a significant effect of secondary task on sternohyoid activations for both sensors, left sensor: $F(3, 123) = 3.12$, $p < .05$, $\eta_p^2 = .07$; right sensor: $F(3, 123) = 6.58$, $p < .001$, $\eta_p^2 = .14$. Post hoc pairwise comparisons for the right sensor yielded significant differences between the phonatory interference task and the control task ($p = .010$), as well as between the articulatory interference task and the control task ($p = .002$). For the left sensor, no pairwise contrasts were statistically significant, and only the contrast between phonatory interference and control yielded a *p* value less than .10 ($p = .055$). These sEMG data suggest that participants engaged in subvocal pitch control of the sternohyoid muscle (as anticipated) and also lip movements during the phonatory interference task. No evidence for subvocal pitch control during the phonatory suppression task (i.e., exhalation) was evident, which is consistent with the conceptualization of this task as suppressing vocal pitch control.

Lastly, bicep motor activity was not expected to differ across secondary tasks due to its irrelevance in vocal phonation and articulation. Consistent with predictions, a one-way ANOVA determined that the bicep's muscle activity did not vary across tasks, $F(3, 123) = 1.23$, $p = .30$, $\eta_p^2 = .03$. This finding replicates our previous work showing that bicep muscle activity is not recruited during preparatory imagery prior to vocal imitation (Greenspon et al., 2023; Pruitt et al., 2019).

We next addressed associations between our psychophysiological (sEMG values) and behavioral measures (absolute pitch deviations) within the experiment. Three participants were excluded from these analyses due to missing data for at least one measure. For this set of analyses, we focused on the sternohyoid muscle given this muscle's theorized role in pitch control. Associations between sternohyoid activity and absolute pitch deviations within a task were not significant for either the left or right sternohyoid muscle (all *p* values > .1).

Discussion

The present study tested whether subvocalization would influence pitch imitation accuracy, by having participants mentally rehearse pitch sequences while completing secondary tasks designed to prohibit the use of muscles that may be involved in subvocalization prior to singing the target sequence. We considered two directions for the influence of subvocalization on singing accuracy: subvocalization as a facilitative mechanism and subvocalization as a maladaptive mechanism. Results suggest that overall, imagery-related subvocalization facilitates imitation. This claim is driven by our findings that disrupting subvocalization by having participants engage in task-irrelevant phonation (i.e., quietly humming a single tone) reduced imitation accuracy compared with performance in the control task. Muscle contractions that are associated with the control of pitch therefore may be a component of preparatory auditory imagery used to plan vocal motor sequences. In this way, the present study offers an important causal link that builds upon previous experiments that provided correlational evidence of associations between subvocal muscle activity and musical imagery (Brodsky et al., 2008; Greenspon et al., 2023; Pruitt et al., 2019).

We measured muscle activity during the preparatory imagery period to verify that the targeted muscles were recruited during the different secondary tasks. Although muscle activity was not limited to a single task, we found evidence that the targeted muscles were engaged during the intended tasks. As expected, we observed the greatest level of lip activity during the articulatory interference task compared with the other three tasks. We also observed greater laryngeal activity, based on sEMG measures of the sternohyoid muscles, during the phonatory interference task compared with the control task. In addition, we found evidence that subvocalization was not limited to the phonatory interference task in that sternohyoid activity was also significantly greater during articulatory interference compared with the control task. And as anticipated, bicep muscle activity did not vary by task.

We found that the effect of motor interference on behavioral performance was selective despite sternohyoid activity being similarly recruited during the phonatory and articulatory interference tasks. Although articulatory interference nominally increased pitch deviations in later production, this difference was not statistically significant after a family-wise alpha correction. This is an important result because one could imagine a scenario in which any motor interference task serves to distract the participant, and the distraction then disrupts production. In this context, it is worth noting that all interference tasks were designed to be as simple as possible, thus not leading to the kind of distracting effect one might find in a standard dual-task methodology. An alternative explanation is that subvocalization may recruit task-specific activity of the lips during the imagery period prior to vocalization similar to how silent reading recruits covert lip muscle activity (McGuigan & Winstead, 1974). However, unlike silent reading, the current study had participants imagine auditory representations that relied on a single syllable so articulatory demands were minimal and therefore covert lip movements during the imagery period were not anticipated to facilitate performance, in line with our current results.

The selectivity of the effect of phonatory interference on imitation behavior rules out an account based on the idea that any generalized arousal influences muscle contractions in a manner that impacts performance; rather, our results show that subvocalization in vocal pitch imitation is muscle specific. In this way, the present work complements studies showing that articulatory interference disrupts verbal auditory imagery (Aleman & van't Wout, 2004; Reisberg et al., 1989; Smith et al., 1995). Furthermore, we found that the disruptive effect of phonatory interference was related to singing ability, such that less accurate singers tended to be more disrupted by phonatory interference compared with more accurate singers. We did not include a nonvocal interference task given the expectation that muscles involved in the vocal system are more likely to disrupt performance than muscles independent of the vocal system. The specificity of motor interference on vocal pitch imitation could be addressed in future work.

When considering singing ability on a continuum (Pfordresher & Larrouy-Maestri, 2015), we find that the effectiveness of subvocalization may relate to overall singing ability. More specifically, we suggest that less accurate singers may rely more strongly on subvocal muscle activity during preparatory imagery than more accurate singers (Greenspon et al., 2023; Pruitt et al., 2019); but this is not necessarily an effective strategy for less accurate singers and does not result in better singing performance. This result informs the findings from Pruitt et al. (2019) and Greenspon et al. (2023), who reported that increased subvocal activity was related to *less* accurate vocal pitch imitation accuracy. This association was not replicated in the current paper, though it is worth noting that the attempt to replicate this association resulted in lower statistical power due to limiting these analyses to a subset of trials (i.e., trials within the control task rather than across all trials of the experiment).

Interestingly, when subvocal activity was prevented in the current study through phonatory interference (i.e., humming), less accurate singers tended to exhibit greater disruption to pitch accuracy than more accurate singers, as shown in Fig. 3A, indicating that preventing the use of subvocalization did not improve performance for these singers. On the other hand, more accurate singers may have access to two strategies for vocal pitch imitation which may not be available to less accurate singers. One strategy may involve the use of subvocal muscle activity during preparatory imagery. This claim is motivated by the overall disruptive effect of phonatory interference found in the current experiment. The second strategy may rely on motor representations driven by cortical activity in the supplemental motor area (or possibly other areas; see below) when imagining the target melody. Accurate singers may rely more on this second strategy when subvocalization is prevented; such was the case in the phonatory interference task. This claim is motivated by the finding that more accurate singers tended to be less disrupted by phonatory interference than less accurate singers. Furthermore, previous research has shown that individual differences in self-reported and behavioral measures of auditory imagery are related to singing accuracy (Greenspon & Pfordresher, 2019; Pfordresher & Halpern, 2013) and are related to structural and functional measures of the supplemental motor area when imagining the human voice (Lima et al., 2015). Further research is needed to fully understand the role of subvocalization for individuals with differing levels of singing ability.

In contrast to phonatory interference, it is worth noting that preventing contractions of muscles involved in phonation through the phonatory suppression task (i.e., exhaling) did not significantly disrupt performance relative to the control task. More accurate singers may have been able to use mental representations of pitch to partially compensate for effects of preventing subvocalization using phonatory interference and phonatory suppression in the current study. On the other hand, less accurate singers may not be able to effectively use subvocalization during preparatory imagery and therefore any subvocal muscle activity during the imagery phase may have contributed to less accurate performance regardless of whether the muscle activity is related to the current task or not, whereas simply preventing muscle engagement during subvocalization was not as disruptive for these singers. Such a claim is supported by the findings that the sternohyoid muscle was recruited during the articulatory interference task in addition to the correlation between articulatory disruption and overall singing accuracy displayed in Fig. 3B showing that less accurate singers tended to be more disrupted by articulatory interference than more accurate singers. Directly addressing the effects of phonatory interference, phonatory suppression, and articulatory interference across different levels of singing ability is needed to clarify this argument.

It is worth noting that the association between auditory imagery and motor planning is subtle, possibly indirect, and may help address why we did not observe reliable associations between performance in the experiment and self-reported auditory imagery scores as measured by the BAIS. It is also possible this null effect may be partially attributed to the precision of our measure of singing accuracy (i.e., pitch deviations vs. percentage correct). In support of the argument of an indirect relationship between auditory imagery and motor planning, neuroimaging results during auditory imagery reveal activation in motor planning areas such as the supplementary motor area and premotor cortex but no significant activity in the primary motor cortex (Halpern, 2015; Lima et al., 2016). The task used here likely had strongest effects on the primary motor cortex and likely only modest effects for motor planning given the simplicity of the task. Moreover, subvocal muscle contractions are themselves subtle and often hard to detect. It is also important to note that the sternohyoid muscle is associated with changing vocal pitch (Roubeau et al., 1997). Participants in the phonatory interference task sustained a single pitch during the imagery period and were discouraged to change pitch either within or across trials. For this reason, the fact that such subtle peripheral activity had any effect on subsequent imitation in our view is noteworthy.

In conclusion, the present results add to the literature suggesting that motor movements play a role in the generation of an auditory image (Brodsky et al., 2008; Pruitt et al., 2019; Smith et al., 1995; Wood et al., 2020; but see Weiss et al., 2021). Importantly, we suggest that the role of these motor movements may differ across levels of singing ability such that less accurate singers may rely more strongly on subvocal muscle activity during preparatory imagery than more accurate singers. Accurate singers may be able to rely additionally on accurate sensorimotor associations that are supported by cortical motor activity (Lima et al., 2015) rather than peripheral motor movements. Even with the different potential support networks for motor activity (peripheral vs. cortical) across singing ability, together our work supports the view that vocal pitch imitation relies on the formation of a sensorimotor image that facilitates motor planning via priming of the motor system (Pfordresher et al., 2015).

Acknowledgements We thank the following undergraduate research assistants for their assistance with data collection and analysis: Gina George, Aaron Carter, Kayden Koh, Johnathan Liow, Chantel Fatorma, Andy Cruz, Elijah Cone, and Faith Moka.

Funding This work was supported by the NSF BCS-1848930.

Availability of data and materials The experiment was not preregistered. Materials for the experiment are available upon request and data for the experiment is available (https://osf.io/d2hbr/?view_only=6afca97e0b4c418abb84d9e4e14e6e74).

Code availability Custom codes will be shared upon reasonable request.

Declarations

Conflicts of interest None.

Ethics approval The research presented was approved by the University at Buffalo, SUNY Institution Review Board.

Consent to participate All participants were 18 years or older and provided informed consent to participate in the experiment.

Consent for publication All participants were 18 years or older and provided informed consent to have their data published.

References

Aleman, A., & van't Wout, M. (2004). Subvocalization in auditory-verbal imagery: Just a form of motor imagery? *Cognitive Processing, 5*, 228–231. https://doi.org/10.1007/s10339-004-0034-y

Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology, 5*, 69–83. https://doi.org/10.2478/v10053-008-0068-2

Berkowska, M., & Dalla Bella, S. (2013). Uncovering phenotypes of poor-pitch singing: The Sung Performance Battery (SPB). *Frontiers in Psychology, 4*, 714. https://doi.org/10.3389/fpsyg.2013.00714

Brodsky, W., Kessler, Y., Rubinstein, B., Ginsborg, J., & Henik, A. (2008). The mental representation of music notation: Notational audiation. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 427–445. https://doi.org/10.1037/0096-1523.34.2.427

Bruder, C., & Wöllner, C. (2021). Subvocalization in singers: Laryngoscopy and surface EMG effects when imagining and listening to song and text. *Psychology of Music, 49*, 567–580. https://doi.org/10.1177/030573561988368

De Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America, 111*, 1917–1930. https://doi.org/10.1121/1.1458024

Demorest, S. M., Pfordresher, P. Q., Dalla Bella, S., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015). Methodological perspectives on singing accuracy: An introduction to the special issue on singing accuracy (Part 2). *Music Perception, 32*, 266–271. https://doi.org/10.1525/mp.2015.32.3.266

Greenspon, E. B., & Pfordresher, P. Q. (2019). Pitch-specific contributions of auditory imagery and auditory memory in vocal pitch imitation. *Attention, Perception, & Psychophysics, 81*, 2473–2481. https://doi.org/10.3758/s13414-019-01799-0

Greenspon, E. B., Pfordresher, P. Q., & Halpern, A. R. (2017). Pitch imitation ability in mental transformations of melodies. *Music Perception, 34*, 585–604. https://doi.org/10.1525/mp.2017.34.5.585

Greenspon, E. B., Pfordresher, P. Q., & Halpern, A. R. (2020). The role of long-term memory in mental transformations of pitch. *Auditory Perception & Cognition, 3*, 76–93. https://doi.org/10.1080/25742442.2020.1866428

Greenspon, E. B., Gentile, A. M., Pruitt, T. A., Halpern, A. R., & Pfordresher, P. Q. (2023). Subvocalization during preparatory and non-preparatory auditory imagery. *Auditory Perception & Cognition, 6*, 108–127. https://doi.org/10.1080/25742442.2022.2163582

Halpern, A. R. (2015). Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music Mind and Brain, 25*, 37–47. https://doi.org/10.1037/pmu0000081

Hutchins, S., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology: General, 141*, 76–97. https://doi.org/10.1037/a0025064

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews, 26*, 235–258. https://doi.org/10.1016/S0149-7634(01)00068-9

Kubin, L. (2016). Neural control of the upper airway: Respiratory and state-dependent mechanisms. *Comprehensive Physiology, 6*, 1801–1850. https://doi.org/10.1002/cphy.c160002

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America, 100*, 2425–2438. https://doi.org/10.1121/1.417951

Lima, C. F., Lavan, N., Evans, S., Agnew, Z., Halpern, A. R., Shanmugalingam, P., & Scott, S. K. (2015). Feel the noise: Relating individual differences in auditory imagery to the structure and function of sensorimotor systems. *Cerebral Cortex, 25*, 4638–4650. https://doi.org/10.1093/cercor/bhv134

Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends in Neurosciences, 39*, 527–542. https://doi.org/10.1016/j.tins.2016.06.003

Livingstone, S. R., Vezer, E., McGarry, L. M., Lang, A. E., & Russo, F. A. (2016). Deficits in the mimicry of facial expressions in Parkinson's disease. *Frontiers in Psychology, 7*, 780. https://doi.org/10.3389/fpsyg.2016.00780

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action–perception mismatch in tone-deafness. *Current Biology, 18*, R331–R332. https://doi.org/10.1016/j.cub.2008.02.045

Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition, 127*, 177–202. https://doi.org/10.1016/j.cognition.2012.12.008

McGuigan, F. J., & Winstead, C. L. (1974). Discriminative relationship between covert oral behavior and the phonemic system in internal information processing. *Journal of Experimental Psychology, 103*, 885–890.

Peretz, I., & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics, 25*, 625–630. https://doi.org/10.1038/ejhg.2017.15

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness." *Music Perception, 25*, 95–115. https://doi.org/10.1525/mp.2007.25.2.95

Pfordresher, P. Q., & Demorest, S. M. (2020). Construction and validation of the Seattle Singing Accuracy Protocol (SSAP): An automated online measure of singing accuracy. In F. Russo, B. Ilari, & A. Cohen (Eds.), *Routledge companion to interdisciplinary studies in singing: Development* (vol. 1, pp. 322–333). Routledge.

Pfordresher, P. Q., & Demorest, S. M. (2021). The prevalence and correlates of accurate singing. *Journal of Research in Music Education, 69*(1), 5–23. https://doi.org/10.1177/0022429420951630

Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review, 20*, 747–753. https://doi.org/10.3758/s13423-013-0401-8

Pfordresher, P. Q., & Larrouy-Maestri, P. (2015). On drawing a line through the spectrogram: How do we understand deficits of vocal pitch imitation? *Frontiers in Human Neuroscience, 9*, 271. https://doi.org/10.3389/fnhum.2015.00271

Pfordresher, P., & Mantell, J. (2009). Singing as a form of vocal imitation-Mechanisms and deficits. In J. Louhivouri, T. Eerola, S. Saarikallio, T. Himberg, & P.-S. Eerola (Eds.), *ESCOM 2009: 7th Triennial Conference of European Society for the Cognitive Sciences of Music* (pp. 425–430). University of Jyväskylä.

Pfordresher, P. Q., & Nolan, N. P. (2019). Testing convergence between singing and music perception accuracy using two standardized measures. *Auditory Perception & Cognition, 2*, 67–81. https://doi.org/10.1080/25742442.2019.1663716

Pfordresher, P. Q., Halpern, A. R., & Greenspon, E. B. (2015). A mechanism for sensorimotor translation in singing: The Multi-Modal Imagery Association (MMIA) model. *Music Perception, 32*, 242–253. https://doi.org/10.1525/mp.2015.32.3.242

Pruitt, T. A., Halpern, A. R., & Pfordresher, P. Q. (2019). Covert singing in anticipatory auditory imagery. *Psychophysiology, 56*, e13297. https://doi.org/10.1111/psyp.1329

Reisberg, D., Smith, J. D., Baxter, D. A., & Sonenshine, M. (1989). "Enacted" auditory images are ambiguous; "pure" auditory images are not. *The Quarterly Journal of Experimental Psychology, 41*, 619–641. https://doi.org/10.1080/14640748908402385

Roubeau, B., Chevrie-Muller, C., & Lacau Saint Guily, J. (1997). Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta Oto-Laryngologica, 117*, 459–464.

Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia, 33*, 1433–1454. https://doi.org/10.1016/0028-3932(95)00074-D

Stepp, C. E. (2012). Surface electromyography for speech and swallowing systems: Measurement, analysis, and interpretation. *Journal of Speech, Language, and Hearing Research, 55*, 1232–1246. https://doi.org/10.1044/1092-4388(2011/11-0214)

Weiss, M. W., Bissonnette, A. M., & Peretz, I. (2021). The singing voice is special: Persistence of superior memory for vocal melodies despite vocal-motor distractions. *Cognition, 213*, 104514. https://doi.org/10.1016/j.cognition.2020.104514

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

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: Musical imagery and auditory cortex. *Neuron, 47*, 9–12. https://doi.org/10.1016/j.neuron.2005.06.013