Received: 17 May 2018 | Revised: 16 September 2018 | Accepted: 19 September 2018

DOI: 10.1111/psyp.13297

ORIGINAL ARTICLE

WILEY PSYCHOPHYSIOLOGY SPR

Covert singing in anticipatory auditory imagery*

Tim A. Pruitt[1] | **Andrea R. Halpern**[2] | **Peter Q. Pfordresher**[1]

[1]Department of Psychology, University at Buffalo, The State University of New York, Buffalo, New York

[2]Department of Psychology, Bucknell University, Lewisburg, Pennsylvania

Correspondence
Tim A. Pruitt, Department of Psychology, University at Buffalo, North Campus, 242 Park Hall, Buffalo, NY 14260.
Email: tapruitt@buffalo.edu

Funding information
Natural Science Foundation BCS grant (1256964)

Abstract

To date, several fMRI studies reveal activation in motor planning areas during musical auditory imagery. We addressed whether such activations may give rise to peripheral motor activity, termed subvocalization or covert singing, using surface electromyography. Sensors placed on extrinsic laryngeal muscles, facial muscles, and a control site on the bicep measured muscle activity during auditory imagery that preceded singing, as well as during the completion of a visual imagery task. Greater activation was found in laryngeal and lip muscles for auditory than for visual imagery tasks, whereas no differences across tasks were found for other sensors. Furthermore, less accurate singers exhibited greater laryngeal activity during auditory imagery than did more accurate singers. This suggests that subvocalization may be used as a strategy to facilitate auditory imagery, which appears to be degraded in inaccurate singers. Taken together, these results suggest that subvocalization may play a role in anticipatory auditory imagery, and possibly as a way of supplementing motor associations with auditory imagery.

KEYWORDS

auditory imagery, phonation, singing accuracy, sternohyoid muscle, subvocalization, surface electromyography (sEMG)

1 | INTRODUCTION

Singing in humans serves as a form of emotional expression (Juslin & Laukka, 2003), a mechanism for social bonding (Brown, 2000), a means to facilitate learning (Schön et al., 2008), a form of intrinsic enjoyment that may promote health (Judd & Pooley, 2014; Kreutz, Bongard, Rohrmann, Hodapp, & Grebe, 2004; Stewart & Lonsdale, 2016), and a vehicle for the treatment of neurological disorders (Wan, Rüber, Hohmann, & Schlaug, 2010). Despite the central role of singing in daily life, little is known about a critical process involved in singing: the vocal imitation of pitch. The vast majority of singing involves reproducing a melody from memory, most often based on an auditory representation (singing from notation alone, called sight singing, is less common and requires training). Although a good deal is known about the control of laryngeal muscles and auditory pitch perception, the transition from perception to vocal action planning is not well understood. This is a critical issue in music cognition given that most poor-pitch singers seem to suffer from deficient sensorimotor translation of this sort, rather than disorders specific to perceptual or to motor processes (Hutchins & Peretz, 2012; Pfordresher & Brown, 2007; Pfordresher & Mantell, 2014). Beyond the domain of music, vocal pitch imitation is an important component of language learning, particularly for tone languages (Kuhl, 2004) but also relevant to production of prosody in nontone languages.

We have recently proposed that audiovocal sensorimotor translation is driven by the formation of a multimodal mental image that integrates auditory and motor imagery (Greenspon, Pfordresher, & Halpern, 2017; Pfordresher & Halpern, 2013; Pfordresher, Halpern, & Greenspon, 2015). Neuroimaging studies have shown that motor planning areas are activated during auditory speech processing (Liebenthal,

*Portions of this research were presented at the 2015 meeting of the Society for Music Perception and Cognition.

Sabri, Beardsley, Mangalathu-Arumana, & Desai, 2013; Tremblay & Small, 2011), familiar (Herholz, Halpern, & Zatorre, 2012) and unfamiliar song listening (Brown & Martinez, 2007; Chen, Rae, & Watkins, 2012), physical actions (Gazzola, Aziz-Zadeh, & Keysers, 2006), and nonverbal vocalizations (McGettigan et al., 2013; Warren et al., 2006). Such results suggest that engaging in auditory imagery might also prime motor planning or vice versa, as happens when one prepares to sing a note or melody. Consistent with this idea, individuals who suffer from a vocal pitch imitation deficit (VPID), which affects the accuracy of sung pitch, report less vivid auditory imagery (Greenspon et al., 2017; Pfordresher & Halpern, 2013) and are less responsive to tasks that involve reproducing mental transformations of melodies (Greenspon et al., 2017). Singing in general may therefore draw on mental imagery to guide sensorimotor translation from a target (either just heard or stored in memory), and deficiencies in this process may underlie VPID and thus poor singing.

If auditory imagery used to prepare vocal motor responses truly engages motor planning, then it should be possible to observe subtle muscle movements at the periphery that are related to vocal motor production. Such activity is called subvocalization and, in the present context, may be considered a form of "covert" singing. During subvocalization, vocal muscles are engaged in the absence of any perceivable vocal production at a time when the participant is not intending to vocalize. Subvocalization has been observed via surface electromyography (sEMG) during reading (Hardyck & Petrinovich, 1970), as well as when trained musicians internally simulate the sound of a melody based on reading notation (Brodsky, Kessler, Rubinstein, Ginsborg, & Henik, 2008). Behavioral studies have shown interfering effects of task-irrelevant subvocalizations on auditory imagery tasks (Aleman & Van't Wout, 2004; Smith, Wilson, & Reisberg, 1995). Furthermore, studies of poor readers (Hardyck & Petrinovich, 1970) and remedial writers (Williams, 1987) have observed that these groups engage in more subvocalization compared to their normal performing counterparts. Yet, no studies have addressed whether a similar compensatory use of subvocalization occurs for music. Nor is it known whether auditory imagery elicits subvocalization more than other mental imagery tasks do, such as visual imagery, which may also be linked to vocal motor responses. We tested the possibility that subvocalization might facilitate pitch imitation generally, in which case the better pitch imitators would show more subvocalization. Finally, previous studies have focused solely on recording sites around the larynx and have overlooked the possible role of other facial muscles during auditory imagery. For instance, participants exhibit subtle facial muscle activity that mimics singers who are engaged in emotional song during both audiovisual exposure (Chan, Livingstone, & Russo, 2013) and in preparation to subsequently imitate the singer (Livingstone, Thompson, & Russo, 2009).

Accordingly, in the present research we measured muscle activity at various sites, including muscles engaged in control of pitch, while participants engaged in auditory or visual imagery. The auditory imagery task involved initial exposure to a melody, followed by an imagery period, and finally vocal reproduction of that melody. The visual imagery task involved initial exposure to an array of objects that resist verbal description ("greebles"; Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002; Tarr, 2016), followed by an imagery period in which participants retained the visual image in memory, and finally a probe question to assess the fidelity of their visual image. We also included trials designed to measure resting state of sEMG. Recording sites included the left and right sternohyoid muscles. These are laryngeal muscles that lower laryngeal cartilage and thus play a role in pitch control (Belyk & Brown, 2017; Roubeau, Chevrie-Muller, & Saint Guily, 1997; Stepp, 2012; Stepp et al., 2011; Vilkman, Sonninen, Hurme, & Körkkö, 1996). We elected to record activity at the lip and corrugator (i.e., eyebrow) muscles based on previous EMG studies showing activity at these sites in preparation for singing (Livingstone et al., 2009). Lastly, activity of the participant's nondominant bicep was recorded to reduce demand characteristics and capture spurious upper body movements. See Figure 1 for an illustration of all recording sites.

2 | METHOD

2.1 | Participants

Forty-six students from the University at Buffalo, SUNY, participated in exchange for course credit in introduction to psychology. The sample was predominantly young adult and musically inexperienced: The mean age was 19 years old (range: 18–24), 24 participants (52%) were female, mean years of instrumental and singing experience were, respectively, 2.5 years (max = 12) and 0.4 years (max = 7). Only

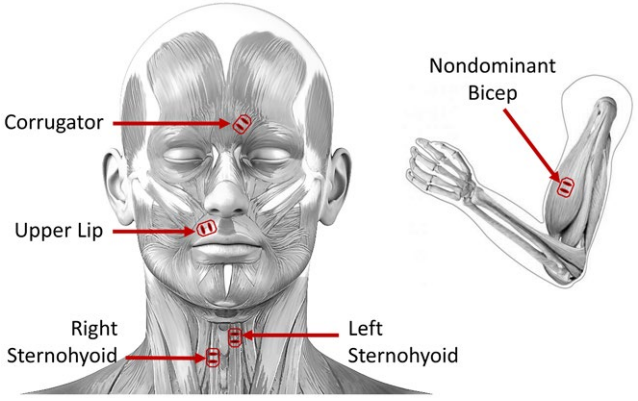

FIGURE 1 Placement sites for surface electromyographic sensors

five participants (10%) had more than 1 year of singing experience, and in every case this experience was participation in a choir rather than private vocal training. Participants were screened to represent a wide range of vocal pitch matching ability, in order to assess contributions of muscle activity to different levels of success.

2.2 | Materials and equipment

In order to compare musical and visual imagery directly, we designed tasks to have similar overall trial structures with temporally comparable imagery periods. Figure 2 illustrates the time course of the two imagery tasks. The auditory imagery task (Figure 2a) required participants to listen to a four-note target melody and then imagine the just-heard melody for 4 s. After this imagery period, a probe question appeared in which the participant reported the vividness of their musical image on a scale from 1 (*no sound*) to 5 (*like real sound*). After their response, the participant sang the melody aloud on the syllable "doo" (/dʊ/).

Eight target melodies (previously used in Pfordresher & Brown, 2007) were created in Praat (Boersma & Weenink, 2013) by concatenating single note recordings of trained male and female vocalists singing on the syllable /dʊ/. We equated amplitude across notes to reduce the perception of intensity contours. The presentation rate of each note within a melody

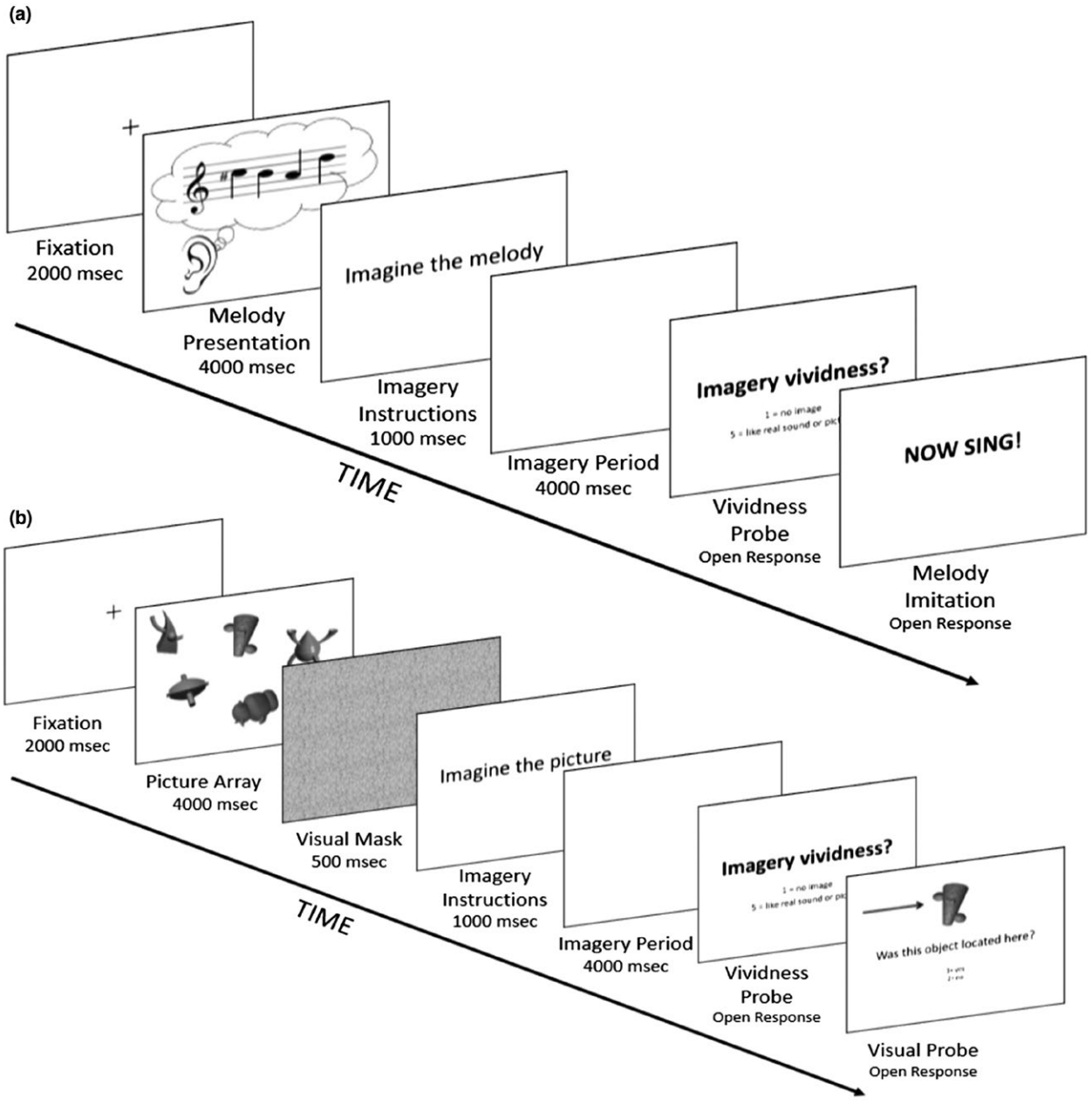

FIGURE 2 Illustration of trial phases for auditory imagery (a) and visual imagery (b) tasks

was set at an interonset interval of 1,000 ms (60 bpm). Both male and female voice stimuli were created so participants would imitate melodies based on a model of their own pitch range. Moreover, target melodies were centered at different musical keys so as to be close to participants' comfort pitches (male: A2, D3, F3; female: F3, A3, D4). Participants imitated only melodies with musical keys closest to their comfort pitch to minimize possible vocal strain resulting from singing outside of their vocal range.

The visual imagery task (Figure 2b) required participants to study a picture of five novel objects for 4 s, followed by the presentation of a visual mask for 500 ms. Participants were instructed to then imagine the picture they had just seen for 4 s. After this imagery period, the participant reported the vividness of his or her image on a scale from 1 (*like no picture at all*) to 5 (*like real picture*). Finally, participants were presented with a single object from the original array and reported whether the object was presented in its original position.

Sixteen visual stimuli were used in this task. Each visual stimulus was composed of five objects placed at different locations in the picture. The objects were selected and adapted from a repository of asymmetric greebles and complex geons (Tarr, 2016). These objects were used to minimize the participant's use of verbal labels that could be subvocalized during the imagery period. Half of the visual trials presented probe questions that showed the same object in its correct position from the original array. The other half of visual probe questions used an object that was in the original array, but in the incorrect location.

sEMG was acquired via the Trigno Mini Wireless system (Delsys Trigno Wireless EMG Systems, Boston, MA), which comprises single-differential, parallel-bar EMG sensors (25 mm × 12 mm × 7 mm, see Figure 3) and four silver

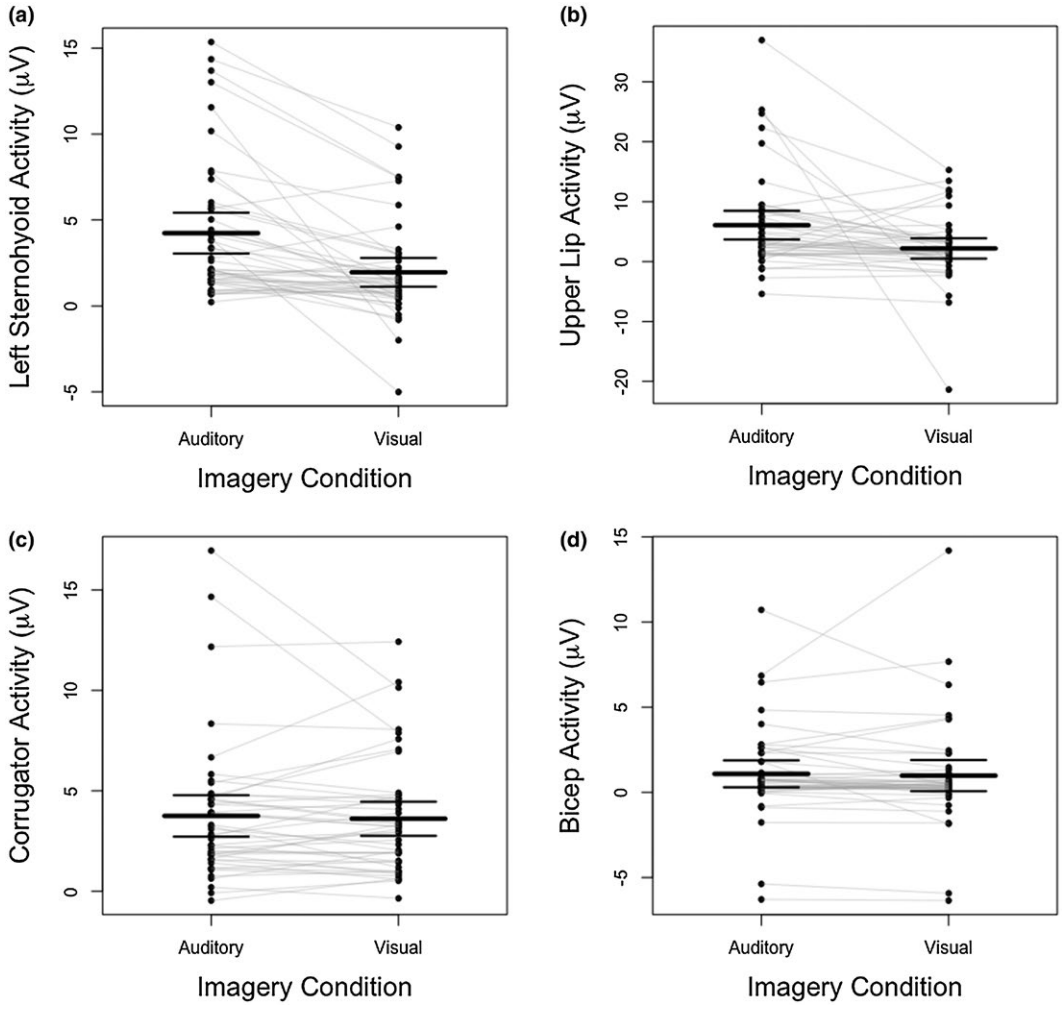

FIGURE 3 Spaghetti plots illustrating the sEMG activity contrasts (e.g., imagery minus rest) across imagery conditions for left sternohyoid (a), upper lip (b), corrugator (c), and bicep (d) sensors. Closed circles represent individual participant means. Bold horizontal lines correspond to the mean of each condition. Lines above and below the mean line represent upper and lower limits of 95% CIs. *Y* axes differ in magnitude due to differences in muscle morphology, where larger muscles (e.g., upper lip) recruit more motor units when activated than smaller muscles (e.g., corrugator)

contacts for local electrical reference in the main sensors' body. sEMG data were converted from analog to digital at a sampling rate of 1,925 Hz with 16-bit resolution using the EMG Works Acquisition and Analysis program (Delsys, Boston, MA).

Audio recordings were captured in a WhisperRoom SE 2000 sound-attenuated booth (Whisper Room Inc., Morristown, TN) at a sampling rate of 22,050 Hz with 16-bit resolution using a Shure PG58 dynamic microphone connected to a Lexicon Omega preamplifier and digitally stored as.wav files. A Dell computer with a 3.6 GHz processor ran Matlab (Mathworks, Inc., Natick, MA) for stimulus presentation and vocal data acquisition purposes. Visual cues throughout the experiment were displayed on a Dell 15-inch LCD computer screen placed directly in front of the participants. Auditory stimuli were played through a pair of Mackie CR3 series Multimedia Monitors (LOUD Technologies, Woodinville, WA), which flanked the LCD computer screen on each side.

During the experiment, participants were seated in a semireclined position in a comfortable chair while resting their head on the chair's headrest. This posture helped keep their head stable in order to maintain a static head position in order to facilitate sEMG recordings.

2.3 | Procedure

2.3.1 | Screening task

All participants were screened about 1 week prior to the main experimental session. We wanted to ensure that participants reflected a broad range of accuracy of pitch matching in singing, and to eliminate participants with hearing deficits as well as participants from whom reliable sEMG data could not be obtained (e.g., individuals with facial hair).

The screening task included four phases and was implemented using in-house MATLAB programs. First, participants completed a series of vocal warm-up tasks including singing the song "Happy Birthday to You" from memory in the key of their choosing. Participants then selected and sang a single pitch they felt was comfortable producing, to determine their comfort pitch. Next, participants listened to and then sung eight different four-note melodies. Each melody was similar to those used in the primary experiment (described above). The experimenter scored the accuracy of each sung performance by using a MATLAB plot that displayed the participant's F0 relative to the target melody, with boundaries of ±50 cents surrounding each note. Any pitch with more than 50% of the trace falling outside these boundaries was scored as an error. Third, participants completed an adaptive pitch discrimination task modeled after Loui, Gunther, Mathys, and Schlaug (2008) designed to identify their discrimination threshold for pitch. Any participant with a threshold less than 200 cents qualified to participate in the main experiment. We elected to use this criterion because the initial comparison of the screening's discrimination task was a 300-cent difference between pitches, and more fine-grained changes were under 200 cents. Thus, participants who exhibited a threshold at or greater than 300 cents were likely either guessing or did not understand the task. Of the 277 participants who were screened, only 30 (10.83%) individuals exhibited thresholds greater than the 200-cent criterion. Finally, participants completed the Bucknell Auditory Imagery Scale (BAIS), which is a self-report measure in which participants form an auditory image and then rate the vividness of that image (vividness subtest), and then attempt to alter that image and rate the ease with which they can do that (control subtest). Imagery items in the BAIS include music, speech, and environmental sounds (Halpern, 2015).

After the screening task, the experimenter invited any eligible participant to receive more course credit by participating in the main experimental task in the following week.[1] The experimenter showed the sEMG equipment to participants and described the experience of being in such a study so that participants could agree to participate with as much knowledge as possible about the procedure.

2.3.2 | Experimental task

Upon arrival, participants were outfitted with the sEMG sensors. Anatomical sites were cleaned with 70% alcohol pads in order to exfoliate dead skin and ensure secure sensor attachment. Sensors were then affixed to the skin using customized, double-sided adhesives (Trigno adhesive, Delsys).

As shown in Figure 1, we recorded sEMG from five sites. The most critical sites were those associated with laryngeal control of pitch on the left and right sternohyoid muscle (m. sternohyoideus). Although laryngeal muscles such as the cricothyroid play a more direct role in pitch control (Ludlow, 2005), its positioning behind other muscles and cartilage precludes reliable surface recording. By contrast, the sternohyoid muscle has the advantage of being superficially positioned in the neck, which limits the degree of signal contamination from other extrinsic laryngeal muscles (Stepp, 2012). Following Stepp, Hillman, and Heaton (2010), we positioned each sensor by first identifying the space between the thyroid and cricoid cartilages of the larynx. For the left sternohyoid sensor, we then moved 1 cm lateral and 1 cm superior to that reference point. The right sternohyoid sensor's placement was 1 cm lateral and 1 cm inferior to the reference

[1]Fifty-one participants completed the experimental task of this study. However, four participants were mistakenly invited and took part in the experiment despite their pitch discrimination thresholds exceeding the 200-cent criterion. Another single participant's data were excluded from analysis due to poor sEMG signal quality. Thus, 90% of the data collected from the sEMG experiment are reported in Results.

point.[2] Reference sensors for both left and right sternohyoid sEMG were placed on the corresponding clavicle. We measured activity on the right upper lip (m. orbicularis oris superioris) via a sensor positioned just lateral of the philtrum and adjacent to the vermilion border. The reference sensor was placed on the mastoid process located behind the participant's right ear. Using Fridlund and Cacioppo's (1986) protocols, the corrugator sensor (m. corrugator supercilii) was placed directly above the medial end of the left eyebrow with the reference sensor placed behind the left ear on the mastoid process. Finally, we measured activity in the participant's nondominant bicep muscle (m. biceps brachii) as a further control measure, as shown in Figure 1. Once the sensors were affixed to the participant, he or she was then seated in the sound booth and instructed to remain as still as possible in the chair throughout the session. The participant then completed vocal warm-up trials similar to those used in the screening procedure.

The primary experimental tasks followed. The participants completed 64 trials involving mental imagery, with visual and auditory trials randomly intermingled such that no more than four trials in a row involved the same imagery modality. These trials were arranged into blocks of 16 trials, and after every block there was a rest trial in which the participant was instructed to sit quietly and not move for 30 s. Rest trials were used as control trials in order to obtain baseline measures of sEMG activity throughout the experiment.

2.4 | Data processing

2.4.1 | sEMG data processing

On a trial-by-trial basis, the sEMG signal was converted from volts to microvolts (μV), and then a de-trending procedure was used to remove potential direct current offsets. Merletti and Hermens (2004) indicate that typical sEMG signals power frequency ranges from 0–450 Hz. However, movements create artifacts within the 0–20 Hz range, so it is advisable to apply a high-pass filter with cutoffs around 10–20 Hz (Stepp, 2012). Therefore, the signals acquired in the current experiment were smoothed with a Butterworth band-pass filter with a 20–450 Hz bandwidth. We then applied an infinite impulse response (IIR) notch filter centered at 60 Hz to remove potential contaminant signals from electrical power lines. The signal was then full-wave rectified, and research assistants visually inspected trials for remaining motion artifacts. Motion artifacts were determined by identifying cases of extreme upward deflections in the sensor's signal. In such cases, the onset and offset of the extreme peak value was marked and the signal's data were removed between these two boundaries. Trials contaminated with multiple (i.e., five or more) motion artifacts were completely removed from subsequent analysis (total of 0.1% of all trials). The experimenter also documented instances where participants coughed, sneezed, cleared their throats, or erroneously vocalized during the imagery portion of a trial. Approximately 0.9% of all trials were excluded based on these criteria. The remaining trials were divided into imagery and singing phases based on the timing of different trial phases and, where necessary, sEMG activity (which is more prominent during singing). Lastly, a linear envelope of the sEMG signal was created by passing a 250-ms wide moving average window across the duration of the trial. The maximum value within each 250-ms window was recorded and averaged across the duration of the trial phase to estimate peak muscle activity.

sEMG data from rest trials were processed in a similar fashion. The 250-ms wide window passed over the first 5 s of the signal, which was equal to the length of the auditory and visual tasks' imagery phases. The maximum values from the rest condition windows were aggregated across the four trials. Rest condition data were then used to generate contrasts by subtracting rest activity from task-specific activity (e.g., auditory imagery minus rest condition). This was conducted for normalization purposes in order to control for potential morphological differences between participants.

An initial inspection of the data revealed that the right sternohyoid sensor yielded very noisy activations. This was likely due to the right sensor's inferior location relative to the left sensor, where its lower placement on the neck is positioned over more subcutaneous fat, which distorts sEMG signals. More problematic was the consistent presence of cardiac pulse contaminations in the right sensor due to the close proximity to the anterior jugular vein. Based on this assessment, we focused analyses on the left sternohyoid sensor.

2.4.2 | Auditory data processing

Sung F0 was extracted from digital audio files using the pitch tracking algorithm, YIN (De Cheveigné & Kawahara, 2001), which runs on MATLAB. Boundaries between sung notes were then identified via a semiautomated MATLAB procedure in which initial estimates were determined based on fluctuations in vocal intensity that are associated with syllabification, followed by any necessary manual corrections by the experimenter. The pitch of each sung note was then estimated based on the median F0 in the central 50% of samples between note boundaries, in order to exclude contamination from vocal "scoops" at the beginning or end of sung notes. Sung pitch errors were based on the absolute difference between each of these pitch values and the target pitch for each note. Absolute differences that were greater than 50 cents (half a semitone) were classified as errors.

[2]Although Stepp (2012) suggests a superior position relative to the larynx for the sternohyoid's sensor, we elected to use an inferior placement of the right sensor for comparative purposes. We ultimately found that the right sensor's data contained more noisy signals, either from cross talk from muscles below the sternohyoid or the sensor's proximity to the right carotid artery.

3 | RESULTS

We first determined whether sEMG activity differed across auditory and visual imagery conditions, as well as baseline activity during rest, using a within-subject analysis of variance (ANOVA), computed separately for each sensor, followed by pairwise tests between means based on familywise $\alpha = 0.05$ and significance determined using the Holm-Bonferroni correction (see Table 1). Mauchley's test indicated violations of sphericity for each ANOVA, but all reported effects remained significant after applying the Greenhouse-Geisser corrections. We did not compare sensors directly at this initial stage of analysis, based on the possibility that spurious statistical effects would emerge based on simple differences in the target muscle morphologies. The upper lip sensor did not yield reliable data for two participants, and analysis of effects at this sensor were conducted with the remaining 44 participants.

In the second analysis stage, we computed difference scores for each participant and condition by contrasting sEMG during each imagery condition to activity during the rest condition. These differences were compared across sensors directly while controlling for variability due to morphological differences across sensors and individuals. A two-way repeated measures ANOVA was conducted to verify sensor-specific effects of imagery conditions by observing a significant Sensor × Task interaction. The results produced a significant main effect of imagery task, $F(1, 45) = 14.91$, $p < 0.001$, $\eta_p^2 = 0.26$. There was a significant main effect of sensor as well, $F(3, 132) = 9.28$, $p < 0.001$, $\eta_p^2 = 0.18$. Mauchly's test determined sphericity violations, but this main effect of sensor remained significant after corrections ($\varepsilon = 0.71$, $p < 0.001$). This was an unsurprising main effect given the anticipated use of these target muscles in the experimental tasks. Nevertheless, a series of pairwise tests using the Holm-Bonferroni correction verified that this main effect resulted from differences between the bicep's sensor activities with all other sensors ($ps < 0.01$). No other sensor activities were found to be significantly different from one another ($ps > 0.37$). Crucially, there was a significant Sensor × Task interaction, $F(3, 132) = 6.47$, $p < 0.001$, $\eta_p^2 = 0.13$, that remained significant after the sphericity correction ($\varepsilon = 0.43$, $p < 0.001$). The presence of this interaction suggests that sEMG activity differed at specific sensor sites across the visual and auditory imagery tasks.

To tease apart these sensor-specific effects, we then conducted a series of planned contrasts between visual and auditory imagery condition within each sensor (see Figure 3). Left sternohyoid activity was significantly greater during auditory ($M = 4.24$, $SD = 4.00$) than visual imagery ($M = 1.96$, $SD = 2.82$), $t(45) = 4.98$, $p < 0.001$, $d = 0.67$. Likewise, lip sEMG activity was greater during auditory ($M = 6.30$, $SD = 8.17$) than visual ($M = 2.12$, $SD = 5.84$) imagery, $t(43) = 2.82$, $p = 0.007$, $d = 0.58$. However, corrugator ($p = 0.61$) and bicep ($p = 0.64$) activity did not differ as a function of imagery condition.

Given imagery's introspective nature, it was necessary to verify that participants were indeed utilizing imagery in their completion of these tasks. Recall that on each trial participants were instructed to self-report their imagery's

TABLE 1 Summary of statistical analyses examining sEMG activity during auditory imagery, visual imagery, and rest conditions

Sensor and imagery comparison	*t*	*p*	*F*	df_1, df_2	Greenhouse-Geisser ε	Corrected *p*	η_p^2
L. sternohyoid			36.88	2, 90	0.838	0.000	0.45
Auditory vs. rest	7.17	0.000					
Visual vs. rest	4.70	0.000					
Auditory vs. visual	4.94	0.000					
Upper lip			13.96	2, 86	0.784	0.000	0.25
Auditory vs. rest	5.12	0.000					
Visual vs. rest	2.52	0.047					
Auditory vs. visual	2.82	0.002					
Corrugator			52.80	2, 90	0.709	0.000	0.54
Auditory vs. rest	7.32	0.000					
Visual vs. rest	8.55	0.000					
Auditory vs. visual	0.52	0.999					
Bicep			5.21	2, 90	0.696	0.016	0.11
Auditory vs. rest	2.76	0.025					
Visual vs. rest	2.15	0.109					
Auditory vs. visual	0.46	0.999					

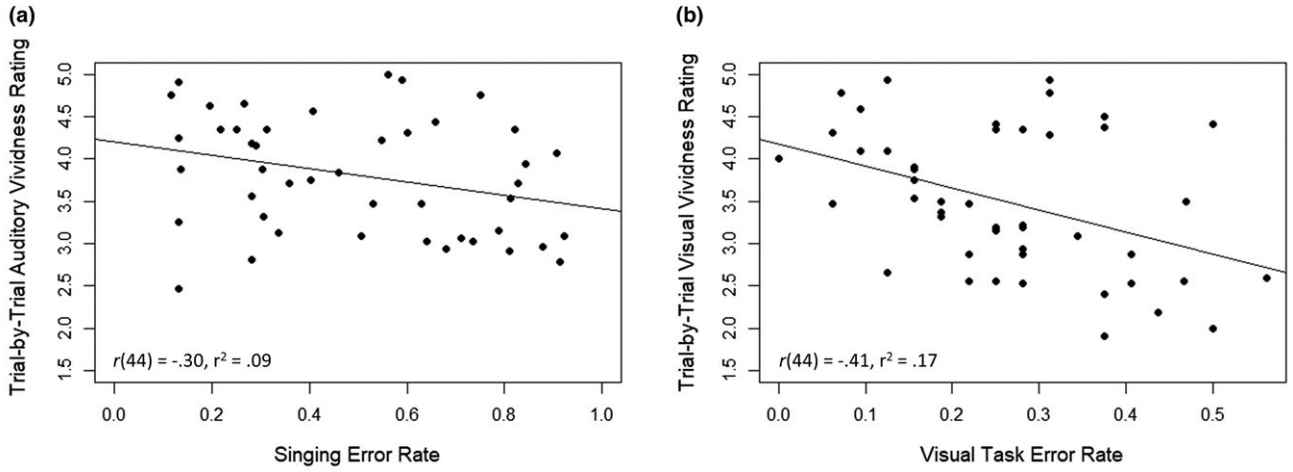

FIGURE 4 Scatter plots of the relationship between task performance (*x*) and trial-by-trial vividness rating (*y*). (a) Relationship between proportion of sung note errors and mean auditory trial vividness ratings. (b) Relationship between proportion of visual task errors and mean visual trial vividness ratings. Each point represents the mean across trials for an individual

vividness on a scale from 1 (*low vividness*) to 5 (*high vividness*). The central issue at hand involves how such vividness self-reports are related to behavioral measures. A positive correlation between vividness ratings and behavioral performance suggests that the effective use of imagery contributes to the success of mental rehearsal. Figure 4 depicts the association between participants' mean vividness rating and their task performance. As can be seen in Figure 4a, there was a significant negative correlation between auditory vividness ratings and proportion of sung note errors, $r(44) = -0.30$, $p = 0.04$, $r^2 = 0.09$. Poor auditory imagery vividness on a particular trial was associated with inaccurate singing accuracy. A similar relationship emerged between mean visual vividness rating and proportion of visual task errors (see Figure 4b), $r(44) = -0.41$, $p = 0.004$, $r^2 = 0.17$. Less vivid visual images were associated with poorer visual task performance. Taken together, the significant associations between imagery vividness and task performance in both imagery tasks suggest the employment of mental imagery in navigating these tasks. Such results cohere with previous findings showing that poorer auditory imagery is associated with poor-pitch imitation (Pfordresher & Halpern, 2013) and trial-by-trial vividness ratings correlate with neural activity (Leaver, Van Lare, Zielinski, Halpern, & Rauschecker, 2009).

As discussed earlier, a critical question in the present research was whether subvocalization relates to singing accuracy, based on the theory that VPID partly originates from problems in generating a multimodal image of the sequence. We focused on the left sternohyoid and lip sensors for this analysis given that these sensors were shown to be significantly more active during auditory imagery compared to other conditions. Moreover, these muscles have an integral role in overt vocalization. We first calculated the correlation between the auditory imagery contrast (imagery activity minus rest activity) for both sensors and the proportion of sung note errors. The correlation between left sternohyoid imagery activity and singing error rates was statistically significant, $r(44) = 0.30$, $p = 0.04$, $r^2 = 0.09$ (see Figure 5a), suggesting that inaccurate singers engage in more covert laryngeal activity during auditory imagery. However, upper lip activity during imagery was not significantly associated with error rates ($p = 0.59$). Neither of the other correlations of sEMG sensor with sung error rates during singing were significant ($ps > 0.20$ in both cases).

We calculated the same series of correlations between sEMG activity during imagery at each sensor and visual task error rate. Crucially, we found that left sternohyoid activity during visual imagery was not associated with visual task performance, $r(42) = 0.01$, $p = 0.94$, $r^2 = 0.00$ (see Figure 5b), which highlights the selectivity of this muscle during auditory imagery. Likewise, we found no significant correlation between the upper lip activity during visual imagery and visual task performance, $r(42) = 0.07$, $p = 0.61$, $r^2 = 0.00$. Lastly, visual task error rates were not correlated with visual imagery corrugator activity ($p = 0.64$) or bicep activity ($p = 0.49$).

4 | DISCUSSION

We have reported, for the first time, evidence that subvocalization plays a role in auditory imagery used during mental rehearsal of a melody before vocal production for those muscles that play the most central role in vocal pitch production. We also report evidence that this activity is related to singing accuracy, with less accurate singers

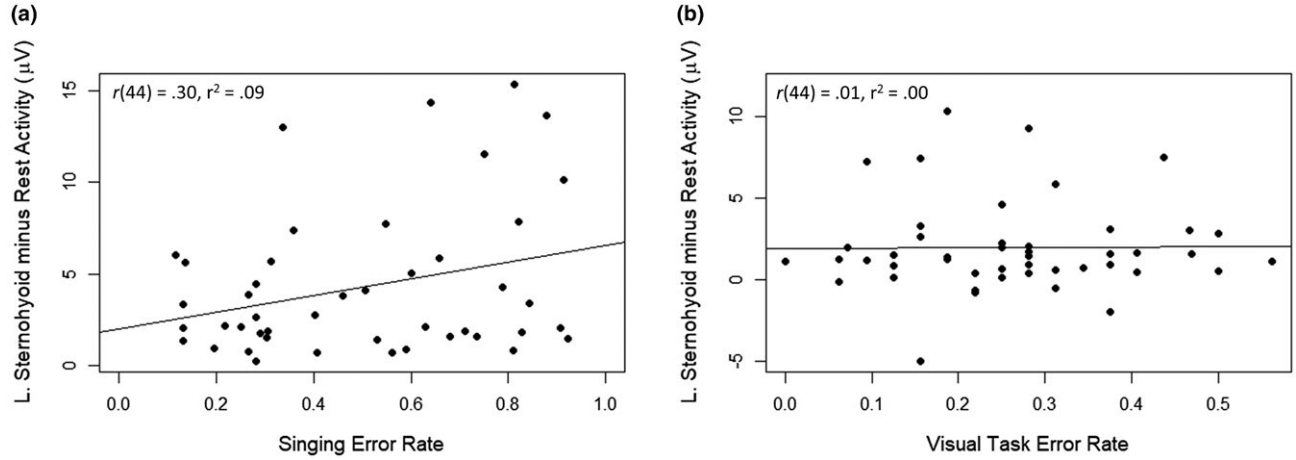

FIGURE 5 (a) Relationship between proportion of pitch errors (*x*) and sEMG auditory imagery contrast (auditory minus rest activity) of the left sternohyoid muscle (*y*). (b) Relationship between proportion of visual task errors (*x*) and sEMG visual imagery contrast (visual minus rest activity) of the left sternohyoid muscle. Each dot represents the mean across trials for an individual

engaging in larger subvocal muscle contractions than accurate singers.

A basic but important contribution of this work is to show that sEMG can be used to measure even subvocal responses. Although some past studies have measured subvocalization using sEMG, this is the first study we know of to demonstrate greater subvocal activity with this measure for auditory as opposed to visual imagery. This is a valuable contribution for researchers who want to measure subvocal responses without using hook-wire electrodes, which require invasive intramuscular implantation. sEMG also has an advantage over electroglottography (quantifying vocal fold contact) in that sEMG can measure both overt and covert (i.e., subvocal) responses. Furthermore, these results suggest that measurements of the sternohyoid muscle can reliably detect activity associated with pitch control. This is important because intrinsic laryngeal muscles primarily control the adduction and abduction of the vocal folds, whereas extrinsic muscles such as the sternohyoid exhibit indirect effects by altering the elevation of the larynx in the neck (Ludlow, 2005; Vilkman et al., 1996). However, the superficial location of the sternohyoid makes it more easily accessible for sEMG, and thus it is important to show that this muscle can measure phonatory activity at a gross level.

A second contribution of this research is in demonstrating a relationship between auditory imagery and peripheral muscle movements. This adds further support for the view that auditory imagery elicits motor planning activity in the brain, leading to a multimodal representation. As such, mental imagery can be conceptualized as involving multiple components, some unimodal and others multimodal (McNorgan, 2012). The extent to which individuals differ in their degree of central multimodal representations is a question for future study. We here show that individuals do differ with respect to peripheral engagement of vocal muscles during auditory imagery. The present results demonstrate that subvocalization, which also occurs during "notational audiation" (a form of auditory imagery used when reading music notation rather than emerging from memory; Brodsky et al., 2008) is a general phenomenon and not the result of an unusual skill learned only by expert musicians.

The selective nature of these results is worth further consideration. Although it is not surprising that the bicep muscle was not active during auditory imagery, it seems plausible that we might have found differing activity of the corrugator muscle. Huron and colleagues (Huron & Shanahan, 2013; Huron, Dahl, & Johnson, 2009) showed that, as participants raise or lower the pitch of their voice, their eyebrows rise and fall with the corresponding directional changes. These authors proposed that a common central motor process may control movements of the eyebrow and vocal folds. As such, one would expect a similar relationship to emerge during the covert rehearsal of a melody. However, the results reported here showed no difference in corrugator activity across auditory and visual imagery conditions. The absence of such differences may reflect a more specified use of these muscles for overt singing, and it is possible that these movements are not necessary for imagining and planning phonatory gestures.

The upper lip muscles were selectively active during auditory imagery, but there was no relationship between this activity and singing accuracy. Although the lips have an inherent role in vocalization, they do not have a critical role in pitch control. The fact that lip movements during auditory imagery did not predict singing accuracy may in part reflect the lack of articulatory variability across syllables in sung melodies, where each sung note was produced using the syllable /dʊ/. If participants were to instead imagine tunes with lyrics or melodies containing phonetic variations, we would expect there to be a stronger relationship between lip movement and singing accuracy, particularly given that pairing

pitch and syllable information leads to memory enhancements (Berkowska & Dalla Bella, 2009; Racette & Peretz, 2007; Schön et al., 2008).

An interesting though surprising result was the fact that inaccurate singers engaged in more subvocalization than did accurate singers. This finding aligns with neuroimaging observations that show expertise garners processing efficiencies, which often correspond to decreases in cortical activity (cf. Chen et al., 2012; Kelly & Garavan, 2004). As such, an accurate singer's skill affords similar efficiencies that may not require the use of subvocalization when completing our imagery and pitch imitation tasks. However, other researchers have shown that expert efficiency gains and reduced cortical activity are sometimes dependent on experimental task (Landau & D'Esposito, 2006). We therefore set out with a two-tailed prediction for this novel area of research, and the obtained results go against a parsimonious assumption that auditory imagery influences motor planning in a unidirectional manner. In such a framework, the formation of auditory imagery leads directly to an associated motor image, with the strength of that association being determined by the accuracy and precision of sensorimotor mapping (Pfordresher et al., 2015). Peripheral muscle activity then reflects this unidirectional mapping and is stronger when the mapping of auditory imagery to motor imagery is more accurate and precise.

The present data, however, suggest that people may engage in both auditory and motor imagery, and that inaccurate imitators may use motor imagery as a way of trying to enact a vague auditory image. It is also possible that this type of enactment actually interferes with the efficient functioning of the sensorimotor loop. This hypothesis could be tested by adding a condition where the laryngeal movements are suppressed or blocked altogether. In short, multimodal imagery may involve bidirectional activations of imagery across associated auditory and motor representations, even when an individual may only be consciously aware of forming an image within one modality. These bidirectional associations may serve a strategic purpose. However, the mapping between modalities may be imprecise for a given person, or on a given trial, leading to variability in primary motor activation during imagery tasks in neuroimaging studies (de Lange, Roelofs, & Toni, 2008; McNorgan, 2012). We note that this use for subvocalization is not without precedent. More subvocal activity is found during reading among poor readers as well as when reading difficult passages (Hardyck & Petrinovich, 1970). Likewise, Williams (1987) found greater subvocal activity during writing tasks in participants with below-average language skills.

Overall, the results of this study further illustrate the link between subvocal activity and auditory imagery for musical stimuli. More important is the fact that this research shows that subvocalization associated with auditory imagery is not limited to conditions of reading musical score nor is the product of specialized musical training. Furthermore, this study demonstrates such laryngeal activity can be captured using sEMG, which lays the groundwork for future research examining subvocalization in a variety of contexts.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation grant BCS-1256964 and was conducted in compliance with the guidelines defined by the University at Buffalo's Institutional Review Board. We would like to express our gratitude to our visiting undergraduate research assistants from Bucknell University, Taylor Reeh and Gabby Gottschall, who were instrumental in the formulation of the study, and who were supported by the REU portion of the NSF grant. We thank Dr. Chris McNorgan and Dr. Larry Hawk for their valuable guidance and consultation on methodology and analytical techniques. Lastly, we deeply appreciate the hard work by the following undergraduates for collecting and processing the data used in this research: Andrew Kothan, Cailin Shupbach, Thomas Gadelrab, Anthony Nagib, Garnettha Sari, and Karen Li.

REFERENCES

Aleman, A., & Van't Wout, M. (2004). Subvocalization in auditory-verbal imagery: Just a form of motor imagery? *Cognitive Processing*, *5*(4), 228–231. https://doi.org/10.1007/s10339-004-0034-y

Belyk, M., & Brown, S. (2017). The origins of the vocal brain in humans. *Neuroscience & Biobehavioral Reviews*, *77*, 177–193. https://doi.org/10.1016/j.neubiorev.2017.03.014

Berkowska, M., & Dalla Bella, S. (2009). Reducing linguistic information enhances singing proficiency in occasional singers. *Annals of the New York Academy of Sciences*, *1169*(1), 108–111. https://doi.org/10.1111/j.1749-6632.2009.04774.x

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer* (Version 5.3. 51) [Computer program]. Amsterdam, Netherlands: Institute of Phonetic Sciences/University of Amsterdam. Retrieved from https://www.praat.org

Brodsky, W., Kessler, Y., Rubinstein, B. S., Ginsborg, J., & Henik, A. (2008). The mental representation of music notation: Notational audiation. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 427–445. https://doi.org/10.1037/0096-1523.34.2.427

Brown, S. (2000). Evolutionary models of music: From sexual selection to group selection. *Perspectives in ethology* (pp. 231–281). Boston, MA: Springer.

Brown, S., & Martinez, M. J. (2007). Activation of premotor vocal areas during musical discrimination. *Brain and Cognition*, *63*(1), 59–69. https://doi.org/10.1016/j.bandc.2006.08.006

Chan, L. P., Livingstone, S. R., & Russo, F. A. (2013). Facial mimicry in response to song. *Music Perception: An Interdisciplinary Journal*, *30*(4), 361–367. https://doi.org/10.1525/mp.2013.30.4.361

Chen, J. L., Rae, C., & Watkins, K. E. (2012). Learning to play a melody: An fMRI study examining the formation of auditory-motor associations. *NeuroImage*, *59*(2), 1200–1208. https://doi.org/10.1016/j.neuroimage.2011.08.012

De Cheveigné, A. D., & Kawahara, H. (2001). *Comparative evaluation of F0 estimation algorithms*. In 7th European Conference on Speech Communication and Technology, Aalborg, Denmark.

de Lange, F. P., Roelofs, K., & Toni, I. (2008). Motor imagery: A window into the mechanisms and alterations of the motor system. *Cortex*, *44*(5), 494–506. https://doi.org/10.1016/j.cortex.2007.09.002

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, *23*(5), 567–589. https://doi.org/10.1111/j.1469-8986.1986.tb00676.x

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology*, *16*(18), 1824–1829. https://doi.org/10.1016/j.cub.2006.07.072

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

Halpern, A. R. (2015). Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain*, *25*(1), 37–47. https://doi.org/10.1037/pmu0000081

Hardyck, C. D., & Petrinovich, L. F. (1970). Subvocal speech and comprehension level as a function of the difficulty level of reading material. *Journal of Verbal Learning and Verbal Behavior*, *9*(6), 647–652. https://doi.org/10.1016/S0022-5371(70)80027-5

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience*, *24*(6), 1382–1397. https://doi.org/10.1162/jocn_a_00216

Huron, D., Dahl, S., & Johnson, R. (2009). Facial expression and vocal pitch height: Evidence of an intermodal association. *Empirical Musicology Review*, *4*(3), 93–100. https://doi.org/10.18061/1811/44530

Huron, D., & Shanahan, D. (2013). Eyebrow movements and vocal pitch height: Evidence consistent with an ethological signal. *Journal of the Acoustical Society of America*, *133*(5), 2947–2952. https://doi.org/10.1121/1.4798801

Hutchins, S. M., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology: General*, *141*(1), 76–97. https://doi.org/10.1037/a0025064

Judd, M., & Pooley, J. A. (2014). The psychological benefits of participating in group singing for members of the general public. *Psychology of Music*, *42*(2), 269–283. https://doi.org/10.1177/0305735612471237

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kelly, A. C., & Garavan, H. (2004). Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex*, *15*(8), 1089–1102. https://doi.org/10.1093/cercor/bhi005

Kreutz, G., Bongard, S., Rohrmann, S., Hodapp, V., & Grebe, D. (2004). Effects of choir singing or listening on secretory immunoglobulin A, cortisol, and emotional state. *Journal of Behavioral Medicine*, *27*(6), 623–635. https://doi.org/10.1007/s10865-004-0006-9

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Landau, S. M., & D'Esposito, M. (2006). Sequence learning in pianists and nonpianists: An fMRI study of motor expertise. *Cognitive, Affective, & Behavioral Neuroscience*, *6*(3), 246–259. https://doi.org/10.3758/CABN.6.3.246

Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *Journal of Neuroscience*, *29*(8), 2477–2485. https://doi.org/10.1523/JNEUROSCI.4921-08.2009

Liebenthal, E., Sabri, M., Beardsley, S. A., Mangalathu-Arumana, J., & Desai, A. (2013). Neural dynamics of phonological processing in the dorsal auditory stream. *Journal of Neuroscience*, *33*(39), 15414–15424. https://doi.org/10.1523/JNEUROSCI.1511-13.2013

Livingstone, S. R., Thompson, W. F., & Russo, F. A. (2009). Facial expressions and emotional singing: A study of perception and production with motion capture and electromyography. *Music Perception: An Interdisciplinary Journal*, *26*(5), 475–488. https://doi.org/10.1525/mp.2009.26.5.475

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology*, *18*(8), R331–R332. https://doi.org/10.1016/j.cub.2008.02.045

Ludlow, C. L. (2005). Central nervous system control of the laryngeal muscles in humans. *Respiratory Physiology & Neurobiology*, *147*(2–3), 205–222. https://doi.org/10.1016/j.resp.2005.04.015

McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., & Scott, S. K. (2013). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cerebral Cortex*, *25*(1), 246–257. https://doi.org/10.1093/cercor/bht227

McNorgan, C. (2012). A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. *Frontiers in Human Neuroscience*, *6*, 285. https://doi.org/10.3389/fnhum.2012.00285

Merletti, R., & Hermens, H. J. (2004). Detection and conditioning of the surface EMG signal. In R. Merletti & P. Parker (Eds.), *Electromyography: Physiology, engineering and non-invasive applications* (pp. 107–131). Hoboken, NJ: Wiley-IEEE.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of tone deafness. *Music Perception: An Interdisciplinary Journal*, *25*(2), 95–115. https://doi.org/10.1525/mp.2007.25.2.95

Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review*, *20*(4), 747–753. https://doi.org/10.3758/s13423-013-0401-8

Pfordresher, P. Q., Halpern, A. R., & Greenspon, E. B. (2015). A mechanism for sensorimotor translation in singing: The multimodal imagery association (MMIA) model. *Music Perception: An Interdisciplinary Journal*, *32*(3), 242–253. https://doi.org/10.1525/mp.2015.32.3.242

Pfordresher, P. Q., & Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive Psychology*, *70*, 31–57. https://doi.org/10.1016/j.cogpsych.2013.12.005

Racette, A., & Peretz, I. (2007). Learning lyrics: To sing or not to sing? *Memory & Cognition*, *35*(2), 242–253. https://doi.org/10.3758/BF03193445

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science*, *13*(3), 250–257. https://doi.org/10.1111/1467-9280.00446

Roubeau, B., Chevrie-Muller, C., & Saint Guily, J. L. (1997). Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta Oto-laryngologica*, *117*(3), 459–464. https://doi.org/10.3109/00016489709113421

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, *106*(2), 975–983. https://doi.org/10.1016/j.cognition.2007.03.005

Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, *33*(11), 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*(4), 1232–1246. https://doi.org/10.1044/1092-4388(2011/11-0214)

Stepp, C. E., Heaton, J. T., Stadelman-Cohen, T. K., Braden, M. N., Jetté, M. E., & Hillman, R. E. (2011). Characteristics of phonatory function in singers and nonsingers with vocal fold nodules. *Journal of Voice*, *25*(6), 714–724. https://doi.org/10.1016/j.jvoice.2010.06.003

Stepp, C. E., Hillman, R. E., & Heaton, J. T. (2010). The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research*, *53*(5), 1220–1226. https://doi.org/10.1044/1092-4388(2010/09-0234)

Stewart, N. A. J., & Lonsdale, A. J. (2016). It's better together: The psychological benefits of singing in a choir. *Psychology of Music*, *44*(6), 1240–1254. https://doi.org/10.1177/0305735615624976

Tarr, M. J. (2016). *TarrLab Stimulus Repository*. [Center for Neural Basis of Cognition Novel Object database]. Retrieved from https://wiki.cnbc.cmu.edu/Novel_Objects

Tremblay, P., & Small, S. L. (2011). On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *NeuroImage*, *57*(4), 1561–1571. https://doi.org/10.1016/j.neuroimage.2011.05.067

Vilkman, E., Sonninen, A., Hurme, P., & Körkkö, P. (1996). External laryngeal frame function in voice production revisited: A review. *Journal of Voice*, *10*(1), 78–92. https://doi.org/10.1016/S0892-1997(96)80021-X

Wan, C. Y., Rüber, T., Hohmann, A., & Schlaug, G. (2010). The therapeutic effects of singing in neurological disorders. *Music Perception: An Interdisciplinary Journal*, *27*(4), 287–295. https://doi.org/10.1525/mp.2010.27.4.287

Warren, J. E., Sauter, D. A., Eisner, F., Wiland, J., Dresner, M. A., Wise, R. J., … Scott, S. K. (2006). Positive emotions preferentially engage an auditory–motor "mirror" system. *Journal of Neuroscience*, *26*(50), 13067–13075. https://doi.org/10.1523/JNEUROSCI.3907-06.2006

Williams, J. D. (1987). Covert linguistic behavior during writing tasks: Psychophysiological differences between above-average and below-average writers. *Written Communication*, *4*(3), 310–328. https://doi.org/10.1177/0741088387004003005

How to cite this article: Pruitt TA, Halpern AR, Pfordresher PQ. Covert singing in anticipatory auditory imagery. *Psychophysiology*. 2019;56:e13297. https://doi.org/10.1111/psyp.13297