The Effect of Visual and Auditory Feedback on Adult Poor-Pitch Remediation

Jacob Berglin

Bienen School of Music, Northwestern University

Peter Q. Pfordresher

Department of Psychology, University at Buffalo State University of New York

Steven Demorest

Bienen School of Music, Northwestern University

Author Note

Jacob Berglin https://orcid.org/0000-0002-3036-541X

Peter Pfordresher https://orcid.org/0000-0002-7024-0785

Jacob Berglin is now at the School of Music, Florida International University.

Steven Demorest, formerly of Northwestern University, passed away on September 22nd, 2019.

We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Jacob Berglin, 10910 SW 17th St, Miami, FL 33199. Email: jberglin@fiu.edu

Abstract

Previous research has led to the hypothesis that poor-pitch singing is the result of a weakness in the auditory/vocal loop. The present study evaluated this hypothesis in a training paradigm that used visual feedback to augment potentially faulty auditory-vocal associations. Following pretest with the Seattle Singing Accuracy Protocol (SSAP), participants were randomly assigned to one of three 20-minute training conditions: 1) Visual feedback training with auditory doubling, in which participants could both see and hear real-time feedback showing the relationship between their sung pitch and the target, 2) Auditory feedback training, where participants relied only on airborne auditory feedback from their own voice, and 3) Control training which involved imitation of speech from a foreign language instruction recording. After training, the SSAP was administered again as a posttest measure. There was a general improvement from pre to post-test across all groups. However, the effect of training was only significant for participants who received visual feedback training, with greater gains in visual training than either of the other conditions. This pattern of results was particularly pronounced for performance on 4-note melodies in the SSAP, in contrast to single pitch matching. Visual feedback may facilitate accuracy by substituting for inaccurate auditory-motor associations. The fact that training, even over a very short time-span, can have significant effects on singing underlines the importance of practice and supports the hypothesis that singing is a learned skill that can benefit from experience and may not simply reflect an inherited talent.

Keywords: Singing accuracy, Augmented feedback, Skill development

Introduction

Recent research suggests that most healthy adults have the capability to match pitch accurately while singing (Dalla Bella et al., 2007; Pfordresher & Demorest, 2021). While most adults do indeed display this ability to sing “accurately” – that is, a dominant tendency to match pitch within a music semitone – a significant minority of the population fails to meet this criterion under testing conditions (Welch, 1979). While estimates of the frequency of poor-pitch singing vary depending on the measure used (Berkowska & Dalla Bella, 2013; Hitchins & Peretz, 2012; Pfordresher et al., 2010; Pfordresher & Larrouy-Maestri, 2015), recent estimated suggest that nearly 30% of the population may fit this description (Pfordresher & Demorest, 2021). Although poor-pitch singing is often referred to as “tone deafness,” and while poor-pitch singers typically describe themselves in this way (Cuddy et al., 2005; Wise & Sloboda, 2008), only a small portion of this group likely suffers from true tone deafness (congenital amusia), which comprises approximately 1.5% of the population (Peretz & Vuvan, 2017). Most poor singing likely reflects difficulty mapping perceptual representations to vocal motor patterns (Hutchins & Moreno, 2013; Hutchins & Peretz, 2015; Pfordresher et al., 2015; Pfordresher & Mantell, 2014). The remediation of poor-pitch singing to date has been challenging for educators, especially in large group settings, who may feel “forced” to resort to solutions that are demeaning to the singer and suppress future participation in music (Welch, 2006). In the present study, we test an approach to remediation that is theoretically grounded and based on scientific literature.

A critical question in the development of interventions for inaccurate singers is in the general effectiveness of different teaching strategies. While some recent research indicates that singing ability may be linked to one’s genotype (Tan, McPherson, Peretz, Berkovic, & Wilson, 2014), other studies have shown that singing accuracy in children may be improved by intervention (Demorest, Nichols, & Pfordresher, 2018). While there may be truth to the suggestion that singing reflects a talent that exists independently of any practice-based effects (Howe, Davidson, & Sloboda, 1998; McPherson & Williamon, 2015), it is important to understand that singing accuracy is associated with musical self-concept (Demorest, Kelley, & Pfordresher, 2017), and that singing accuracy changes across the lifespan in a way that may be associated with musical participation (Demorest & Pfordresher, 2015b). Thus, for a large segment of the populace, singing accuracy may be understood as a learned motor skill that is reliant on continued use, in a “use it or lose it” fashion.

Past remediation research has focused on classroom instructional strategies such as individual and small group attention, a focus on musculature development, or pedagogical approaches such as “speech to song,” where novice singers work first to gain control of the speaking voice before the instructor identifies a personal note from which to extend the singing range (Svec, 2017). Still, researchers have hypothesized that poor-pitch singing is the result of a weakness in the auditory/vocal loop (Berkowska & Dalla Bella, 2009; Pfordresher, Demorest, et al., 2015). One possible means of helping inaccurate singers is to provide feedback during imitation to augment their perceptual acuity. While reviews of previous studies have found mixed benefits from auditory feedback (Nichols, 2018), research focusing on visual feedback has been more promising (Hoppe, Sadakata, & Desain, 2006; Paney & Tharp, 2021; Welch, Howard, & Rush, 1989; Wilson, Lee, Callaghan, & Thorpe, 2008). This may be a result of the theory that in order to sing accurately, an individual must have an *internal model* of the auditory-vocal system, something not fully developed in all individuals (Pfordresher, Halpern, & Greenspon, 2015; Pfordresher & Mantell, 2014). If poor singers have a missing, under-developed, or inaccurate internal model, visual feedback may help correct that model and the overall link between vocal/motor output and auditory feedback.

Still, due to limitations in effect size and study design, much extant research involving feedback suggests the need for a more controlled examination of interventions. For instance, the highly variable nature of a school setting involving thousands of student participants and multiple teachers administering interventions, makes conclusions difficult to determine. Likewise, studies without a randomization element are limiting. Addressing these design concerns, the purpose of this pretest-posttest control group design was to compare improvements in singing accuracy after a brief practice period in which participants matched pitch patterns with visual and auditory feedback, matched the same patterns with only auditory feedback, or experienced a control condition involving the imitation of a foreign language. We hypothesized that pitch accuracy among less accurate singers would improve after training with visual and auditory feedback, but not for other training conditions.

Method

Participants

Participants (=84) were recruited at two research sites: the University at Buffalo (n = 56), and Northwestern University (n = 28). Participants from Buffalo received credit for Introduction to Psychology, whereas participants from Northwestern participated in exchange for monetary compensation. All were adults (M age = 21.75, SD = 8.02, range = 18-70). Mean reported years of training on an instrument was 0.49 (SD = 1.35, range = 0-7), and 84% of participants (n = 67) reported no musical training. Mean reported years of training in singing was 0.13 (SD = 0.6, range = 0-4), and 94% of participants (n = 75) reported no singing training. Participants self-reported gender for the purposes of identifying a comfortable register for their singing voice: 43

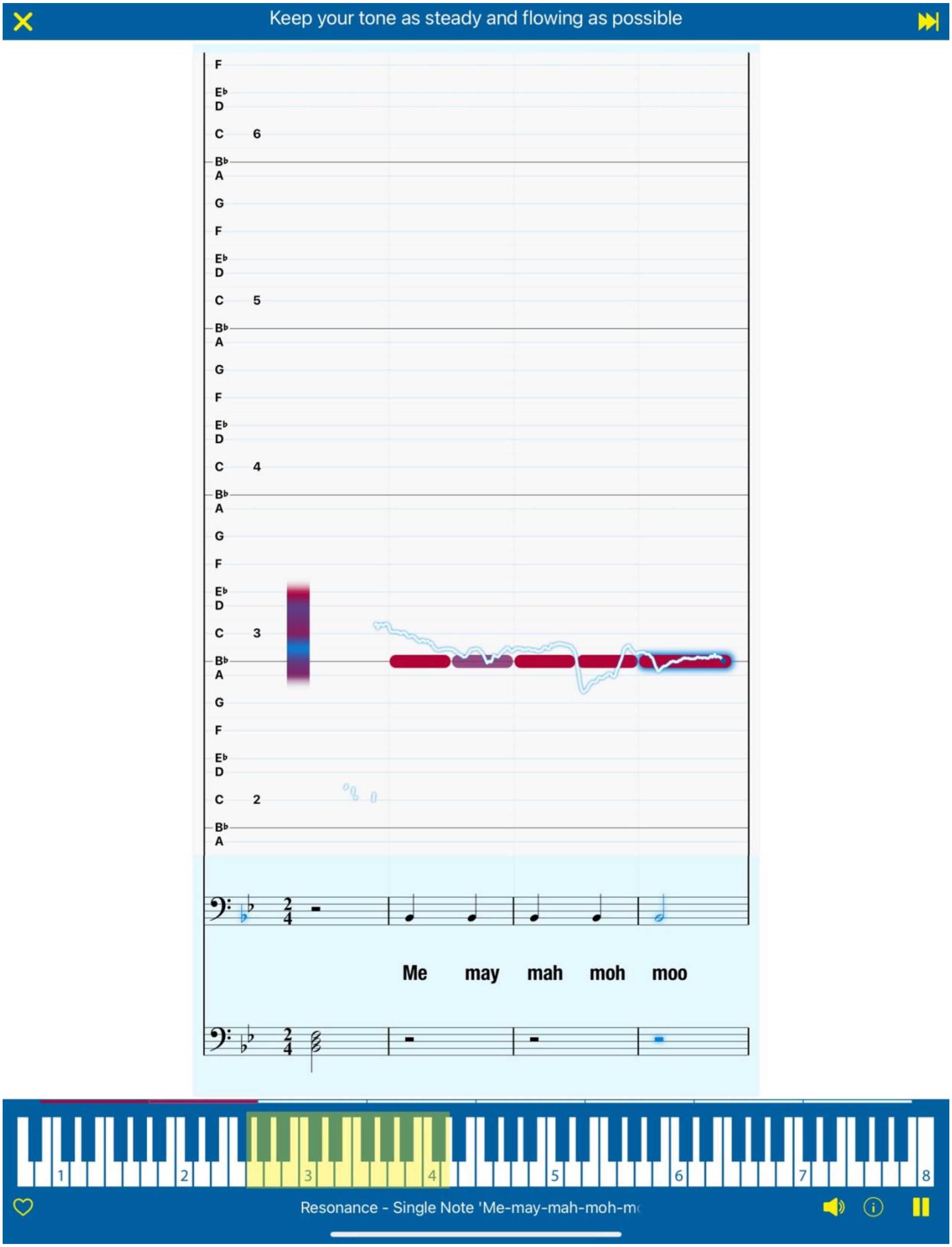
reported being female and 37 reported being male. Criteria for inclusion were based on self-reporting no history of hearing loss, vocal motor disorder, or other neurological disorder. Included participants also demonstrated a pitch discrimination threshold of less than 100 cents (see Materials and Procedure), given that larger thresholds may have indicated a pitch perception deficit such as congenital amusia (Hyde & Peretz, 2004). We originally identified *below 85% accuracy* on the Seattle Singing Accuracy Protocol (SSAP, explained below) as the inclusion threshold for participation; however, after analyzing participant inclusion criteria during data collection, we realized that an *80% accuracy threshold* would allow for more compelling findings. After applying these inclusion criteria, 75 participants were retained for data analyses reported here.

After qualifying for the study, participants were randomly assigned to one of the three 20-minute training conditions that involved vocal imitation of an auditory target: 1) *Visual feedback training* (n = 25) in which participants could both hear and view real-time visual feedback showing the relationship between their sung pitch and the target, 2) *Auditory feedback training* (n = 23) which was identical to the first condition with the exclusion of visual feedback, 3) *Control training* (n = 27) which involved imitation of speech from a foreign language instruction recording (Italian).

Materials and Procedure

Participants first completed the SSAP as a pretest measure. The SSAP is a standardized measure of singing accuracy developed by a team of researchers (Demorest & Pfordresher, 2015a; Demorest et al., 2015; Pfordresher & Demorest, 2020), that runs online (https://ssap.music.northwestern.edu) or via Matlab (Mathworks, Natick, MA). Individuals taking the SSAP first indicate their gender for purposes of vocal doubling (a female vocal model for all except post-pubescent male voices). Next, individuals complete warm-up tasks used to establish their comfortable vocal range. Three imitation tasks follow and constitute the most critical SSAP tasks: Imitation of single vocal pitches, imitation of single piano pitches, and imitation of 4-note vocal melodies. Next, individuals sing a familiar song from memory (first using text, and then a second time using “doo” for each syllable), complete an adaptive pitch discrimination task using a staircase procedure designed to identify one’s threshold for discrimination (Loui, Guenther, Mathys, & Schlaug, 2008), and complete a series of demographic and singing self-concept questions. The main focus of the present study is on imitation tasks, which provide the most direct measure of pitch accuracy in singing (Pfordresher & Demorest, 2020).

The first experimental group involved Visual training. These participants used an iPad app called *Erol Singers’ Studio* to practice singing with augmented visual feedback. Visual feedback came in the form of piano roll notation showing target pitches, on which was superimposed a continuous line displaying change in the participants vocal fundamental frequency (f_0). Participants were able to use the visual discrepancy between the line representing their pitch and target pitches for self-correction. After an initial assessment of vocal range involving basic singing exercises, participants were placed into one of five voice-class categories: High (beginning on C4), Middle-High (A3), Middle (F3), Middle-Low (C3), and Low (A2). Following a 10-minute training session on use of the app, participants followed by performing four sets of vocal exercises in a randomly counterbalanced order: two based on repetition of a single pitch based on a MIDI piano model, one involving successive pitches in stepwise motion, and one based on arpeggiated movement. Figure 1 is a screenshot from the Erol Singing Studio app showing a single-pitch matching exercise, while Figure 2 demonstrates a triadic exercise. Training took approximately 20 minutes to complete and was largely self-directed, with an experimenter in the room to assist at the participant's request.



Keep your tone as steady and flowing as possible

Figure 1: A screenshot from the Erol Singing Studio app demonstrating a single-pitch matching exercise. The uneven line represents the singer's attempt at matching the reference pitch.

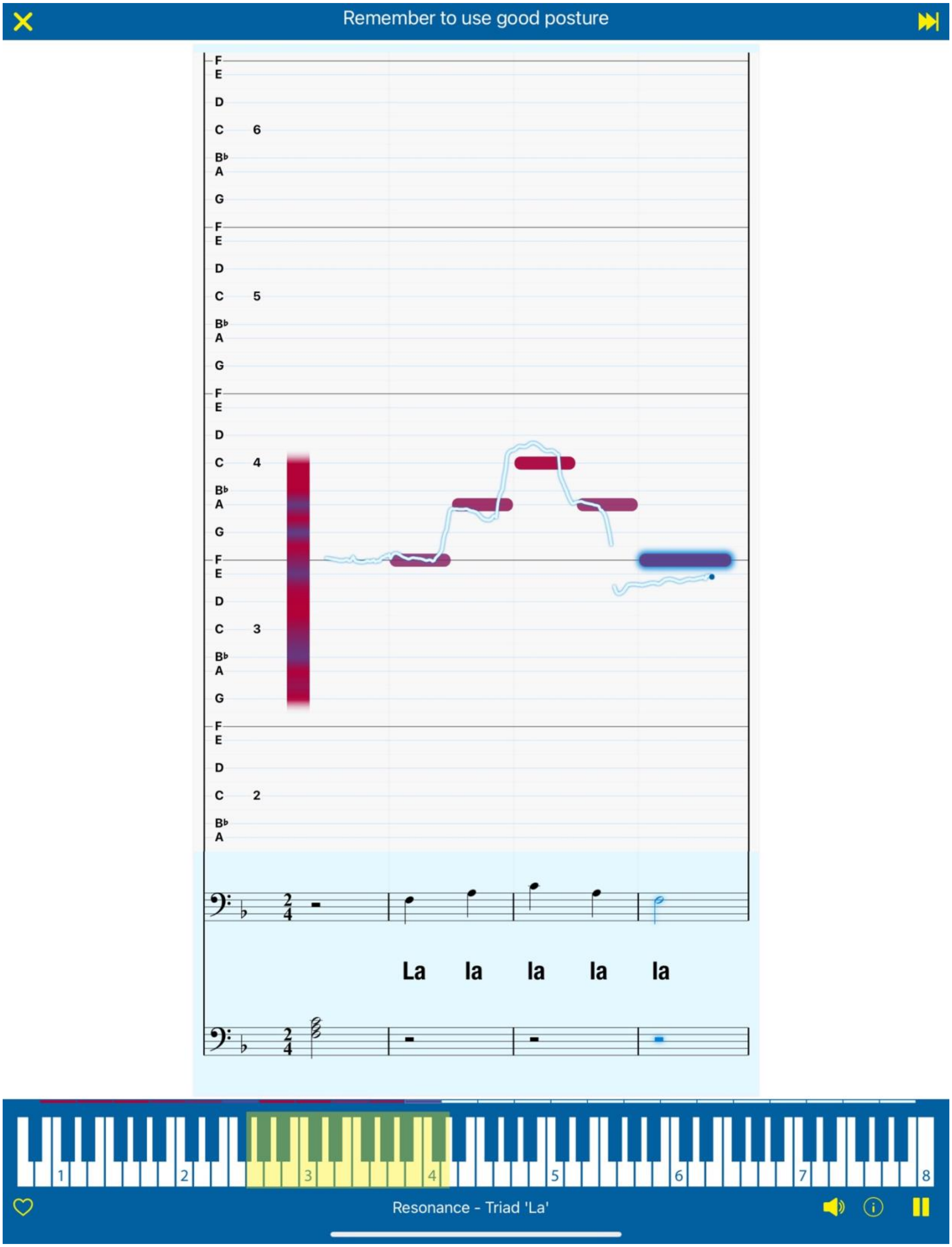


Figure 2: A screenshot from the Erol Singing Studio app demonstrating a triadic matching exercise. The uneven line represents the singer’s attempt at matching the reference pitch.

The second experimental group involved specifically Auditory training. This treatment was identical to the Visual training group with the single exception that participants did not view the iPad while performing, and thus received no visual feedback. Instead, after range identification identical to the Visual condition, exercises were placed in a playlist on the iPad in mp3 format and participants sang along with only the MIDI piano audio. In this condition, participants' only opportunity for feedback was airborne auditory feedback from their own voice.

The third experimental group involved Control training. In this condition, participants listened to recordings of Italian speech from a language learning program and were instructed to imitate this speech. This was designed to be an active control condition that involved auditory-vocal imitation but did not involve musical pitch matching.

Each participant engaged in one of the three training conditions. Immediately after training, participants completed the SSAP a second time as a posttest measure. No corrective feedback of any sort was provided to participants after the pretest or the posttest.

Equipment

Experimental sessions at the Buffalo site were carried out in a quiet room that contained a WhisperRoom sound attenuated booth (WhisperRoom, Inc.) that was used to record pre and posttest recordings of the SSAP. The experimenter ran the SSAP over the internet on a PC computer, and gave the participants verbal instructions from the protocol's visual interface. Participants heard the experimenter's instructions and auditory stimuli through Sennheiser HD 280 headphones, and wore a Shure WH30 headset microphone for recordings. During training trials, participants exited the sound booth and sat next to a table that had an iPad stand that contained an iPad Pro, used to run the Erol Singers Studio app or launch the mp3 playlists containing the auditory-only exercises or the Italian language learning audio, depending on condition.

Experimental sessions at Northwestern were carried out in a quiet room with the experimenter present in the room with the participant. Participants used a Macbook laptop set up by the experimenter to take the SSAP, using a high-quality computer headset with microphone. Instructions were given verbally during the instruction period prior to training. During training, participants used the same headset/microphone attached to an iPad Pro while using the Erol Singers Studio app or listening to the mp3 playlists, depending on condition.

Analysis

We measured singing accuracy at pre- and post-test by aggregated produced errors across all pitch imitation trials in the SSAP. The median fundamental frequency (f_0) for the central portion of each sung note was extracted and contrasted with the corresponding intended f_0 value from the target using an automated procedure (for details, see Pfordresher & Demorest, 2020). If the absolute difference between these values exceeded 50 cents (100 cents = 1 semitone), the sung pitch was categorized as an error. The mean of these discrete values (1 or 0 for each sung note) across all produced notes in all imitation trials yields the mean percent of errors produced, which was our primary measure of accuracy. We also analyzed pitch deviation scores (the mean absolute difference between target and imitated f_0 value), which yielded a similar pattern of results to those reported here.

The critical questions under exploration here were whether practice in vocal pitch matching improves singing accuracy, and whether visually augmenting feedback (as executed by the Erol program) enhances this improvement. Our analyses include omnibus Analysis of Variance across all treatment conditions, planned contrast designed to test whether improvement was found within any of the three treatment conditions, and planned contrasts between each condition that involves vocal pitch training (auditory or visual) with the control condition.

Results

We begin by presenting error rates across all imitation trials on the SSAP before and after the brief training session. Figure 3 represents these results in a boxplot, with individual lines representing change from pre to post-test for individual participants.

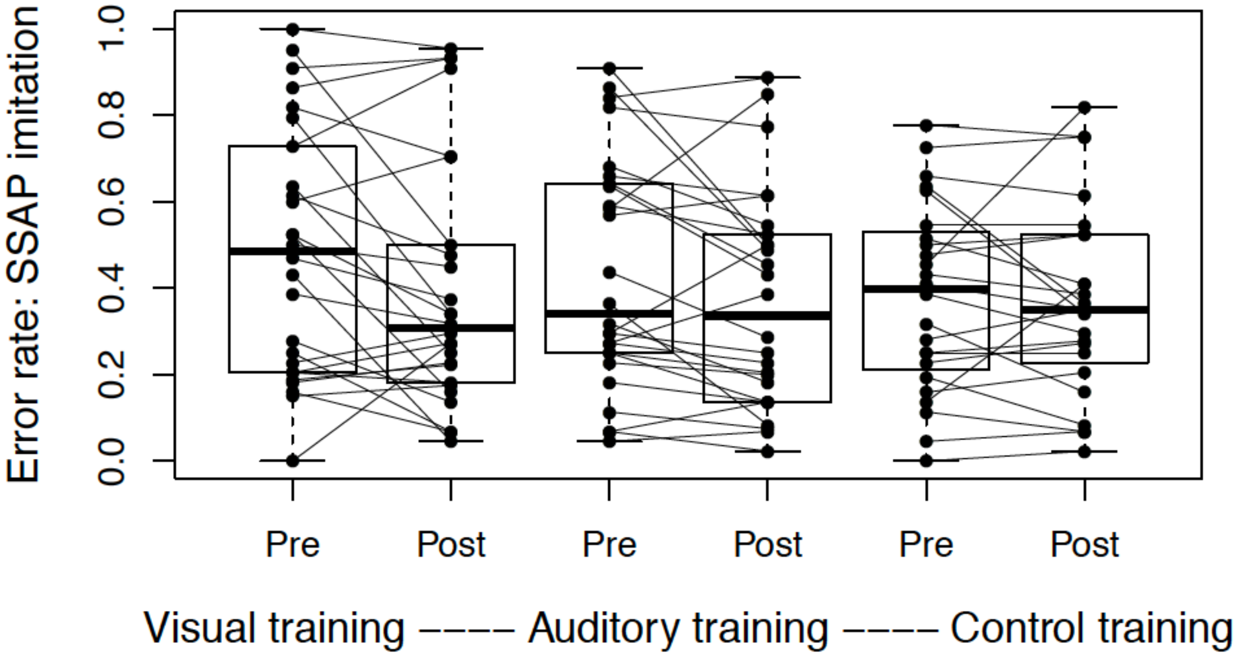


Figure 3: Boxplots displaying error rates as a function of training condition and pre vs. posttest. Dots indicate means for individual participants and lines connecting dots show change from pre to post for individuals.

As can be seen, there was a noticeable tendency for errors to decrease (i.e., for performance to improve) from pre to post-test for participants engaging in visual training, whereas general improvement was not as apparent in the other training groups. At the same time, there was considerable variability across participants, including variability within the pre-test condition on its own. The difference in the frequency of pre-test errors did not differ significantly across the three training conditions ($p = .38$); nevertheless, we considered it prudent to reduce the influence of this variability source by analyzing improvement relative to performance at pre-test.

As such we analyzed *percent improvement* from pre to posttest by computing the difference in error rates across pre and post-test (positive indicating reduced error rates after training) and dividing that difference by the error rate exhibited at pre-test. A score of .5 indicates 50% improvement from pre-test (e.g., an error rate of .5 at pre-test and an error rate of .25 at posttest). Two participants (one from Visual treatment, one from Control treatment) produced no errors at pre-test and had to be excluded (the resulting percent improvement measure would be infinity); it was also prudent to remove participants like this, given that improvement from training would be impossible.

Figure 4 displays percent improvement for each participant grouped by condition, with random variability added to the x-axis to avoid overlapping data points. Means and 95% confidence intervals are superimposed over the data points for each condition, and a guideline at zero is displayed for reference.

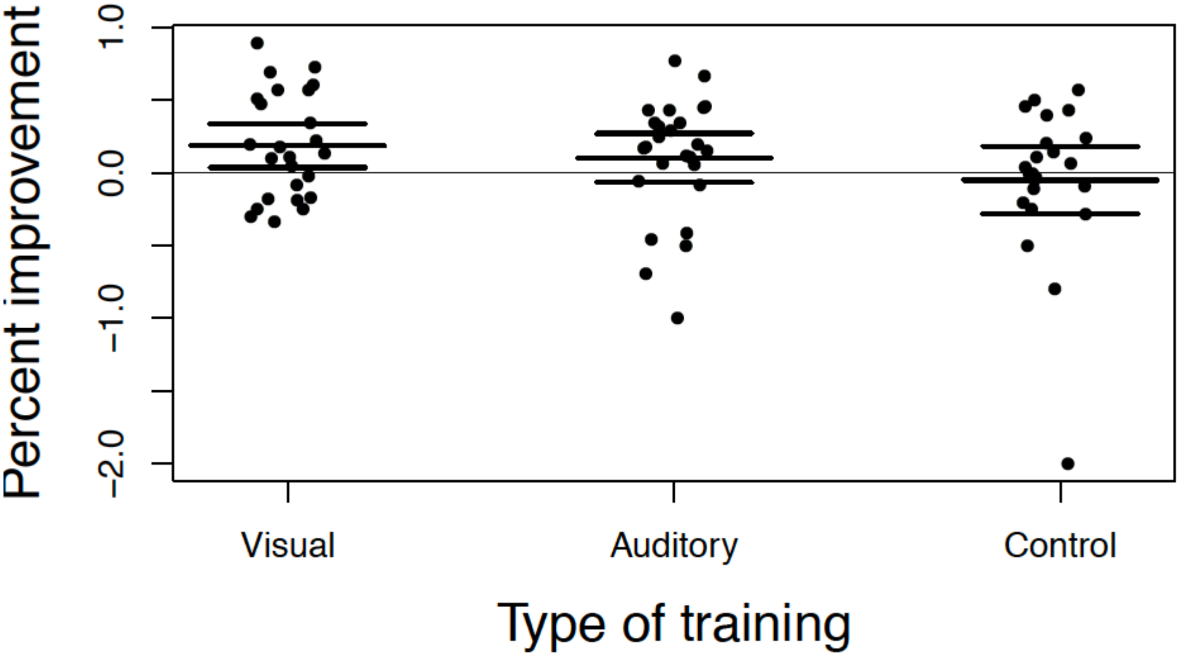


Figure 4: Strip chart displaying percent improvement scores (pre/post difference normalized by pretest) for individual participants (dots). Horizontal lines for each treatment condition display means (middle) and 95% confidence intervals (upper/lower). The light horizontal line across the graph separates scores showing improvement (positive) from those showing declines (negative) from pre to post.

As can be seen, the visual and auditory training conditions improved performance on average, whereas mean performance did not improve for the control condition. In addition, the 95% confidence intervals for the visual training condition did not overlap zero, which suggests that the true population mean for this condition would indicate improvement. In keeping with these observations, mean performance for the visual training group was significantly greater than zero, $t(24) = 2.55$, p = .009, whereas the others were not (auditory training, $p = .12$, control $p = .67$). With respect to comparisons across conditions, the contrast between visual training and control was significant, $t(46) = 1.8$, p = .039, $d = 0.49$, whereas the contrast between auditory training and control was not significant ($p = .14$, $d = 0.29$). However, the one-factor ANOVA comparing all three conditions was not significant and indicated a small effect size, $F(2,71) = 1.75$, $p = 0.181$, $\eta_p^2 = 0.047$. As observed before, the effects of training were highly variable, which likely accounts for the mixed results.

In evaluating the data in Figure 4, one is struck by the outlier in the control condition whose performance accuracy diminished by nearly 200% at posttest (this participant produced a .14 error rate at presets and .41 at posttest). As noted previously, the inclusion of participants who make very few errors at pretest constitutes a non-optimal test of how training improves pitch accuracy. Both of these issues were addressed in a subsequent analysis that limited the sample to participants making at least 20% errors at pretest across all pitch-imitation tasks, which did not reduce the overall sample size too greatly (Visual N = 21, Control N = 18, Auditory N = 21). Results from this subset are shown in Figure 5, which is formatted like Figure 4 and uses the same measure of performance (percent improvement). Note the different Y-axis scaling.

Figure 5: Strip chart formatted as Figure 4, but for those participants who made at least 20% errors at pretest.

As was the case for the whole sample, significant improvement was found for participants who engaged in visually augmented training, *t*(20) = 2.85, p = 0.005, but not for participants in the control condition (*p* = .17). The contrast between visual training and control was also significant, *t*(37) = 1.75, *p* = .044, *d* = 0.54. Unlike results from the whole sample, significant improvement was found for participants who engaged in auditory training, *t*(20) = 1.83, *p* = .041, although this improvement was not significantly greater than the control condition (*p* = .170, *d* = 0.36), thus further supporting the interpretation that visual and auditory training yields greater benefits than auditory only.

The SSAP includes three different kinds of vocal imitation conditions: Imitation of single pitches based on a vocal timbre, imitation of single pitches based on a piano timbre, and imitation of 4-note patterns based on a vocal timbre. We next addressed the influence of training on these different imitation tasks. Because auditory and visual training tasks used a vocal timbre, it is possible that training may transfer selectively to vocal imitation tasks in the SSAP. Likewise, because most of the visual and auditory training trials involved imitating pitch patterns, it is possible that training will yield its largest effect on the imitation of 4-note patterns.

Table 1 displays means and 95% confidence interval widths for percent improvement scores among participants making at least 20% errors at pretest. Visual feedback training led to significant improvement for imitation of 1-note and 4-note vocal pitch. However, significant improvement for 1-note vocal pitches was also found for the control condition, though not for 4-note vocal patterns. This led to a significant contrast between Vocal and Control training for the 4-note vocal patterns. No significant improvement within an imitation condition was found for Auditory training. All told, this more detailed analysis suggests that visual feedback training yields a selective advantage for the imitation of 4-note melodies.

Table 1: Means (95% confidence intervals) for percent improvement among participants scoring >= 20% errors at pre-test, along with p-values for pairwise contrasts across training conditions.

	Type of training				
Imitation task	*Visual*	*Auditory*	*Control*	*V > C*	*A > C*
1-note vocal	0.**35** (0.26)	0.14 (0.45)	0.**45** (0.20)	0.731	0.923
1-note piano	0.16 (0.31)	0.12 (0.26)	0.10 (0.29)	0.370	0.438
4-note vocal	0.**22** (0.18)	0.11 (0.17)	-0.07 (0.22)	**0.022**	0.107

Note: Column V>C displays one-tailed p-values for the difference between Visual and Control training; column A>C displays one-tailed p-values for difference between Auditory and Control training conditions. Bold type for these columns indicates significant p-values. Other use of boldface indicates mean percent improvement that is significantly greater than zero (one-tailed) for a given training and imitation condition.

One may reasonably question how dependent these results are on our selection of a 20% criterion for inclusion of data in most of the analyses reported here. We chose this criterion in advance of analyzing data, although we admit that it may seem arbitrary. In order to address this issue, we re-analyzed results for percent improvement using cutoffs ranging from 0% to 40% errors at pre-test, in 2% increments. We stopped at 40% because by this point the sample sizes within each group were approaching n = 10, at which point statistical results become highly questionable. Significant improvement was found for the Visual training condition for each of these groups at $p < .02$ each, whereas no cutoff yielded significant improvement for Control training and significance for Auditory training was found on 57% of cutoffs and never fell below $p = .02$. The significant contrast between Visual and Control training was found only for cutoffs ranging from 20-26%, whereas contrasts between Auditory and Control conditions were not significant at any cutoff. Thus, although there was some dependency of the significant Visual versus Control contrast on the cutoff we used, other results were stable regardless of the cutoff used.

Discussion

We found statistically significant improvement in singing accuracy after a brief training session. Improvement was specific to interventions that involved vocal pitch matching, and thus did not simply reflect a re-test advantage, and were most reliable for training with augmented visual feedback. Effects varied greatly across participants, some of whom did not improve or even worsened, but on average pitch accuracy with visually augmented training improved by approximately 20% relative to pretest. This improvement is considerable given the brevity of training as well as the fact that most participants participated to fulfill course credit and were not necessarily motivated to improve their singing skills. In this final section, we reflect on the significance of these findings as well as on prospects for future research.

Current results support the conceptualization of singing as a complex combination of inborn “gifts” along with environmental and developmental factors (McPherson & Williamon, 2015). While a small number of individuals may indeed suffer from congenital perceptual issues, participants in this study, especially those in the visual condition, displayed a tendency toward improvement over even a very brief period of remediation. This supports the idea that singing is a learned motor skill, which can be remediated even in untrained singers who may not choose to pursue further musical talent development. This complements other research showing effects of training on singing accuracy among children (Demorest et al., 2018), and the positive effects of musical-self-assessments on musical participation and singing accuracy (Demorest et al., 2017). Although it is likely that there is a connection between singing ability and one's genotype (Tan et al., 2014), environmental influences clearly play a large role, and more emphasis on the effect of practice may prove fruitful in understanding vocal motor control as well as more beneficial, given the potentially positive effect of even amateur musical participation (Clift et al., 2010; Kreutz, Bongard, Rohrmann, Hodapp, & Grebe, 2004; Stewart & Lonsdale, 2016).

The most robust and reliable improvement was found when participants received augmented visual feedback during training through the commercial software package Erol Singer's Studio. This finding converges with other studies that found similar effects for training of singing accuracy (Hoppe et al., 2006; Paney & Tharp, 2021; Welch et al., 1989; Wilson et al., 2008), and more generally with research in the motor control literature suggesting an advantage for visual feedback in early learning stages for complex tasks (Sigrist, Rauter, Riener, & Wolf, 2013). This result is important in both practical and theoretical senses. In a practical sense, visual feedback helps clarify the kind of errors a singer might make so that the instructor does not have to use vague terminology ("too high"). The potential of evaluating one's errors through the visual modality also allows the novice singer the ability to process error information outside the context of vocal motor control. Theoretically, these results are consistent with the view that inaccurate singing results from poorly-formed auditory-vocal associations that prevent inverse modeling of vocal motor control necessary to match pitch (Berkowska & Dalla Bella, 2009; Hutchins & Peretz, 2012; Loui, Alsop, & Schlaug, 2009; Pfordresher, Halpern, et al., 2015; Pfordresher & Mantell, 2014).

The control condition in the present study also involved vocal imitation, but of an unfamiliar language. The fact that this condition did not benefit vocal pitch imitation was anticipated but is also important theoretically. The modular framework for music and language proposed by Peretz and Coltheart (2003), suggests that the processing of phonetic information is performed independently of pitch coding, and is consistent with the dissociation of training effects found here. Other research on vocal imitation has found that the use of phonetic information as a segmental cue in imitation does increase the accuracy of pitch imitation, and was interpreted as consistent with a more integrated framework for phonetic and pitch control (Mantell & Pfordresher, 2013). It is possible that the advantage found by Mantell and Pfordresher does not arise from the use of a common imitative mechanism for pitch and articulation, but instead on the meaningfulness of variations in phones and pitch.

Closer analysis of how training influenced specific imitation tasks within the SSAP revealed that the effect of augmented visual feedback selectively enhanced the imitation of 4-note vocal melodies, rather than matching single pitches based on either a piano or a vocal timbre. This makes sense when one considers how participants may relate the visual representation of pitch to vocal motor control. Whereas the absolute vertical position of the visual representation in Erol may be difficult to map onto vocal motor control, comparisons between the relative change in the visual representation of targets and concurrent patterns in the visual representation of vocal pitch may allow easier mapping. As such, benefits of visual feedback training may have been most pronounced on the auditory-vocal mapping of relative pitch patterns more so than absolute pitch height.

A key limitation of the present research involved a lack of control over the process of training. Although we controlled the tasks participants experienced during training, various factors may have influenced the effectiveness of training for a single individual. In any self-directed experimental task, researchers cannot be exactly sure to what level each participant is motivated to succeed. A similar design with a longer treatment may address this, as well as help determine the effectiveness of the Visual training condition. Another potential limitation stems from limitations of the device. The *Erol Singers' Studio* app also uses the piano as an auditory doubling device. Research indicates that some singers respond better to vocal models (Hutchins, Larrouy-Maestri, & Peretz, 2014; Hutchins & Peretz, 2012; Pfordresher & Mantell, 2014; Watts & Hall, 2008). A revised Visual training condition that allows for vocal doubling, or a more song-based approach, might be more effective. Additionally, more research is necessary to identify the underlying causes of poor-pitch singing in order to design better interventions that target specific functional deficits based on the many underlying components of vocal pitch matching (Berkowska & Dalla Bella, 2009; Hutchins & Moreno, 2013; Pfordresher, Demorest, et al., 2015; Pfordresher, Halpern, et al., 2015; Zarate, 2013).

In conclusion, these results suggest that even a 20-minute practice session can improve singing accuracy when one's own auditory feedback is augmented by visual feedback that presents on-line knowledge of results. Visual feedback may facilitate accuracy by substituting for inaccurate auditory-motor associations. The fact that training, even over a very short time-span, can have significant effects on singing testifies to the importance of practice and the hypothesis that singing is a learned skill that can benefit from experience and may not simply reflect an inherited talent (Demorest et al., 2018).

Acknowledgement

This work was supported in part by NSF Grant BCS-1848930 and a Fulbright Canada Research Chair to author PQP. We are grateful to Catherine Moore and Alexia Feiner for help with data collection in Buffalo. Author SD passed away before the writing stage of this work, but took the lead role in designing the study and planning data analyses. Authors JB and PQP have done their best to draft a report that would have met his approval.

References

Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology, 5*, 69-83.

Berkowska, M., & Dalla Bella, S. (2013). Uncovering phenotypes of poor-pitch singing: The Sung Performance Batter (SPB). *Frontiers in Psychology, 4*, 714.

Clift, S., Hancox, G., Morrison, I., Hess, B., Kreutz, G., & Stewart, D. (2010). Choral singing and psychological wellbeing: Quantitative and qualitative findings from English choirs in a cross-national survey. *Journal of Applied Arts and Health, 1*, 19-34.

Cuddy, L. L., Balkwill, L.-L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare: A study of "tone deafness" among university students. *Annals of the New York Academy of Sciences, 1060*, 311-324.

Demorest, S. M., Kelley, J., & Pfordresher, P. Q. (2017). Singing Ability, Musical Self-Concept, and Future Music Participation. *Journal of Research in Music Education, 64*, 405-420.

Demorest, S. M., Nichols, B., & Pfordresher, P. Q. (2018). The effect of focused instruction on young children's singing accuracy. *Psychology of Music, 46*, 488-499.

Demorest, S. M., & Pfordresher, P. Q. (2015a). Seattle Singing Accuracy Protocol - SSAP [Measurement instrument]. https://ssap.music.northwestern.edu/.

Demorest, S. M., & Pfordresher, P. Q. (2015b). Singing accuracy development from K-adult: A comparative study. *Music Perception, 32*, 293-302.

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.

Hoppe, D., Sadakata, M., & Desain, P. (2006). Development of real-time visual feedback assistance in singing training: A review. *Journal of Computer Assisted Learning, 22*, 308-316.

Howe, M. J., Davidson, J. W., & Sloboda, J. A. (1998). Innate talents: reality or myth? *The Behavioral and brain sciences, 21*, 399-407; discussion 407-342.

Hutchins, S., Larrouy-Maestri, P., & Peretz, I. (2014). Singing ability is rooted in vocal-motor control of pitch. *Attention, Perception & Psychophysics, 76*, 2522-2530.

Hutchins, S., & Moreno, S. (2013). The Linked Dual Representation model of vocal perception and production. *Frontiers in Psychology, 4*, 825.

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

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15*, 356-360.

Kreutz, G., Bongard, S., Rohrmann, S., Hodapp, V., & Grebe, D. (2004). Effects of Choir Singing or Listening on Secretory Immunoglobulin A, Cortisol, and Emotional State. *Journal of Behavioral Medicine, 27*, 623-635.

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *The 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*, R331-332.

Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of speech and song *Cognition, 127*, 177-202.

McPherson, G. E., & Williamon, A. (2015). Building gifts into musical talents. In G. E. McPherson (Ed.), *The child as musician: A handbook of musical development* (pp. 340-360.). Oxford: Oxford University Press.

Paney, A. S., & Tharp, K. L. (2021). The effect of concurrent visual feedback on adult singing accuracy. *Psychology of Music, 49*, 360-370.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*, 688-691.

Peretz, I., & Vuvan, D. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics, 25*, 625-630.

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., & Demorest, S. M. (2021). The prevalence and correlates of accurate singing. *Journal of Research in Music Education, 69*, 5-23.

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). London: Routledge.

Pfordresher, P. Q., Demorest, S. M., Dalla Bella, S., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015). Theoretical perspectives on singing accuracy: An introduction to the special issue on singing accuracy (Part 1). *Music Perception, 32*, 227-231.

Pfordresher, P. Q., Halpern, A. R., & Greenspon, E. B. (2015). A Mechanism for Sensorimotor Translation in Singing: The Multi-Modal Imagery Association (MMIA) Model. *Music Perception, 32*, 293-302.

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.

Pfordresher, P. Q., & Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive Psychology, 70* 31-57.

Sigrist, R., Rauter, G., Riener, R., & Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review, 20,* 21-53.

Stewart, N. A. J., & Lonsdale, A. J. (2016). It's better together: The psychological benefits of singing in a choir. *Psychology of Music, 44*, 1240-1254.

Tan, Y. T., McPherson, G. E., Peretz, I., Berkovic, S. F., & Wilson, S. J. (2014). The genetic basis of music ability. *Frontiers in Psychology, 5*.

Watts, C. R., & Hall, M. D. (2008). Timbral influences on vocal pitch-matching accuracy. *Logopedics Phoniatrics Vocology, 33*, 74-82.

Welch, G. F. (1979). Poor pitch singing: A review of the literature. *Psychology of Music*, 7, 50-58.

Welch, G. F. (2006). Singing and vocal development. In G. McPherson (Ed.), *The child as musician: A handbook of musical development* (pp. 311-329). New York: Oxford University Press.

Welch, G. F., Howard, D. M., & Rush, C. (1989). Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychology of Music, 17*, 146-157.

Wilson, P. H., Lee, K., Callaghan, J., & Thorpe, C. W. (2008). Learning to sing in tune: Does real-time visual feedback help? *Journal of Interdisciplinary Music Studies, 2*, 157-172.

Wise, K., & Sloboda, J. A. (2008). Establishing an empirical profile of self-defined 'tone deafness': Perception, singing performance, and self-assessment. *Musicae Scientiae, 12*, 3-23.

Zarate, J. M. (2013). The neural control of singing. *Frontiers in Human Neuroscience, 7*, 237.