Pitch-specific contributions of auditory imagery and auditory memory in vocal pitch imitation

Emma B. Greenspon^{1,2} · Peter Q. Pfordresher¹

Published online: 8 July 2019
© The Psychonomic Society, Inc. 2019

Abstract

Vocal imitation guides both music and language development. Despite the developmental significance of this behavior, a sizable minority of individuals are inaccurate at vocal pitch imitation. Although previous research suggested that inaccurate pitch imitation results from deficient sensorimotor associations between pitch perception and vocal motor planning, the cognitive processes involved in sensorimotor translation are not clearly defined. In the present research, we investigated the roles of basic cognitive processes in the vocal imitation of pitch, as well as the degree to which these processes rely on pitch-specific resources. In the present study, participants completed a battery of pitch and verbal tasks to measure pitch perception, pitch and verbal auditory imagery, pitch and verbal auditory short-term memory, and pitch imitation ability. Information on participants' music background was collected, as well. Pitch imagery, pitch short-term memory, pitch discrimination ability, and musical experience were unique predictors of pitch imitation ability. Furthermore, pitch imagery was a partial mediator of the relationship between pitch short-term memory and pitch imitation ability. These results indicate that vocal imitation recruits cognitive processes that rely on at least partially separate neural resources for pitch and verbal representations.

Keywords Vocal pitch imitation · Auditory imagery · Short-term memory

The ability to imitate sound using one's voice is an early-emerging behavior in humans and is a critical behavior in language acquisition (for a review, see Kuhl, 2010). Beyond language development, the ability to vocally imitate sound, specifically pitch, is an important component in music production (i.e., singing). Singing is a ubiquitous activity that serves social, cultural, and educational functions. Singing emerges spontaneously in early childhood, as is the case with speech. However, unlike speech, singing abilities vary later in life, specifically with respect to vocal pitch imitation. Furthermore, a sizeable minority of the population never acquires the ability to match a sung pitch within a semitone of the target pitch (Dalla Bella, Giguère, & Peretz, 2007; Pfordresher & Brown, 2007; Pfordresher, Brown, Meier,

Belyk, & Liotti, 2010; Pfordresher & Larrouy-Maestri, 2015). The research presented here addresses what basic cognitive processes are related to individual differences in vocal pitch imitation ability.

The vocal imitation of pitch recruits a broad network of processes related to perception, motor control, and cognition (Pfordresher, Halpern, & Greenspon, 2015). The core of this network involves associations between auditory perception and vocal motor control within the *vocal sensorimotor loop* (Berkowska & Dalla Bella, 2009), a simplified version of which is shown in Fig. 1. The research to date suggests that individual differences in vocal pitch imitation may primarily reflect variability in *sensorimotor translation*. Support for this claim has come from the following lines of research. First, inaccurate singers often exhibit precise pitch discrimination skills, indicating that inaccurate vocal imitation is not a direct consequence of impaired perception (Dalla Bella et al., 2007; Pfordresher & Brown, 2007). Second, a purely motor-control account of inaccurate pitch imitation is not sufficient, either, given evidence that inaccurate singers exhibit vocal ranges similar to those of accurate singers during spontaneous vocal production and show precision similar to that of accurate singers

✉ Emma B. Greenspon
egreensp@monmouth.edu; ebgreens@buffalo.edu

¹ Department of Psychology, University at Buffalo, State University of New York, New York, NY, USA

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

Fig. 1 A simplified version of the vocal sensorimotor loop

when sustaining a constant pitch (Pfordresher & Mantell, 2009). Finally, inaccurate singers are able to match pitch outside the vocal motor system (Hutchins & Peretz, 2012). These lines of research have motivated a sensorimotor account of inaccurate pitch imitation, such that inaccurate singers are ineffective at mapping auditory representations onto motor representations of the vocal tract, resulting in inaccurate vocal performance (Hutchins & Peretz, 2012; Pfordresher & Brown, 2007; Pfordresher et al., 2015). Although sensorimotor translation may be the source of individual differences in pitch imitation ability, it is not clear from these observations what processes are necessary for sensorimotor translation to occur.

One process that might facilitate sensorimotor translation is the formation of an accurate auditory image. Various lines of evidence suggest that auditory imagery integrates both perceptual and motor representations, and thus might provide a basis for sensorimotor translation. Importantly, auditory imagery recruits brain activity in both secondary auditory areas and motor-planning areas (Leaver, Van Lare, Zielinski, Halpern, & Rauschecker, 2009; Lima et al., 2015; McNorgan, 2012; Zatorre & Halpern, 2005). Motor components of auditory images have also been established using vocal–motor interference tasks, such as subarticulation, which can interfere with auditory imagery (Smith, Wilson, & Reisberg, 1995). Finally, the assumption of a link between imagery and performance has a long-standing history in music education, in which *notational audiation*, or the formation of an auditory image from notation, is assumed to be critical for sight-reading (Gordon, 1975). Consistent with these observations, inaccurate singers report experiencing less vivid auditory images than do accurate singers (Greenspon, Pfordresher, & Halpern, 2017; Pfordresher & Halpern, 2013). However, no studies to date have related vocal pitch imitation to the accuracy with which auditory images can be formed online.

Although some pitch imitation tasks involve the reproduction of single pitches, which have a low short-term memory load, most singing involves listening to and then reproducing

sequences of pitches in a melody: a task involving high short-term memory demands. Therefore, individual differences in pitch imitation accuracy might be due in part to differences in short-term memory capacity. However, research to date has been mixed regarding the relationship between auditory short-term memory capacity and vocal pitch imitation ability. Pfordresher and Brown (2007) found that inaccurate singers do not exhibit a greater deficit for sequences with more complex pitch patterns. In fact, the largest deficits among this group are found for imitating single pitches (Pfordresher & Brown, 2007). On the other hand, a more recent study did uncover an association between backward digit span and vocal pitch imitation, but only when participants were required to produce mental transformations of pitch sequences (Greenspon et al., 2017). In support of the role of short-term memory capacity in vocal performance, Christiner and Reiterer (2013) reported a positive correlation of both forward and backward digit span with a measure of singing ability, but it is not clear how closely associated the measure of singing was to accuracy in pitch imitation. In sum, although auditory short-term memory is a theoretically critical process involved in pitch imitation, the relationship between short-term capacity and pitch imitation ability is currently not well-defined. More important, to date, no studies have focused specifically on the relationship between pitch imitation and memory span for pitch information (Williamson & Stewart, 2010). As such, it remains to be determined whether these variable results reflect measurements of a memory store that is not critical for singing or differences in the operationalizing of short-term memory and singing ability across the studies.

Thus, in the present study we analyzed the correlations between accuracy of vocal pitch imitation and online measures of both auditory imagery and short-term memory, as is shown in Fig. 2. Although it is difficult to separate imagery and short-term memory, given their likely integration (Baddeley & Andrade, 2000), we separated these processes as clearly as possible on the basis of task demands. As such, our imagery tasks were designed to place minimal demands on memory

Fig. 2 Critical measures and the associated hypotheses

capacity, while requiring participants to simulate a perceptual experience based on imagery. In contrast, our memory tasks involved the physical presentation of stimuli—thus placing minimal demands with respect to imagery—but could include many items to store, thus placing demands on capacity. In this way, in the present study we evaluated the degree to which auditory imagery and auditory short-term memory are involved in sensorimotor translation between auditory pitch perception and vocal–motor control of pitch.

In addition to inconsistencies in how auditory imagery and auditory short-term memory have been operationalized in the past singing literature, previous studies have also been inconsistent with respect to the dimensions (e.g., pitch or phonetic) that are reflected in measures of auditory imagery, auditory short-term memory, and vocal imitation ability (Christiner & Reiterer, 2013; Greenspon et al., 2017; Pfordresher & Halpern, 2013). For instance, studies investigating the role of auditory imagery in pitch imitation have used self-report measures of imagery that include verbal items (Greenspon et al., 2017; Pfordresher & Halpern, 2013), and studies investigating the role of short-term memory in pitch imitation have used verbal span tasks to measure short-term memory capacity (Christiner & Reiterer, 2013; Greenspon et al., 2017). However, the degree to which pitch imitation relies on pitch-specific processes with regard to imagery and memory has not been directly addressed—a primary aim of the present study, as we outline in Fig. 2. If pitch imitation does involve pitch-specific processes, then the inconsistent relationships between short-term memory and pitch imitation in the current literature could be attributed to the reliance on verbal measures of short-term memory capacity in previous studies.

Current evidence for the domain specificity of music and language processing has been mixed (for reviews, see Peretz, 2009; Peretz, Vuvan, Lagrois, & Armony, 2015; Schulze & Koelsch, 2012). On the one hand, the modular approach, motivated by case studies, states that music and language rely on separate neural resources (Fodor, 1983; Peretz & Coltheart, 2003). Recently, neuroimaging studies have shown that pitch and verbal representations rely on overlapping neural regions, but the patterns of activation within these regions differ for speech and music (Norman-Haignere, Kanwisher, & McDermott, 2015; Peretz et al., 2015; Rogalsky, Rong, Saberi, & Hickok, 2011; Schulze, Zysset, Mueller, Friederici, & Koelsch, 2011). On the other hand, Mantell and Pfordresher (2013) found that individuals who were inaccurate at imitating pitch in melodies tended also to be inaccurate at imitating pitch in speech, a result that aligns with previous work suggesting that pitch processing relies on shared resources for speech and music, as opposed to an encapsulated system (Mantell & Pfordresher, 2013;

Patel, 2003; Pfordresher & Brown, 2009; Mantell & Pfordresher, 2013).

To address the question of dimension specificity in pitch imitation, we incorporated measures of imagery and short-term memory that focused on either pitch or verbal processing. For our auditory imagery measures, we included tasks that required the formation of pitch images or images based on the sequencing of phonemes into pronounceable nonwords (verbal imagery). To measure short-term memory, we included memory span tests for digits (verbal memory) as well as for pitch. We used the Seattle Singing Accuracy Protocol (Demorest & Pfordresher, 2015; Demorest et al., 2015) to measure pitch imitation ability and pitch discrimination thresholds and to assess music background. Figure 2 lays out four tests of basic cognitive processes (boxes) that could be related to the accuracy of vocal pitch imitation. With this design, we focused on two key hypotheses. One hypothesis is that vocal pitch imitation might participate in a network of *process-specific* mechanisms: Correlations between pitch imitation ability and performance on auditory imagery tasks—both pitch-based and verbal—but not performance on memory span tasks would support this hypothesis. The second hypothesis is that vocal pitch imitation might participate in a network of *dimension-specific* mechanisms: Correlations between pitch imitation ability and performance on tasks that involve the processing of pitch—either imagery or short-term store—and not tasks that involve the processing of verbal information would support this hypothesis. Finally, we also addressed whether auditory imagery mediates the relationship between auditory short-term memory and pitch imitation errors. This analysis was motivated by the proposed roles of auditory imagery and an auditory store in the sensorimotor loop shown in Fig. 1, in combination with the close association between self-reported imagery vividness and working memory that has been proposed by Baddeley (Baddeley, 2012; Baddeley & Andrade, 2000).

Method

Participants

A sample of 285 participants were recruited through the Introductory Psychology subject pool at the University at Buffalo and completed the study in exchange for course credit. All participants reported normal hearing, speech, and reading abilities. Eleven of the participants were removed from the sample who exhibited extremely high pitch discrimination thresholds above 1200 cents (range: 3435–9187 cents). Thresholds in this range imply that the participant did not understand the instructions. An

additional 58 participants were removed because they lacked data from at least one predictor measure, due to coding or experimenter error. In the final sample of 216 participants, 33 of the participants reported having vocal training (15%), 61 reported having six or more years of music experience (28%), and 62 reported no music experience (29%). All participants were native English speakers, and only three speakers reported experience with a language other than English (French, Russian, and Yoruba).¹

Apparatus

Participants completed vocal production tasks in either an acoustically treated room or a sound-attenuated recording booth (Whisper Room SE 2000). Auditory stimuli were presented over Sennheiser HD 280 Pro headphones, while vocal performance was recorded using a Shure WH30 headset microphone connected through a Lexicon Omega I/O box.

Materials

Seattle Singing Accuracy Protocol (SSAP) This protocol consists of a battery of measures designed to assess singing accuracy, pitch perception, and musical background (Demorest & Pfordresher, 2015; Demorest et al., 2015). Participants first completed tasks designed to measure their comfortable singing range, which consisted of singing a familiar song and producing a sustained “comfort pitch” using the syllable “doo” for several seconds.

Participants then completed pitch imitation tasks in which they would listen to a target stimulus and then reproduce it by singing. The trials were blocked by the type of task and progressed from trials in which the target stimulus comprised a single vocal pitch (ten trials), to a single piano pitch (ten trials), and then to a four-note vocal pitch sequence (six trials). Vocal pitches were based on recordings of a male or a female trained singer (vocal gender matched to the participant) using the syllable “doo.” The range of sung pitches was adapted to the participant’s comfort pitch as exhibited in the warm-up phase.

After the vocal imitation tasks, participants sang a familiar song from memory (first with lyrics and then on “doo”), completed an adaptive pitch discrimination task (Loui, Alsop, & Schlaug, 2009; Loui, Guenther, Mathys, & Schlaug, 2008), and completed a musical background questionnaire. The SSAP analyzes pitch errors automatically by extracting the F0 trajectory with the YIN algorithm (De De Cheveigné & Kawahara, 2002), identifying tone onsets on the basis of

amplitude, removing the beginning and ends of tones so as to avoid the influence of “scoops,” and taking the median F0 from inside each tone. Errors are defined as pitches with a difference from the target greater than 50 cents (half a musical semitone).²

Short-term memory measures We estimated short-term memory span for pitch and verbal sequences using an adaptive two-up, one-down staircase procedure based on Williamson and Stewart (2010). Participants heard a sequence of items that were randomly sampled from the set of ten items, followed by a second sequence of the same length. The second sequence was either the same as the first or contained two items that were reversed in serial position. The tone sequences started at a length of two tones, and the digit sequences started at a length of four digits, reflecting the different baseline levels of difficulty for the tasks as determined by Williamson and Stewart. The task ended after eight reversals of sequence length (increasing to decreasing or vice versa). Each participants’ final span length was calculated by averaging across sequence lengths for the last six reversals.

The *pitch span task* used ten triangle-waveform tones between 262 Hz (C4) and 741 Hz (F#5), such that each tone was separated by at least one whole step (200 cents) in the scale. The rise and fall times of each tone were linear over a duration of 20 ms, and total tone durations were 500 ms. The sequences were atonal, thus preventing chunking strategies.

The *digit span task* used digits ranging from 0 to 9, spoken by a male speaker in American English with a Midland dialect for male participants, or by a female speaker with an Inland North dialect for female participants. Each digit was resynthesized to have a constant fundamental frequency of 102 Hz, for male stimuli, or 193 Hz, for female stimuli, and a constant intensity. The duration of each spoken digit was approximately 500 ms.

Imagery measures We estimated the accuracy with which participants could generate auditory images online on the basis of pitch or phonetic (verbal) qualities.

The *pitch imagery task* was adapted from Gelding, Thompson, and Johnson (2015); an example is shown in Fig. 3. Each trial began with a C-major scale that provided a tonal context (each tone used the same acoustic dimensions as in the pitch span task), followed by a single cue tone. Participants then heard a sequence of three pitches, accompanied by visually presented arrows that pointed up or down, with the arrow direction corresponding to the direction of the change in pitch from the previous to the current tone. Next, participants observed a

¹ Performance on the verbal imagery task was above a cutoff of 1 *SD* below the mean (.73; 73% correct) for each of the multilingual speakers: Russian speaker, 1 (100%); Yoruba speaker, .79 (79%); French speaker, .96 (96%).

² The ± 50 -cent cutoff marks the categorical boundary surrounding a target pitch in equal-tempered tuning.

Fig. 3 Illustration of a trial in the pitch imagery task

sequence of one to three more arrows and had to imagine what pitch should accompany these arrows, based on the tonal context and the preceding auditory pitches. Finally, participants heard a probe tone (with no accompanying arrow) and judged whether that pitch matched the last pitch that was imagined (the correct response for the example in Fig. 3 would be “yes”).

The *verbal imagery task* used line drawings (Duñabeitia et al., 2018) representing high-frequency monosyllabic words that consisted of three phonemes. For half of the words, the phonological and orthographic representations were congruent (e.g., *bus*), and for the other half the phonological and orthographic representations were incongruent (e.g., *thumb*), to prevent the use of orthographic strategies. Participants first completed a familiarization phase, in which they were presented a line drawing on a computer screen while the word illustrated by the picture was spoken over headphones. In an experiment trial, participants observed the three line drawings presented on a computer screen in a sequential, additive order (see the example in Fig. 4). Beneath each line drawing was a row of three dashes, representing three phonemes, with the phoneme to be imagined indicated by an asterisk. Participants were asked to combine the phonemes from each word in order to imagine a pronounceable nonword. Participants then heard a spoken nonword over headphones (produced by the same talker as for the male digit span stimuli) and were asked to indicate whether this stimulus matched the imagined nonword.

Procedure

After providing informed consent, participants were seated in an acoustically treated experiment room or sound-attenuated vocal booth. They completed the SSAP after being instructed by an experimenter on the appropriate posture for singing. Participants then completed the imagery and short-term memory tasks, with the order of these tasks counterbalanced through a Latin square design. Then they completed music and language background questionnaires, as well as the Bucknell Auditory Imagery Scale (BAIS; Halpern, 2015), a self-report measure of auditory imagery ability consisting of Vividness and Control subscales. Finally, participants were debriefed and given course credit for their participation.

Results

Descriptive statistics for each task are presented in the Appendix Table 3. We focus here on correlations across the measures, given that the primary purpose of this study was to measure what core cognitive abilities are associated with singing accuracy. Bivariate correlations across all measures are shown in Table 1. Pitch imitation error rates were significantly correlated with all variables, with the exceptions of digit span and BAIS Control. Correlations between the predictor variables were also examined.

Next we evaluated the relative contributions of significant the predictor variables for pitch error rates that were associated with the theoretical framework shown in Fig. 2 (this was limited to the online measures, and thus excluded the BAIS subtests). We conducted a five-step hierarchical linear regression with pitch imitation error rate as the dependent variable. The variables were ordered according to their conceptual status, with those that had a less critical theoretical role or that already enjoyed strong support from previous research being entered before variables that had a more critical and hypothetical role. The results are shown in Table 2. Not surprisingly, musical experience (Step 1) and pitch discrimination threshold (Step 2) were strong predictors of pitch error rates during singing. More importantly, short-term memory for pitch (Step 3) and online pitch imagery (Step 5) contributed significantly beyond the other predictors. However, performance on the verbal imagery task (Step 4) did not contribute beyond the predictors entered at earlier steps, despite having a significant bivariate relationship (Table 1). This significant bivariate relationship may in large part reflect the shared variance between verbal imagery performance and short-term memory for pitch (also shown in Table 1).³ Other orderings of the predictors, carried out after the reported ordering for purposes of testing alternatives, yielded the same results.

Finally, we tested the relationship between short-term memory and imagery from the vocal sensorimotor loop model

³ We did not include the BAIS self-report measures in the hierarchical regression, in order to focus on measures of online processing. However, a follow-up exploratory analysis with BAIS Vividness entered before verbal imagery showed that self-reported vividness did not account for a significant portion of variance in pitch imitation errors.

Fig. 4 Illustration of a trial in the verbal imagery task

(Fig. 1) through a mediation analysis. We reasoned that auditory imagery is necessary to make pitch representations in short-term memory accessible for associations with vocal motor planning, on the basis of the hypothesis that pitch imagery supports sensorimotor translation. This analysis was predicated on the fact that pitch imagery and pitch span were correlated with each other, in addition to the finding that both pitch imagery and pitch memory were unique predictors of pitch imitation error rates. A nonparametric bootstrapping procedure with 1,000 simulations revealed a significant indirect effect, $B = -.01$, 95% CI = $-.02, -.003$. Large pitch spans were associated with high scores for pitch imagery, which in turn were associated with low rates of pitch imitation errors; see Fig. 5. The proportion of the indirect effect with respect to the total effect was .20. In addition, the direct effect was also significant, $B = -.04$, 95% CI = $-.09, -.03$; high pitch span thresholds were associated with low rates of pitch imitation errors, above and beyond the contribution of pitch imagery. In sum, these results indicate that pitch imagery was a partial mediator of the effect of pitch span on pitch imitation errors.

Discussion

We have reported associations between pitch imitation ability and measures of basic cognitive functions that are hypothesized to play a role in the vocal sensorimotor loop, shown in Fig. 1 (Berkowska & Dalla Bella, 2009; cf. Pfordresher et al., 2015). Accurate imitators exhibited larger short-term memory spans for pitch and greater accuracy in generating auditory images of musical pitch than did inaccurate imitators. These relationships were found to be independent of other factors, including pitch perception and musical experience, which, not surprisingly, also correlated with pitch imitation ability. We also showed that pitch imagery partially mediates the relationship between pitch span capacity and pitch imitation errors, providing evidence for a possible pathway by which auditory short-term memory and auditory imagery may be involved in vocal pitch imitation.

The significance of the correlations of pitch imitation ability with imagery and memory span measures is two-fold. First, these associations draw connections between the theoretical construct of the vocal sensorimotor loop model (Fig. 1) and basic cognitive functions. Our results support auditory imagery's role in sensorimotor translation of auditory perception and vocal motor planning (Pfordresher et al., 2015) and suggest that short-term memory is necessary for maintaining a representation of a target sequence for purposes of planning a series of motor actions (cf. Palmer & Pfordresher, 2003). Whereas other studies have shown an association between singing accuracy and self-report of auditory imagery vividness (Greenspon et al., 2017; Pfordresher & Halpern, 2013), this is the first study we know of to have reported an association with the accuracy of pitch images formed online. Importantly, we also established the unique contribution of pitch short-term memory capacity in pitch imitation. Second, the general cognitive capacities we outlined—and the measures we used here—are not limited to music performance contexts. Vocal imitation of pitch may therefore not reflect an encapsulated ability, but may instead reflect the integration of several coordinated functions.

Although the cognitive functions associated with vocal pitch imitation serve broader purposes, it is also important to note that these associations are specific to the representation of pitch and are not measures of imagery and memory for phonological information. At first glance, these results appear consistent with claims that music cognition, whether with respect to perception or performance, relies on domain-specific resources that are not shared with speech (e.g., Peretz & Coltheart, 2003). However, pitch also plays an important role in language processing. The present design does not distinguish between language-based and music-based uses of pitch, in that the tasks only involved musical pitch processing. Other studies of vocal pitch imitation have suggested that imitations of sung and spoken pitch rely on shared resources (e.g., Mantell & Pfordresher, 2013; Pfordresher & Brown, 2009; Wisniewski, Mantell, & Pfordresher, 2013). Therefore,

Table 1 Bivariate correlations between tasks

	Music Experience	Pitch Discrimination	Digit Span	Pitch Span	Verbal Imagery	Pitch Imagery	BAIS Vividness	BAIS Control
Pitch Errors	-.23***	.22**	-.07	-.25***	-.19**	-.26***	-.16*	-.11
Music Exp.		-.13	-.01	.09	.05	.22**	.18**	.18**
Pitch Disc.			-.07	-.17*	-.33***	-.11	-.28***	-.17*
Digit Span				.14*	.20**	.12	-.04	-.03
Pitch Span					.33***	.26***	.07	.14*
Verbal Imagery						.26***	.01	.01
Pitch Imagery							.01	.03
BAIS Vividness								.63***

$N = 216$. * $p < .05$, ** $p < .01$, *** $p < .001$.

we only make claims about *dimension*-specific processing of pitch, rather than *domain*-specific processing of music. In this context, it is worth noting that an earlier study did report an association of verbal span with singing accuracy, which was not found here (Greenspon et al., 2017; cf. Christiner & Reiterer, 2013). However, the association was found using backward span and was only found for singing tasks that involved a particularly demanding reproduction task—that is, reproduction of mental transformations of melodies. Here we found that simple reproduction of melodies correlates specifically with pitch span.

Interestingly, there was a general tendency for participants to perform more accurately with verbal than with pitch stimuli in both the imagery and memory span tasks (cf. Benassi-Werke, Queiroz, Araujo, Bueno, & Oliveira, 2012). It is possible that these differences relate to the nature of the representations underlying pitch and verbal items. With respect to the span task, digits are highly familiar items and reflect both visual (numeric and orthographic) and auditory (phonological) representations. Pitches, on the other hand, may reflect both

visual (musical notation) and auditory representations for musicians (≥ 6 years music experience, 28% of the sample); however, for nonmusicians (72% of the sample), pitches reflect auditory-only representations. Similarly, phonological images, even for the nonwords used here, may relate to orthographic visual representations that enhance performance. At this point, however, it is hard to draw firm conclusions from these tasks, given that the tests we used were not calibrated with respect to the perceptual salience of the individual items.

In conclusion, the present research helps clarify the role of basic cognitive processes in the vocal imitation of pitch. The role of auditory imagery in vocal imitation has been established by recent research (Greenspon et al., 2017; Pfordresher & Halpern, 2013; Pfordresher et al., 2015). However, a novel finding from the present research is that auditory imagery relies on pitch-specific resources that uniquely relate to pitch imitation ability. In addition, the present study established a pitch-specific role of auditory short-term memory in vocal pitch imitation, further supporting the notion that pitch and phonological representations may rely on separate memory stores within the phonological loop (Schulze et al., 2011).

Table 2 Five-step hierarchical linear regression of significant predictors on pitch imitation errors

Step/Predictor	Step 1 β	Step 2 β	Step 3 β	Step 4 β	Step 5 β
Musical experience	-.23**	-.20**	-.19**	-.19**	-.16*
Pitch discrimination		.20**	.16*	.14*	.14*
Pitch span			-.20**	-.18**	-.15*
Verbal imagery				-.07	-.03
Pitch imagery					-.16*
Adjusted R^2	.05**	.08***	.12***	.12***	.13***
F for ΔR^2	—	9.28**	9.79**	0.99	5.49*

$N = 216$, * $p < .05$, ** $p < .01$, *** $p < .001$, reporting standardized regression estimates and adjusted R^2

Fig. 5 Mediation analysis with pitch imagery as the mediator of the relationship between pitch span and pitch imitation. Unstandardized regression estimates are presented, along with standard errors in parentheses. ** $p < .01$, *** $p < .001$

Open Practices Statement This research was supported by National Eye Institute grant no R01 EY025275.

Appendix

Table 3 Descriptive statistics for all measures

Task	Scale	Min	Q1	Median	Q3	Max	Mean	SD
Pitch reproduction	Error rate	0	.17	.38	.63	1	.41	.27
Pitch discrimination	Cents	1.33	10.57	21.08	82.82	1002.14	70.83	135.1
Music experience	Years	0	0	3	6	15	3.62	3.68
Digit span	#Digits	4	5.33	6	7.04	14.33	6.29	1.42
Pitch span	#Pitches	2.16	3.49	4.33	5.16	10	4.48	1.21
Pitch imagery	Proportion correct	.28	.52	.60	.65	1	.59	.13
Verbal imagery	Proportion correct	.46	.79	.87	.92	1	.84	.12
BAIS Vividness	Likert 1:7	2.57	4.36	5	5.8	7	5.05	0.98
BAIS Control	Likert 1:7	1.64	4.5	5.07	5.93	7	5.1	1

References

- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, 129, 126–145.
- Benassi-Werke, M. E., Queiroz, M., Araujo, R. S., Bueno, O. F., & Oliveira, M. G. M. (2012). Musicians' working memory for tones, words, and pseudowords. *Quarterly Journal of Experimental Psychology*, 65, 1161–1171.
- Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology*, 5, 69–83.
- Christiner, M., & Reiterer, S. M. (2013). Song and speech: Examining the link between singing talent and speech imitation ability. *Frontiers in Psychology*, 4, 874. <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, 1182–1189.
- De Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, 111, 1917–1930.
- Demorest, S. M., & Pfordresher, P. Q. (2015). Seattle Singing Accuracy Protocol—SSAP [Measurement instrument]. Available at <https://ssap.music.northwestern.edu/>
- Demorest, S. M., Pfordresher, P. Q., Dalla Bella, S., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015). Methodological perspectives on singing accuracy: An introduction to the special issue on singing accuracy (Part 2). *Music Perception*, 32, 266–271.
- Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, 71, 808–816. <https://doi.org/10.1080/17470218.2017.1310261>
- Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
- 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, e0121809. <https://doi.org/10.1371/journal.pone.0121809>
- Gordon, E. E. (1975). *Learning theory, patterns, and music*. Buffalo, NY: Tometic Associates.
- Greenspon, E. B., Pfordresher, P. Q., & Halpern, A. R. (2017). Pitch imitation ability in mental transformations of melodies. *Music Perception*, 34, 585–604.
- Halpern, A. R. (2015). Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music, Mind and Brain*, 25, 37–47.
- 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, 76–97.
- Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron*, 67, 713–727.
- 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, 2477–2485.
- Lima, C. F., Lavan, N., Evans, S., Agnew, Z., Halpern, A. R., Shanmugalingam, P., ... Scott, S. K. (2015). Feel the noise: Relating individual differences in auditory imagery to the structure and function of sensorimotor systems. *Cerebral Cortex*, 25, 4638–4650. <https://doi.org/10.1093/cercor/bhv134>
- Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *Journal of Neuroscience*, 29, 10215–10220.
- Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology*, 18, 331–332.

- Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition*, 127, 177–202. <https://doi.org/10.1016/j.cognition.2012.12.008>
- 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>
- Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, 88, 1281–1296.
- Palmer, C., & Pfordresher, P. Q. (2003). Incremental planning in sequence production. *Psychological Review*, 110, 681–712. <https://doi.org/10.1037/0033-295X.110.4.683>
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6, 674–681.
- Peretz, I. (2009). Music, language and modularity framed in action. *Psychologica Belgica*, 49, 157–175.
- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6, 688–691. <https://doi.org/10.1038/nn1083>
- Peretz, I., Vuvan, D., Lagrois, M. É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B*, 370, 20140090.
- Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of tone deafness. *Music Perception*, 25, 95–115.
- Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics*, 71, 1385–1398. <https://doi.org/10.3758/APP.71.6.1385>
- Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *Journal of the Acoustical Society of America*, 128, 2182–2190.
- Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review*, 20, 747–753.
- Pfordresher, P. Q., Halpern, A. R., & Greenspon, E. B. (2015). A mechanism for sensorimotor translation in singing. *Music Perception*, 32, 242–253.
- Pfordresher, P. Q., & Larrouy-Maestri, P. (2015). On drawing a line through the spectrogram: How do we identify deficits of vocal pitch imitation? *Frontiers in Human Neuroscience*, 9, 271. <https://doi.org/10.3389/fnhum.2015.00271>
- Pfordresher, P. Q., & Mantell, J. T. (2009). Singing as a form of vocal imitation: Mechanisms and deficits. In J. Louhivouri, T. Eerola, S. Saarikallio, T. Himberg, & P.-S. Eerola (Eds.), *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009)* (pp. 425–430). Jyväskylä, Finland: University of Jyväskylä.
- Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. *Journal of Neuroscience*, 31, 3843–3852.
- Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences*, 1252, 229–236.
- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, 32, 771–783. <https://doi.org/10.1002/hbm.21060>
- Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, 33, 1433–1454.
- Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory*, 18, 657–669.
- Wisniewski, M. G., Mantell, J. T., & Pfordresher, P. Q. (2013). Transfer effects in the vocal imitation of speech and song. *Psychomusicology: Music, Mind, and Brain*, 23, 82–99.
- Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: Musical imagery and auditory cortex. *Neuron*, 47, 9–12. <https://doi.org/10.1016/j.neuron.2005.06.013>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.