

Subvocalization during Preparatory and Non-preparatory Auditory Imagery

Emma B. Greenspon, Anna M. Gentile, Tim A. Pruitt, Andrea R. Halpern & Peter Q. Pfordresher

To cite this article: Emma B. Greenspon, Anna M. Gentile, Tim A. Pruitt, Andrea R. Halpern & Peter Q. Pfordresher (2023) Subvocalization during Preparatory and Non-preparatory Auditory Imagery, *Auditory Perception & Cognition*, 6:1-2, 108-127, DOI: [10.1080/25742442.2022.2163582](https://doi.org/10.1080/25742442.2022.2163582)

To link to this article: <https://doi.org/10.1080/25742442.2022.2163582>

 View supplementary material [↗](#)

 Published online: 01 Jan 2023.

 Submit your article to this journal [↗](#)

 Article views: 58

 View related articles [↗](#)

 View Crossmark data [↗](#)



Subvocalization during Preparatory and Non-preparatory Auditory Imagery

Emma B. Greenspon^a, Anna M. Gentile^b, Tim A. Pruitt^c, Andrea R. Halpern^d
and Peter Q. Pfordresher^c

^aDepartment of Psychology, Monmouth University, West Long Branch, NJ, United States; ^bSchool of Education, University at Albany, Albany, NY, United States; ^cDepartment of Psychology, University at Buffalo, Buffalo, NY, United States; ^dDepartment of Psychology, Bucknell University, Lewisburg, PA, United States

ABSTRACT

Given previous results showing that auditory imagery is associated with subvocal muscle movements related to pitch control, the present study addressed whether subvocalization of pitch is differentially involved during imagery that precedes the execution of an imagined action as compared to non-preparatory imagery. We examined subvocal activity using surface electromyography (sEMG) during auditory imagery that preceded sung reproduction of a pitch sequence (preparatory) or recognition of a pitch sequence (non-preparatory). On different trials, participants either imagined the sequence as presented, or imagined a mental transformation of that sequence. Behavioral results replicated previous findings of poorer reproduction and recognition of transformed sequences compared to sequences in their original form. Physiological results indicated that subvocal activity was significantly above baseline for all conditions, greater than activity observed for the bicep control site, and greater for longer sequences, but did not reliably scale with transformation type. Furthermore, greater subvocal activity during preparatory imagery was associated with greater subvocal activity during non-preparatory imagery for muscles involved in pitch control and articulation. Muscle activity involved in pitch control was similarly recruited for both preparatory and non-preparatory auditory imagery. In contrast, muscle activity involved in vocal articulation was most strongly recruited during motor preparation. Our findings suggest that pitch imagery recruits subvocal muscle activity regardless of whether the imagined action is intended to be effected.

ARTICLE HISTORY

Received 16 February 2022
Revised 28 October 2022
Accepted 22 December 2022

KEYWORDS

Subvocalization; singing; auditory imagery; short-term memory

Although singing is a near-universal behavior, singing accuracy varies from person to person. Inaccurate singing, defined as deviating from a target pitch by more than a semitone on average, occurs in about a third of the adult population (Berkowska & Dalla Bella, 2013; Pfordresher & Demorest, 2021; Hutchins & Peretz, 2012; Pfordresher & Brown, 2007; Pfordresher & Larrouy-Maestri, 2015; Pfordresher et al., 2010). Inaccurate singing can co-exist with accurate pitch discrimination ability, suggesting that this

CONTACT Emma B. Greenspon  egreensp@monmouth.edu  Department of Psychology, Monmouth University, West Long Branch, NJ, United States

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/25742442.2022.2163582>.

© 2023 Informa UK Limited, trading as Taylor & Francis Group

behavior does not necessarily result from a deficit in perceptual pitch processing (Dalla Bella et al., 2007; Pfordresher & Brown, 2007; Pfordresher & Nolan, 2019). Additionally, inaccurate singing is not necessarily tied to impaired vocal motor control (Pfordresher & Brown, 2007; Pfordresher & Mantell, 2009), although inaccurate singers' deficit in pitch matching may be limited to the vocal system (Hutchins et al., 2014; Hutchins & Peretz, 2012). Inaccurate singing has also been tied to a deficit of sensorimotor translation that leads to poor vocal imitation of pitch in both speech and song (Mantell & Pfordresher, 2013), and thus may be considered a general vocal pitch imitation deficit (Pfordresher & Larrouy-Maestri, 2015).

Our recent work has shown that imagining a melody prior to singing results in subvocal activations of muscles involved in pitch control, whereas similar activity is not observed for a visual imagery task (Pruitt et al., 2019). These physiological findings complement behavioral findings in which irrelevant subvocal articulation interferes with auditory imagery of speech (Aleman & Van't Wout, 2004; Smith et al., 1995), suggesting that certain types of motor activations are necessary for the formation of auditory images. Together, these results align with the Multi-Modal Imagery Association (MMIA) model, which suggests that auditory imagery depends on a sensorimotor mechanism involving both auditory and motor processes (Pfordresher et al., 2015). Evidence for sensorimotor processing in auditory imagery as posited by the MMIA model is further supported by neuroimaging research that has found that the act of imagining sound recruits both perceptual and motor planning areas of the brain (Halpern et al., 1999; Herholz et al., 2012; Lima et al., 2016; Zatorre & Halpern, 1993).

Singing is a sensorimotor task because it requires the vocalist to have an accurate representation of the target pitch as well as an accurate motor plan for the vocal system that will result in the vocalization of the target pitch. If both auditory imagery and singing rely on a shared mechanism for sensorimotor processing, then we would expect to observe associations between the ability to form auditory images related to pitch, and singing accuracy. Such associations between imagery and vocal pitch-matching ability have been found using both self-report and behavioral measures of imagery (Greenspon & Pfordresher, 2019; Greenspon et al., 2017, 2020; Pfordresher & Halpern, 2013). In addition, Pruitt et al. (2019) found that laryngeal subvocal motor activations were associated with singing accuracy, albeit in a negative direction. Together, these findings support a sensorimotor account of individual differences in singing ability. However, it is not clear whether subvocal activations are specific to motor preparation prior to singing (Pruitt et al., 2019) or may also be involved more generally in non-preparatory auditory imagery (Brodsky et al., 2008), a question that is addressed by the current study.

Complementing research on auditory imagery and singing ability, research has supported the role of auditory working memory in the vocal reproduction of pitch (Christiner & Reiterer, 2013; Greenspon & Pfordresher, 2019). Recently, Greenspon and Pfordresher (2019) found that a measure of pitch short-term memory, but not digit short-term memory, predicted singing accuracy. In order to reproduce (or recognize) a series of pitches, one must first be able to briefly store those pitches in memory. According to Baddeley's (2012) multicomponent approach to working memory, pitch representations can be maintained in the phonological loop – a short-term store for auditory information (cf., Berz, 1995) – through subvocal motor activations. Therefore, subvocalization may be recruited for both the formation and maintenance of auditory images.

In an attempt to clarify the role of subvocalization during imagery, we adopt the term *preparatory* imagery in the current paper to refer to imagining a specific action prior to engaging in that action. For instance, this type of imagery would be involved when a person rehearses a song in their head before singing the song out loud. In contrast, we use the term *non-preparatory* imagery to refer to forming an auditory image without the intention of actualizing that action; for instance, when a person imagines a melody in their head without the intention of singing the melody out loud.

The present study measured the role of subvocalization during auditory imagery by using surface electromyography (sEMG) during an auditory mental transformation task (Greenspon et al., 2017, 2020; cf., Shepard & Metzler, 1971; Zatorre et al., 2010). The transformation task involves reproducing or recognizing exact repetitions or transformations (transposition = change of key, reversal = retrograde transformation of serial order, serial shift = starting position at second to last event and then looping through the sequence) of three- and four-note novel pitch sequences (Greenspon et al., 2017). Our previous work suggests that auditory images exhibit a degree of inflexibility such that transformations of serial order (shifts and reversals) and key both significantly disrupt performance relative to baseline measures (i.e., exact repetitions of a melody) and the degree of disruption is similar across different types of pitch transformations despite the ubiquity of transpositions in musical performances (Attneave & Olson, 1971).

We addressed whether subvocalization is more strongly recruited during preparatory vs. non-preparatory imagery by comparing subvocal muscle activations across tasks (production vs. recognition). The relationship between subvocalization and the difficulty of manipulating an auditory image was addressed by comparing performance for non-transformed and transformed pitch sequences. In addition, we addressed whether subvocalization scaled with the length of the imagined sequence by comparing performance for three- and four-note sequences.

These manipulations served to test different hypotheses about the role of subvocalization during auditory imagery. If subvocal activations reflect demands related to motor planning, then we would expect greater subvocal activations for preparatory than non-preparatory imagery. Alternatively, if subvocal activations reflect the difficulty of manipulating the auditory image, then we would predict greater subvocal motor activity when participants mentally transform pitch sequences during both preparatory and non-preparatory imagery, as opposed to trials in which participants simply reproduce or recognize the same pattern they just heard. We also considered that if subvocal activations reflect demands related to the length of the imagined sequence, then we should observe greater subvocal activity when participants are asked to rehearse longer compared to shorter sequences. In addition, we left open the possibility that subvocalization could support task demands implemented in more than one way. Finally, we sought to replicate the negative correlation between laryngeal subvocal activation and singing accuracy reported in our previous work (Pruitt et al., 2019).

Method

Participants

Twenty-nine undergraduate students from an introductory psychology course at the University at Buffalo, SUNY participated in this study in exchange for course credit. All

participants completed a baseline assessment procedure to measure singing and perceptual abilities, which was completed on a different day from the experimental procedure in the same academic semester. From this original sample, two participants were removed due to exhibiting a pitch discrimination threshold that was higher than the starting point for the task (see Procedure), which suggests either a pitch perception deficit or a failure to understand instructions. Of the 27 participants who completed the baseline assessment and participated in the full study, one participant was removed because of technical issues of acquiring sEMG data and one participant was removed for not following instructions in the production task.

All behavioral results reported below are based on the remaining 25 participants.¹ The across-sensor analysis of sEMG data used a subset of 19 participants and within-sensor sEMG analyses used subsets that ranged from 22 to 24 participants. Participants were excluded from sEMG analyses due to technical issues of acquiring sEMG data affecting different sensors.² Technical issues could occur when the sensor loses contact with the skin due to a variety of reasons, including participant movement or weak adhesion between the tape on the sensor and the skin. Participants in the final sample (11 female participants, 14 male participants) had a mean age of 19.12 years ($SD = 1.01$ years, range = 18–22 years), an average of 3.4 years of formal music training ($SD = 3.34$ years, range = 0–13 years) and all passed the hearing screening in the range from 1000 Hz to 4000 Hz. Seven participants in the final sample reported six or more years of formal musical training, which, according to the “six-year rule,” would classify them as musicians (Zhang et al., 2020) and six participants in the final sample reported voice as their primary instrument, as such our sample is more representative of an undergraduate population than a population of trained musicians.

Materials

The stimuli used by Greenspon et al. (2017) were used in the current experiment. A set of 16 target pitch sequences were used; half were four notes long and half were three notes long. Target sequences were composed using the C-major scale and began on C or G with equal probability. Sequences were created using the software package Vocaloid: Leon (Zero-G Limited, Okehampton, UK). Notes were sung on the syllable /da/ with an interval of 1 s between notes.

Each stimulus sequence appeared in four different mental transformation conditions: non-transformed, transposed, reversed, and serial shifted. All trials began with the presentation of the pitch sequence in its original form. For preparatory imagery trials in the production task, transformations were indicated via instructions and through the use of an initial cue note that started the reproduction phase (see Procedure). For non-preparatory imagery trials in the recognition task, the initial presentation of the pitch sequence was followed by a silent pause that was 1.5× the duration of the target sequence, and then the presentation of a sequence that either correctly matched the sequence in its transformed state or contained one altered pitch. Altered pitches preserved the contour of the sequence and never occurred on the first or last notes. The majority of altered pitches (86%) fell within 2 semitones of the original pitch, and the largest alteration was up to 5 semitones from the original pitch.

Equipment

Both the assessment and experimental procedures took place in a WhisperRoom SE 2000 sound-attenuated booth (Whisper Room Inc., Morristown, TN) containing a comfortable chair for participants to sit in, as well as a Shure PG58 microphone for vocal recordings. Participants viewed trial instructions on a Dell 15-inch LCD computer monitor positioned in the sound booth which was connected to a computer that the experimenter operated outside of the sound booth. Auditory stimuli were presented to the participants at a comfortable listening volume using a pair of Mackie CR3 series Multimedia Monitor speakers (LOUD Technologies, Woodinville, WA). Matlab (MathWorks Inc., Natick, MA) was used to program the study.

During the experimental trials, participants wore five sEMG sensors as shown in [Figure 1](#). These sensors were applied to the left and right upper lip (*orbicularis oris superioris* muscle), left and right sternohyoid (*m. sternohyoideus*), and the left bicep (*m. biceps brachii*) and were fitted to the participant by a trained research assistant while the participant was seated outside of the sound booth prior to the start of the experiment trials. Sternohyoid sensors were used to measure phonatory subvocalization given previous findings that sternohyoid muscle activity is associated with pitch control (Brodsky et al., 2008; Pruitt et al., 2019; Roubeau et al., 1997). These sensors were placed by first locating the space between thyroid and cricoid cartilages of the neck, and then placing the sensor 1 cm lateral and 1 cm superior to this space (Stepp et al., 2010). Lip sensors were used to monitor articulatory movements. These sensors were positioned just lateral to the philtrum and adjacent to the vermilion border. The bicep sensor was used as a control site. Additionally, a sensor was placed on a loudspeaker in the sound-attenuated booth to record amplitude fluctuations associated with sound onsets in order to synchronize stimuli presentation and vocalizations with sEMG data. sEMG data was captured through Trigno Mini Wireless system (Delsys Trigno Wireless EMG Systems, Boston, MA). These data were converted from analog to digital using the EMG Works Acquisition and Analysis Program (Delsys, Boston MA).

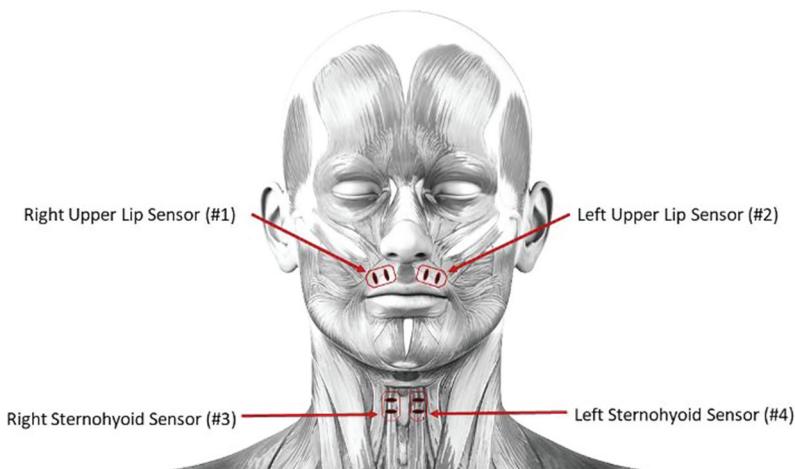


Figure 1. Surface electromyography sensor placement. A fifth sensor was placed on the left bicep as a control site, which is not pictured in the figure.

Procedure

Assessment

For the assessment procedure, participants first completed an audiometry task using a Maico Ma27/MA27e (Maico Diagnostics, Eden Prairie, MN) to assess their hearing ability. Participants were asked to detect 500, 1,000, 2,000, and 4,000 Hz tones played at 20 dB in both their left and right ears. Additionally, participants completed the Seattle Singing Accuracy Protocol (SSAP, Demorest & Pfordresher, 2015; Demorest et al., 2015; Pfordresher & Demorest, 2020), which includes a singing assessment involving the imitative reproduction of single pitches and novel melodies, as well as singing a familiar song from memory. The SSAP also includes an adaptive pitch discrimination task that estimated a participant's smallest detectable change in pure tone frequency. In addition to the SSAP, participants completed an adaptive pitch span task (Williamson & Stewart, 2010), to assess short-term memory for pitch, and a modified pitch imagery arrow task (Gelding et al., 2015; Greenspon & Pfordresher, 2019) to assess pitch imagery accuracy. Finally, participants completed self-report measures: the Bucknell Auditory Imagery Scale (BAIS; Halpern, 2015), the Language Experience and Proficiency Questionnaire (Marian et al., 2007), and the Goldsmith's Musical Sophistication Index (GMSI, Müllensiefen et al., 2014). After completing the assessment procedure, participants were invited to complete the experimental procedure for additional course credit on a separate day within 3 months of the initial assessment.

Experiment

Sensors were first fitted to the participants and then participants were seated in the sound booth prior to the start of the warm-up procedure of the experiment. Warm-up exercises involved singing "Happy Birthday" and reading "The Rainbow Passage" (Fairbanks, 1960) out loud. The ordering of the tasks (i.e., production vs. recognition) followed the procedure used by Greenspon et al. (2017): participants first completed the preparatory imagery trials of the production task followed by the non-preparatory trials of the recognition task. The ordering of tasks provides a conservative test of whether muscle activity is differentially recruited during non-preparatory imagery compared to preparatory imagery based on the reasoning that lower sEMG values during non-preparatory imagery even after participants completed the production task suggest that the muscle activity may be specific to preparation.

Trials within the production and recognition tasks were blocked by transformation condition. Non-transformed trials always came first, followed by blocks of trials for each of the three transformation conditions, with the order of these blocks varying across participants according to a Latin square. Before each block of trials, participants completed practice trials until it was clear to the experimenter that the participant understood the instructions. Each block had 8 trials, resulting in a total of 32 trials.

During preparatory imagery trials (i.e., the production task), participants first heard a recording of the original (non-transformed) pitch sequence in a synthesized vocal timbre (a male voice for male participants, a female voice an octave higher for female participants), using the syllable /da/. Following this, they heard a cue note indicating the starting note to use for production; the pitch class of the cue note varied depending on

the transformation condition. This cue note was followed by a brief imagery period, indicated to the participant by a thought bubble image appearing on the computer monitor. This imagery period was 1.5× as long as the sequence (4 seconds for the three note trials and 6 seconds for the four note trials) in order to give participants ample time to form auditory images and to mentally rehearse the sequence. This thought bubble image was then followed by a green circle, indicating to the participants that they should reproduce the non-transformed or transformed sequence using the syllable /da/. See [Figure 2](#) for an example of what participants were presented during the experimental procedure.

Non-preparatory imagery trials (i.e., the recognition task) were structured to parallel preparatory imagery trials with the exception that the reproduction phase was replaced by the presentation of a pitch sequence that either matched the appropriately transformed or non-transformed version of the original sequence, or differed with respect to one note. Following the presentation of this comparison sequence, participants responded verbally as to whether it was an appropriate match to the current transformation condition (by saying “yes”) or whether it did not (“no”). Participants’ verbal responses were audio recorded and immediately manually entered into the program by the experimenter. Experimenters monitored the participants’ performance through headphones and instructed participants not to overtly vocalize during the imagery phase (mental rehearsal) of either the production or recognition task.

Although sEMG data were collected throughout the entire trial, our focus was on the imagery phase that occurred between the initial presentation of a pitch sequence and the response phase. It is important to note that participants were never told about the meaning or importance of subvocalization.

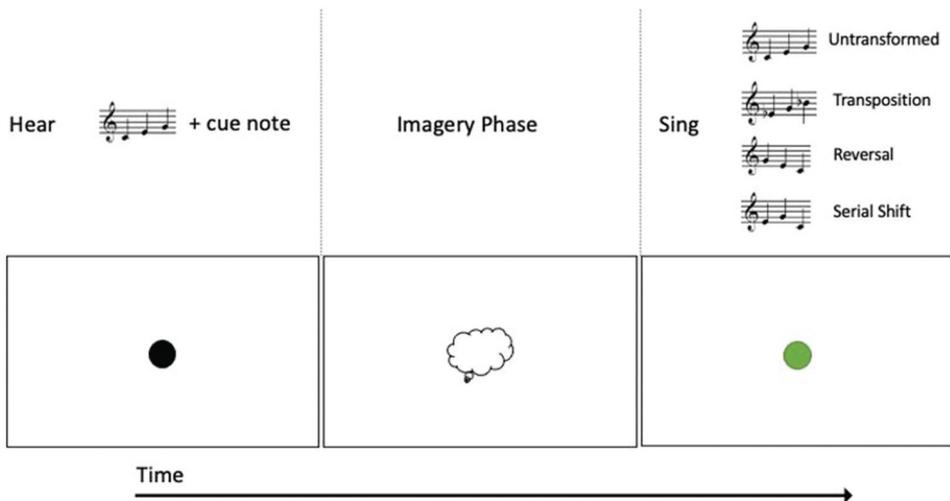


Figure 2. Paradigm for the production task. Participants were instructed to listen to the pitch sequence which played while the black circle was presented on the computer monitor. The thought bubble was used to instruct participants to imagine the sequence and the green circle prompted participants to vocalize the sequence.

Design

Behavioral data for the production (preparatory imagery) and recognition (non-preparatory imagery) tasks were analyzed in separate Analyses of Variance (ANOVAs) given that levels of chance performance differ across these tasks (see “Data Analysis”). Each ANOVA adopted a two-way within-participants design with the factors transformation condition (no transformation, transposition, reversal, serial order shift), and sequence length (three or four notes). Main effects and interactions were further analyzed by pairwise contrasts that incorporated a Holm-Bonferroni correction to preserve familywise alpha of .05. For analyses of significant interactions, pairwise contrasts were performed to compare different transformation conditions within each length condition (length of three- or four-notes) and performed to compare different sequence lengths within each transformation condition.

In order to address whether subvocal motor activity related to pitch control and articulation reflects activity specific to imagery processes rather than general arousal during the imagery phase, we assessed which sensors yielded significant sEMG activity beyond what was found in the control bicep sensor by analyzing mean activation within each sensor and task (production or recognition). Pairwise contrasts across sensors were Holm-Bonferroni corrected. Analyses were also performed within a sensor using a 2 (preparatory and non-preparatory imagery) \times 4 (no transformation, transposition, reversal, and shift) \times 2 (three-note and four-note) within subjects ANOVA. We assessed associations across measures using bivariate correlations.

Data Analysis

We analyzed pitch accuracy for preparatory imagery trials (i.e., the production task) through in-house Matlab scripts (MathWorks Inc., Natick, MA). First, the fundamental frequency (f_0) from recorded digital audio files was extracted through the YIN algorithm (Cheveigné & Kawahara, 2002). Vectors of f_0 values were segmented into notes through a semi-automated procedure in which candidate onsets were identified using peaks in the amplitude contour arising from syllabification (i.e., the initial stop consonant of /da/), followed by correction based on visual and audio inspection. The median f_0 from the central portion of each note was used to estimate the produced pitch for that note. Pitch accuracy was computed by calculating the difference between the produced pitch and the target pitch of the ideal sequence (reflecting the appropriate transformation condition) in cents (100 cents = 1 musical semitone). Deviations with absolute values exceeding 50 cents were categorized as errors, as this degree of error is more proximal to a pitch class other than the target pitch. Sung pitches within 50 cents of the target pitch were categorized as accurate productions.

Reproduction of mentally transformed melodies is surprisingly difficult for participants (Greenspon et al., 2017, 2020). Thus, it is important to establish baseline values for chance performance, reflecting how accurate performance would be if a participant produced pitches within their comfortable range without regard for the correct target note. We computed such estimates using a procedure based on Jacoby et al. (2019; see also Greenspon et al., 2020). We generated 1,000 data sets based on different random permutations of the association between each sung pitch and each target pitch for every

participant. A match between a sung and target pitch that is within a semitone for these data sets thus reflects chance associations, and confidence intervals can be generated based on the probability of different levels of accuracy across all data sets.

sEMG data were analyzed using a procedure based on Pruitt et al. (2019; see also Stepp, 2012). Initial processing of the data proceeded as follows. First, we corrected for DC-offsets in the data through linear detrending. Second, we high-pass filtered the data with 10 Hz cutoff to avoid movement artifacts. The third step involved smoothing the data using a Butterworth band-pass filter with a 20–450 Hz bandwidth to focus on frequencies most reliably associated with muscle contractions. Fourth, we applied an infinite impulse response notch filter centered at 60 Hz to remove potential contaminant signals from electrical power lines. The resulting smoothed and filtered wave was rectified and inspected by lab personnel for artifacts such as head movements, swallowing, or throat-clearing, which lead to highly exaggerated fluctuations in sEMG. In such cases, the corresponding onset and offset of these spurious movements were identified and the intervening data points were removed.

Next, we identified and standardized time points in the trial that may be associated with muscle movements associated with imagery. The imagery phase of each trial in the sEMG data stream was identified using amplitude fluctuations from the sensor placed on the loudspeaker to determine the end point of the stimulus presentation (“listen”) phase, followed by the time designated for mental rehearsal. Within the timeframe associated with the imagery (i.e., mental rehearsal) phase of the trial, we extracted mean sEMG at different time points within a moving window of 200 ms. The highest mean value within a trial was used as a measure of peak activity. We standardized these values to control for individual differences and morphological differences in muscle size by subtracting the mean value associated with the baseline activity phase at the start of each trial. Although Pruitt et al. (2019) used rest trials, in the current study the baseline activity phase was determined by passing a 1 s moving window across the listening phase of the trial and selecting the sample that exhibited average minimum sEMG activity. Within-trial baselines have been used in prior research measuring sEMG from facial muscles (Livingstone et al., 2016) and were used in the current study to maximize consistency in sEMG readings across the experiment. Following the removal of trials that contained spurious data due to experimenter error (<2% of all trials), standardized values for a trial that were greater than $\pm 3 SD$ from the mean were removed from further analysis.³ The resulting standardized peak activations for each trial constitute the data we report in analyses of sEMG.

Results

Behavioral Data

We first addressed whether effects of length (i.e., higher accuracy for three-note than four-note sequences) and transformation (i.e., higher accuracy for non-transformed compared to transformed sequences) were replicated as per Greenspon et al. (2017, 2020) by running a 2 (three- and four-note) \times 4 (no transformation, transposition, serial order shift, and reversal) ANOVA on singing accuracy for preparatory imagery trials (i.e., the production task).⁴ As shown in Figure 3, the ANOVA yielded significant main effects of length, $F(1, 24) = 7.79$, $p = .01$, $\eta^2_p = .25$ and transformation condition, $F(3,$

72) = 24.60, $p < 0.001$, $\eta^2_p = .51$, but no interaction ($p = .13$). The main effect of length was driven by higher accuracy when participants reproduced three-note sequences ($M = .27$, $SD = .19$) compared to four-note sequences ($M = .23$, $SD = .16$). The main effect of transformation, which we interpreted as reflecting the difficulty of manipulating an auditory image, was analyzed using pairwise contrasts with a Holm-Bonferroni correction. Accuracy was higher for reproduction of non-transformed sequences than for all transformation conditions ($p < .001$ for each contrast), whereas none of the transformed conditions differed from each other ($p > .40$ for each contrast). Additionally, accuracy was consistently at or near chance performance for the three types of transformations, where chance is represented by the shaded area in Figure 3. These results replicate results from Greenspon et al. (2017, 2020). Figure 3 displays the proportion of pitches sung correctly in each condition for preparatory imagery trials.

Next, we analyzed the proportion of correctly recognized sequences for the non-preparatory imagery trials as shown in Figure 4. The ANOVA yielded significant main effects of transformation condition, $F(3, 72) = 14.71$, $p < 0.001$, $\eta^2_p = .38$, a significant transformation \times length interaction, $F(3, 72) = 5.78$, $p = .001$, $\eta^2_p = .19$, but no main effect of length ($p = .129$). Pairwise contrasts for the main effect of transformation yielded the same pattern as found for the production data. The interaction arose from the fact that performance was significantly lower for recognition of non-transformed sequences of four-note sequences than for three-note sequences ($p < .001$). The diminished accuracy for recognition of four-note sequences eliminated differences across non-transformed and transformation conditions. For three-note sequences, pairwise contrasts comparing the non-transformed condition to the transformation conditions yielded the same

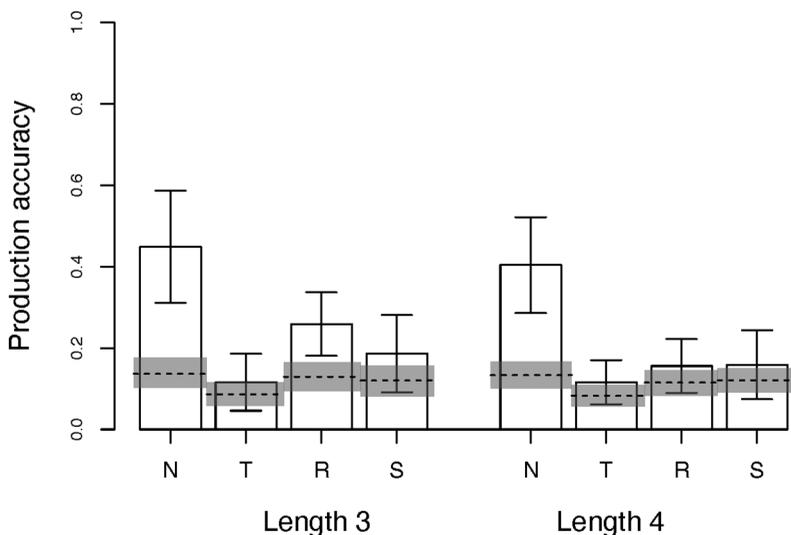


Figure 3. Proportion of correctly produced pitches by sequence length and transformation condition. Labels on X-axis indicate transformation condition where N= no transformation, T = transposition of key, R = reversal of order, S = shift of the starting position. Bars illustrate means surrounded by 95% confidence intervals. The shaded area represents chance performance with the mean denoted by the dashed line and 95% confidence intervals from a null distribution of sample means (see Data Analysis).

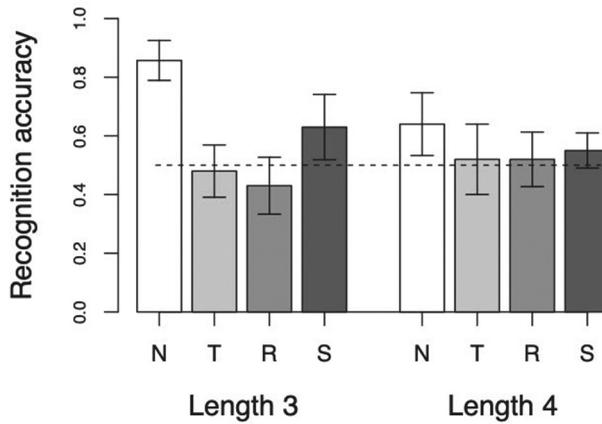


Figure 4. Proportion of correctly recognized sequence by sequence length and transformation condition. Labels on X-axis indicate transformation condition where N = no transformation, T = transposition of key, R = reversal of order, S = shift of the starting position. Bars illustrate means surrounded by 95% confidence intervals. The dashed line represents chance performance (.50 proportion correct).

pattern for recognition as found for production data (i.e., higher accuracy for non-transformed compared to transformed conditions). With respect to comparisons across the three transformation conditions, transformations were recognized more accurately for serial shifts than reversal. No other comparisons across the transformation conditions were significant. In line with results from the production task, transformation performance was again consistently at or near chance performance even when participants were asked to recognize transposed pitch sequences. In sum, behavioral performance for both preparatory and non-preparatory imagery trials supports the robust and reliably disruptive effect of mental transformations of pitch (Greenspon et al., 2017, 2020).

We then assessed whether individual differences in performance were correlated across preparatory imagery (production) and non-preparatory imagery (recognition) trials. As shown in Figure 5, overall accuracy for preparatory auditory imagery was positively related to overall accuracy for non-preparatory auditory imagery (i.e., scores averaged across the four conditions in each task). This relationship held even when limiting the correlations to non-transformed, $r(23) = .55, p < .01$, and transformed trials, $r(23) = .49, p < .01$. In addition, we found that BAIS subscales, pitch imagery ability, pitch span, pitch discrimination, and musical training all correlated with accuracy in the production task, replicating the associations reported in Greenspon and Pfordresher (2019). Unsurprisingly, self-reported singing behavior, which was measured from the GMSI, was also significantly related to production accuracy in the preparatory imagery trials. A full list of correlations between behavioral measures can be found in the Supplemental Material.

sEMG Data

We first assessed whether sEMG activity reflected subvocal motor movements during the auditory imagery phase of the experiment. We did this by contrasting standardized mean peak values across sensors and task modalities, averaging across transformation

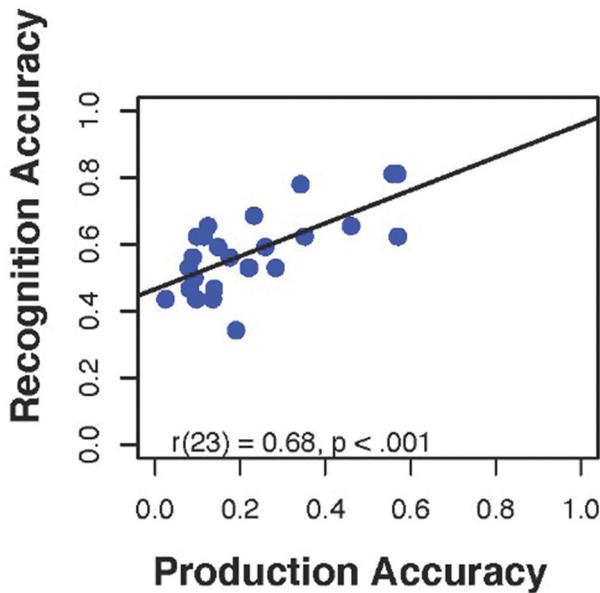


Figure 5. Correlation between accuracy in production and recognition tasks. Each plotting element represents mean performance for a participant across all trials and conditions. The line reflects the best-fitting least-squares regression.

conditions and sequence lengths. For this analysis, the bicep sensor serves as a critical comparison given that bicep contractions are not assumed to play any role in subvocalization.

Mean sEMG activity for each sensor in the production and recognition tasks is shown in [Figure 6](#). The 5 (Bicep, Left and Right Sternohyoid, Left and Right Lip) \times 2 (preparatory and non-preparatory imagery) ANOVA yielded significant main effects of sensor, $F(4, 72) = 14.63, p < 0.001, \eta^2_p = .45$, a significant main effect of task, $F(1, 18) = 14.33, p < 0.001, \eta^2_p = .44$, and a significant interaction, $F(4, 72) = 5.20, p = 0.001, \eta^2_p = .22$. The main effect of sensor arose from the fact that all face and neck sensors yielded higher activation than the bicep sensor ($p < .01$ for all contrasts) and the left sternohyoid muscle yielded higher activation than the left upper lip ($p = .015$). No other contrasts across sensors were statistically significant. The main effect of task reflected greater activation during preparatory imagery prior to production ($M = 3.64, SD = 1.20$) than non-preparatory imagery prior to recognition ($M = 2.69, SD = 1.26$). The interaction reflected the fact that differences across production (preparatory imagery) and recognition (non-preparatory imagery) were only significant for the two lip sensors (left lip $p = .018$, right lip $p = .002$), and not for laryngeal sensors ($p > .50$ for each).

To address the effects of task, length of imagined sequence, and difficulty of image manipulation on sEMG activity we ran separate 2 (preparatory and non-preparatory imagery) \times 2 (three-note and four-note) \times 4 (no transformation, transposition, reversal, and shift) ANOVAs on standardized sEMG values for each sensor site: left sternohyoid muscle, right sternohyoid muscle, left lip muscle, right lip muscle, and bicep muscle. The results of each analysis are shown in [Table 1](#).

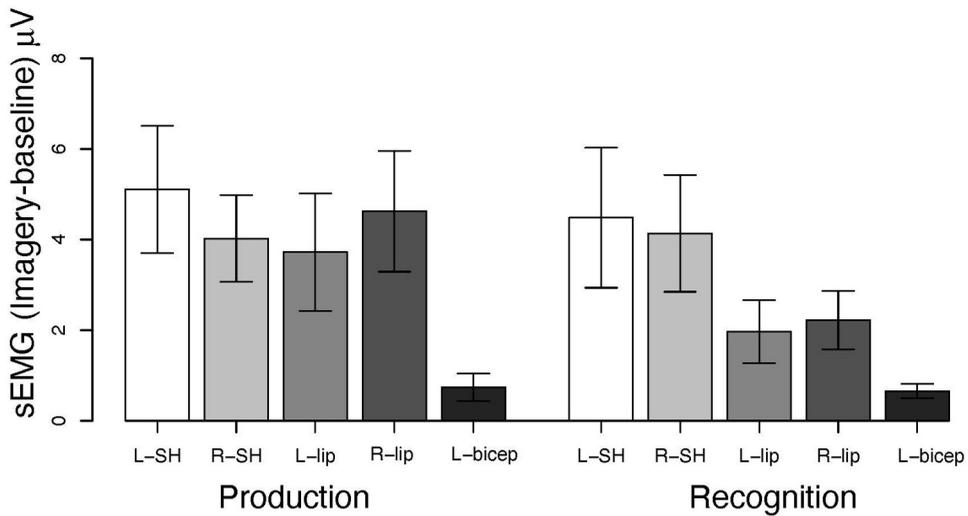


Figure 6. Mean peak sEMG activity across sensors and tasks. Labels on X-axis indicate sensor, where L = left lateralized site, R = right lateralized site, and SH = sternohyoid muscle. Bars illustrate means surrounded by 95% confidence intervals. All values are standardized by differencing rest activation within trial; units are expressed in microvolts.

Table 1. Summary of ANOVA effects for the sEMG sites.

Sensor					
Effect	L SH (<i>n</i> = 24) <i>F</i>	R SH (<i>n</i> = 22) <i>F</i>	L Lip (<i>n</i> = 23) <i>F</i>	R Lip (<i>n</i> = 24) <i>F</i>	L Bicep (<i>n</i> = 24) <i>F</i>
Task	0.53	0.23	12.30**	14.32**	0.00
Length	4.30*	4.28	4.89*	5.99*	2.56
Transformation	1.26	3.15*	0.08	0.21	0.78
Task × Length	0.76	0.49	1.56	1.92	1.75
Task × Transformation	0.02	0.63	0.33	1.28	0.59
Length × Transformation	0.64	1.20	0.41	0.58	0.34
Task × Length × Transformation	1.11	0.96	0.27	1.06	0.58

Note: * $p < .05$, ** $p < .01$. L = left lateralized site, R = right lateralized site, and SH = sternohyoid muscle.

As mentioned earlier, only the upper lip sensors revealed a main effect of task, as shown in Figure 6. Like the effect of task, the effect of length also varied by sensor: longer imagined sequences were associated with greater standardized sEMG activity, and this effect was observed for the lip (left and right) and sternohyoid (left only) sites but not the bicep control site. Finally, only the right sternohyoid muscle revealed a main effect of transformation. Unlike the robust and reliable effect of transformation found for behavioral data, Holm-Bonferroni corrected comparisons revealed that this effect was restricted to one pair of conditions; right sternohyoid activity during the shift condition ($M = 4.97$, $SD = 2.68$) was significantly greater than activity observed during the non-transformed condition ($M = 3.65$, $SD = 2.17$) and this was the only significant difference in standardized sEMG values across the four transformation conditions.

Next, as we did for behavioral data, we assessed whether individual differences in sEMG values for each sensor were correlated across preparatory imagery (production)

Table 2. Correlations across production and recognition sEMG values within each sensor.

Sensor	<i>r</i>	<i>df</i>
Left Sternohyoid	.70***	22
Right Sternohyoid	.57**	20
Left Lip	.51*	21
Right Lip	.46*	22
Bicep	.17	22

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

and non-preparatory imagery (recognition) trials, as shown in Table 2. We found consistent positive associations across muscle activation during preparatory and non-preparatory imagery for the facial sEMG sites. Importantly, the only relationship that did not reach statistical significance was observed for the bicep control site.

In addition to the above within-measure analyses, we were also interested in whether physiological and behavioral measures were related within a task. These analyses focus on the sternohyoid muscle given its theorized role in subvocalization and vocal pitch control (Brodsky et al., 2008; Roubeau et al., 1997). There was a significant negative correlation between sternohyoid sEMG and production accuracy, $r(20) = -.40$, $p < .05$ for the right sternohyoid site. Although the negative correlation between the left sternohyoid sEMG and behavioral accuracy for production was not significant, $r(22) = -.13$, $p = .27$, this association, like the right sternohyoid site, is in the same direction as the association reported by Pruitt et al. (2019). Together, our results suggest that participants who exhibit greater subvocalization tend to exhibit lower singing accuracy than participants who exhibit less subvocalization. Associations between sEMG and recognition accuracy did not reach significance for either sternohyoid site (all $p > .53$). As found for the behavioral data, physiological data were significantly correlated with self-reported imagery as measured by the BAIS and self-reported singing behavior as measured by the GMSI; however, correlations were less consistent for physiological measures than for behavioral measures. BAIS scores and sEMG were positively correlated, and complementing the association reported between overall singing accuracy and sEMG activity, we found that self-reported singing behavior was negatively associated with sternohyoid activity. A full list of correlations across assessment measures and measures from the experiment are presented in the supplemental material.

Discussion

The current study assessed behavioral and sEMG measures within an auditory mental transformation paradigm to determine whether performance accuracy and muscle activity during auditory imagery scaled with task (production vs. recognition), the length of the imagined sequence of pitches, and the difficulty of manipulating the imagined sequence. We report the following three novel sEMG findings. First, we found evidence for subvocalization in preparatory and non-preparatory contexts based on significant muscle activity at all sEMG facial and neck sites (involving articulators and pitch control muscles), in comparison to the bicep sensor as a control site. Importantly, the sternohyoid muscle, a muscle with a suggested role in vocal pitch control (Roubeau et al., 1997), was similarly recruited for preparatory and non-preparatory imagery. In contrast, the

upper lip, a muscle involved in vocal articulation, but not pitch control, exhibited greater activity during preparatory as opposed to non-preparatory tasks. This difference suggests that lip muscles, more than sternohyoid muscles, may reflect subvocal movements used to prepare motor actions and may not exclusively reflect imagery processes. Second, we replicated the negative correlation between subvocalization and singing accuracy reported in Pruitt et al. (2019). Third, we replicated behavioral results of Greenspon et al. (2017) whereby performance accuracy in the current study was related to the difficulty of manipulating an auditory image during preparatory and non-preparatory imagery. Notably, we found a dissociation in that manipulations of difficulty did not have a corresponding effect on physiological measures.

Given this pattern of results, we suggest that the role of subvocal muscle movements of the sternohyoid muscles may reflect motor processing pertaining to a sensorimotor image of a pitch sequence, which supports the role of sensorimotor processing during auditory imagery described in the MMIA model (Pfordresher et al., 2015). Furthermore, the current study highlights the role of motor processes in the peripheral nervous system during auditory imagery and complements neuroimaging research that has reported that imagining pitch recruits cortical motor planning processes at the level of the central nervous system (Halpern et al., 1999; Herholz et al., 2012; Lima et al., 2016; Zatorre & Halpern, 1993). The upper lip muscles may be more strongly recruited for vocal articulation than pitch control, a claim that aligns with previous work on auditory imagery for verbal information (Aleman & Van't Wout, 2004; Smith et al., 1995). By contrast, the role of lip movements in imagery for the present pitch sequences would be minimal, given that every note was articulated using the same syllable.

Given prior support for the sternohyoid's involvement in auditory pitch imagery processes (Pruitt et al., 2019; c.f. Brodsky et al., 2008), the current study also addressed whether sternohyoid activity during subvocalization is differentially recruited for imagined pitch sequences that vary in length or transformation type. Interestingly, we found a main effect of length for the lip and sternohyoid (left only) muscles that was not observed in the bicep control. This effect was driven by greater subvocalization during the imagery phase for longer compared to shorter sequences. These sEMG results align with Baddeley's multicomponent approach to working memory, particularly with respect to the role of subvocalization in maintaining auditory information within the phonological loop (Baddeley, 2012). However, an alternative explanation for this pattern of findings is that a longer sequence may provide a greater chance to obtain unusually high maximal values as a result of a greater number of observations across the imagery phase. This confound could be addressed in future studies by analyzing an imagery phase that has a consistent duration across the three-note and four-note sequences. However, the absence of the effect of length in the bicep control sensor provides evidence against this alternative explanation. A potential limitation of the current study is that the baseline measure for sEMG was defined during the listening phase rather than during a silent phase within a trial, which may account for the positive standardized values observed in the control site (i.e., bicep). It is important to note that the baseline measure was not confounded with the type of transformation performed in a trial and therefore we would still expect to detect a transformation effect in our sEMG data if subvocalization did vary by transformation condition. However, in contrast to the robust behavioral effects found for the transformation conditions, surprisingly, the effect of transformation was not

borne out in sEMG. We observed a main effect of transformation type for the right sternohyoid muscle that was limited to the comparison between the non-transformed and serial shift conditions and thus did not reflect the consistent differences across conditions found in behavior reported here and elsewhere (Greenspon et al., 2017, 2020).

Based on our psychophysiological data, we suggest that subvocalization, as measured by activity of the sternohyoid muscle, may reflect the initiation of a motor plan that is associated with a sensorimotor auditory image (cf., Pfordresher et al., 2015). Whereas muscle contractions associated with the initiation of this “motor image” may reflect subvocal processes involved in the maintenance of a pitch sequence (Baddeley, 2012), this activity does not scale with the difficulty of transforming an imagined sequence and may instead be associated with upstream cortical activity that precedes the peripheral activity we observe here.

We also observed that individuals who exhibited greater sEMG activity for a specific muscle during the imagery phase of production trials tended to exhibit greater sEMG activity for that muscle during the imagery phase of recognition trials. Importantly, this pattern was not found for the bicep control site and was observed only for sEMG recording sites involved in the vocal system (i.e., upper lip and sternohyoid muscles). This pattern of results suggests that the upper lip and sternohyoid muscles are more reliably recruited during preparatory and non-preparatory imagery than the control site and that individual differences in muscle activity during imagery are related across tasks. Furthermore, we also replicated the negative correlation between physiological and behavioral measures reported in Pruitt et al. (2019). Individuals who exhibited greater subvocal activity tended to have poorer singing accuracy compared to individuals who subvocalized less. In the current study, this association was found for the right, but not left (as in Pruitt et al., 2019) sternohyoid muscle. It is worth noting that this association was observed despite using a different task and smaller sample than Pruitt et al. (2019, $N = 46$).

In addition to the novel sEMG findings discussed above, we also report a replication of our prior behavioral data (Greenspon et al., 2017, 2020), by demonstrating that accuracy on both production and recognition tasks was significantly, and similarly, disrupted by the three types of melody transformation as compared to performance for non-transformed melodies. This pattern of results suggests that auditory images reflect a degree of inflexibility, even for melody transpositions, which are a common type of pitch transformation used in music performance. Relatedly, Greenspon et al. (2020) showed that difficulty in manipulating auditory images extends even to auditory images of familiar melodies, where retrieval should involve relatively few cognitive resources given how well those melodies are encoded. With respect to short-term memory demands, we found a disrupting effect of sequence length in both behavioral tasks, with poorer performance associated with longer sequences. The main effect of sequence length in the current study replicated the main effect of length reported in our previous research (Greenspon et al., 2020) and aligns with results from Greenspon and Pfordresher (2019) who found evidence that pitch short-term memory capacity is associated with singing ability. Finally, complementing the within-participant consistencies of sEMG activity for the upper lip and sternohyoid muscles, we found that performance was associated across the production and recognition tasks, regardless of whether participants were imagining a pitch sequence in its original or transformed form. We interpret these psychophysiological and behavioral findings as

suggesting that individual differences in sensorimotor processing of pitch are related across preparatory and non-preparatory imagery.

To summarize, the current study used behavioral and physiological measures to determine whether performance and muscle activity scales with the length of an imagined pitch sequence and the difficulty of manipulating an imagined pitch sequence during preparatory (e.g., Pruitt et al., 2019) and non-preparatory (e.g., Brodsky et al., 2008) auditory imagery. We reported several novel findings pertaining to sEMG measures, including the finding that muscle activations related to pitch control are present during both preparatory and non-preparatory auditory imagery, and that this activity may reflect the initiation of a general motor program of a sensorimotor image of pitch. In addition, whereas subvocal muscle activations may reflect cognitive demands related to memory load, these activations may not be related to varied cognitive demands pertaining to task difficulty. Although we also found that articulatory processes were recruited during both preparatory and non-preparatory imagery, these processes appear to be more specifically related to motor preparation than processing imagined pitch sequences. In addition to replicating behavioral effects of the pitch transformation paradigm (Greenspon et al., 2017, 2020), we also replicated the relationship between the degree of subvocalization and behavioral performance as reported in Pruitt et al. (2019). To further address this relationship, future research involving trained and untrained singers, such as a paradigm that randomly assigns untrained singers to a course of training in pitch-matching, may help clarify the degree to which subvocal activity is related to singing accuracy.

Notes

1. This sample size exceeds the size needed to achieve 80% power for detecting the transformation effect in our behavioral data based on a power analysis performed using G*power 3.1 (Faul et al., 2007) using the effect size for the transformation effect reported in Greenspon et al. (2017).
2. Six participants were excluded from the five-sensor ANOVA, three participants were excluded from analyses using the right sternohyoid sensor, two participants were excluded from analyses using the upper left lip sensor, and all remaining within-sensor analyses excluded one participant.
3. This step resulted in the exclusion of between 2.5% and 6.5% of trials in each task (i.e., production and recognition) for each of the five sensors (Bicep, Left and Right Sternohyoid, Left and Right Lip) due to extreme sEMG values ($>3 SD$ from the mean).
4. A 2 (three- and four-note) \times 4 (no transformation, transposition, serial order shift, and reversal) ANOVA on relative pitch accuracy replicated the transformation effect found for the pitch accuracy measure. No other effects were significant for analyses on relative pitch accuracy, including the main effect of length.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the NSF BCS1848930.

Ethics Statement

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

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>
- Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *The American Journal of Psychology*, 84(2), 147–166. <https://doi.org/10.2307/1421351>
- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63(1), 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- 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>
- Berz, W. L. (1995). Working memory in music: A theoretical model. *Music Perception*, 12(3), 353–364. <https://doi.org/10.2307/40286188>
- 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(2), 427–445. <https://doi.org/10.1037/0096-1523.34.2.427>
- Cheveigné, A. D., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4), 1917–1930. <https://doi.org/10.1121/1.1458024>
- Christiner, M., & Reiterer, S. M. (2013). Song and speech: Examining the link between singing talent and speech imitation ability. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00874>
- Dalla Bella, S., Giguère, J.-F., Peretz, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America*, 121(2), 1182–1189. <https://doi.org/10.1121/1.2427111>
- Demorest, S. M., & Pfordresher, P. Q. (2015). Singing accuracy development from K-adult. *Music Perception*, 32(3), 293–302. <https://doi.org/10.1525/mp.2015.32.3.293>
- Demorest, S. M., Pfordresher, P. Q., Dalla Bella, S., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015). Methodological perspectives on singing accuracy: In introduction to the special issue on singing accuracy (Part 2). *Music Perception*, 32(3), 266–271. <https://doi.org/10.1525/mp.2015.32.3.266>
- Fairbanks, G. (1960). *Voice and articulation drillbook* 2nd ed. Harper.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Gelding, R. W., Thompson, W. F., & Johnson, B. W. (2015). The pitch imagery arrow task: Effects of musical training, vividness, and mental control. *PLoS ONE*, 10(3), e0121809. <https://doi.org/10.1371/journal.pone.0121809>
- Greenspon, E. B., & Pfordresher, P. Q. (2019). Pitch-specific contributions of auditory imagery and auditory memory in vocal pitch imitation. *Attention, Perception, & Psychophysics*, 81(7), 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(5), 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(1–2), 76–93. <https://doi.org/10.1080/25742442.2020.1866428>

- 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>
- Halpern, A. R., Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, 9(7), 697–704. <https://doi.org/10.1093/cercor/9.7.697>
- 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
- Hutchins, S., Larrouy-Maestri, P., & Peretz, I. (2014). Singing ability is rooted in vocal-motor control of pitch. *Attention, Perception & Psychophysics*, 76(8), 2522–2530. <https://doi.org/10.3758/s13414-014-0732-1>
- 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>
- Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology*, 29(19), 3229–3243. <https://doi.org/10.1016/j.cub.2019.08.020>
- Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends in Neurosciences*, 39(8), 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>
- Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition*, 127(2), 177–202. <https://doi.org/10.1016/j.cognition.2012.12.008>
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4), 940–967. [https://doi.org/10.1044/1092-4388\(2007\)067](https://doi.org/10.1044/1092-4388(2007)067)
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLOS ONE* 9(2): e89642. <https://doi.org/10.1371/journal.pone.0089642>
- 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., & 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: Vol. 1 Development* (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., Demorest, S. M., Bella, S. D., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015). Theoretical perspectives on singing accuracy. *Music Perception*, 32(3), 227–231. <https://doi.org/10.1525/mp.2015.32.3.227>
- 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., & 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. Q., & Mantell, J. (2009). Singing as a form of vocal imitation—Mechanisms and deficits. In *ESCOM 2009: 7th Triennial Conference of European Society for the Cognitive Sciences of Music*.

- Pfordresher, P. Q., & Nolan, N. P. (2019). Testing convergence between singing and music perception accuracy using two standardized measures. *Auditory Perception & Cognition*, 2(1–2), 67–81. <https://doi.org/10.1080/25742442.2019.1663716>
- Pruitt, T. A., Halpern, A. R., & Pfordresher, P. Q. (2019). Covert singing in anticipatory auditory imagery. *Psychophysiology*, 56(3), e13297. <https://doi.org/10.1111/psyp.13297>
- Roubeau, B., Chevrie-Muller, C., & L, J. (1997). Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta Otolaryngologica*, 117(3), 459–464. <https://doi.org/10.3109/00016489709113421>
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703. <https://doi.org/10.1126/science.171.3972.701>
- 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](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. [https://doi.org/10.1044/1092-4388\(2011/11-0214\)](https://doi.org/10.1044/1092-4388(2011/11-0214))
- 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\(2010/09-0234\)](https://doi.org/10.1044/1092-4388(2010/09-0234))
- Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory*, 18(6), 657–669. <https://doi.org/10.1080/09658211.2010.501339>
- Zatorre, R. J., & Halpern, A. R. (1993). Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia*, 31(3), 221–232. [https://doi.org/10.1016/0028-3932\(93\)90086-f](https://doi.org/10.1016/0028-3932(93)90086-f)
- Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: A role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, 22(4), 775–789. <https://doi.org/10.1162/jocn.2009.21239>
- Zhang, J. D., Susino, M., McPherson, G. E., & Schubert, E. (2020). The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music*, 48(3), 389–409. <https://doi.org/10.1177/0305735618804038>