BRIEF REPORT

Auditory imagery and the poor-pitch singer

Peter Q. Pfordresher · Andrea R. Halpern

Abstract The vocal imitation of pitch by singing requires one to plan laryngeal movements on the basis of anticipated target pitch events. This process may rely on auditory imagery, which has been shown to activate motor planning areas. As such, we hypothesized that poor-pitch singing, although not typically associated with deficient pitch perception, may be associated with deficient auditory imagery. Participants vocally imitated simple pitch sequences by singing, discriminated pitch pairs on the basis of pitch height, and completed an auditory imagery self-report questionnaire (the Bucknell Auditory Imagery Scale). The percentage of trials participants sung in tune correlated significantly with self-reports of vividness for auditory imagery, although not with the ability to control auditory imagery. Pitch discrimination was not predicted by auditory imagery scores. The results thus support a link between auditory imagery and vocal imitation.

Keywords Auditory imagery · Vocal imitation · Poor-pitch singing · Perception and action

A critical problem that any vocal instructor must face is an extension of the fact that a singer is unable to observe laryngeal motor movements used to control pitch. Even if one were able to observe such movements (i.e., through a laryngoscope), it is doubtful that such visual information would be as useful as, for instance, being able to observe the hand posture of an expert violinist. Vocal instructors therefore incorporate auditory and visual imagery in pedagogy (e.g., "imagine the pitch coming out of the top of your head"). In the music education

P. Q. Pfordresher (✉)
Department of Psychology, University at Buffalo, 355 Park Hall,
Buffalo, NY 14260, USA
e-mail: pqp@buffalo.edu

A. R. Halpern (✉)
Psychology Department, Bucknell University, 1 Dent Drive,
Lewisburg, PA 17837, USA
e-mail: ahalpern@bucknell.edu

community, such use of auditory imagery is commonly referred to as *audiation* (cf. Brodsky, Kessler, Rubinstein, Ginsborg, & Henik, 2008).

The research reported here tested a broader implication of such practices. Specifically, if vocal imitation of pitch depends on auditory imagery and is not just aided by it, individuals who are deficient with respect to vocal imitation, referred to here as *poor-pitch singers* (Hutchins & Peretz, 2012; Pfordresher & Brown, 2007; Welch, 1979), should show deficient auditory imagery abilities. We compared individuals' performance on a simple vocal pitch imitation task (singing a four-note monotone sequence) with responses on a questionnaire concerning the vividness and controllability of auditory images (the Bucknell Auditory Imagery Scale; henceforth, BAIS) and pitch discrimination ability.

Theoretically, the present research rests on the assumption that vocal imitation of pitch relies on an *inverse model* of the auditory–vocal system and that the contents of this inverse model are auditory images (see Fig. 1). An inverse model is an internal model of a perception/action system that guides motor planning on the basis of the anticipated outcomes (goals) of the action, similar to the ideomotor hypothesis of William James (cf. Shin, Proctor, & Capaldi, 2010). Thus, the link from motor planning to perceptual output operates inversely, a process that may or may not be conscious. Because it must incorporate schematic knowledge about sensorimotor associations, the inverse model in Fig. 1 is controlled by motor planning, as well as perceptual input (target pitches). We propose (following Keller, 2012) that mental images (here, auditory images) are the contents of such an inverse model. Thus, in the context of pitch imitation, auditory images function as a bridge between perception and action that serve to guide planning of laryngeal movements. As such, limitations in auditory imagery may indicate a poorly functioning inverse model, leading to poor pitch imitation ability.

These theoretical assumptions are supported by recent neuroimaging research, which suggests that auditory imagery leads to activation in motor planning areas (for
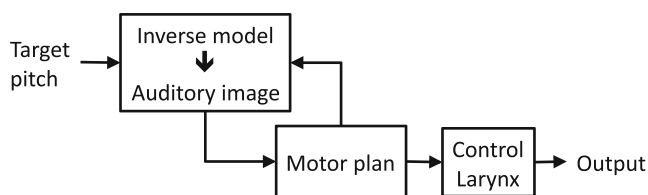
Fig. 1 Representation of the vocal pitch imitation system, including the hypothesized inverse internal model. (Adapted from Hickok, Houde, & Rong, 2011)

reviews, see Halpern, 2001; Zatorre & Halpern, 2005). For instance, a recent study demonstrated enhanced activation in premotor areas and basal ganglia when participants anticipated a forthcoming auditory event, and these activations were positively correlated with self-reported vividness of imagery of the anticipated tune (Leaver, Van Lare, Zielinski, Halpern, & Rauschecker, 2009). Thus, auditory imagery may allow one to carry out the kind of sensorimotor transformations that allow inverse modeling of perception and action (cf. Herholz, Halpern, & Zatorre, 2012; Zatorre, Halpern, & Bouffard, 2010). Behaviorally, past research has shown that better auditory imagery ability facilitates consistency of expressive timing in performance (Clark & Williamon, 2011) and more effective practice (Brown & Palmer, 2012; Highben & Palmer, 2004).

Thus, if vocal imitation of pitch depends on auditory imagery, deficiencies in the vocal imitation of pitch among poor-pitch singers should be correlated with deficiencies in auditory imagery. Poor-pitch singing is a production-related deficit characterized by a general tendency to mistune absolute pitch, compression of interval size for relative pitch, and imprecision of pitch production (for reviews, see Berkowska & Dalla Bella, 2009; Pfordresher & Mantell, 2009; Welch, 1979). Poor-pitch singers typically do not demonstrate deficient pitch discrimination abilities, nor do they appear to be deficient with respect to vocal motor control outside the context of imitation. Thus, it has been hypothesized that a deficient inverse model of the auditory–vocal system may cause poor-pitch singing (Pfordresher, 2011). We collected self-reports of auditory imagery ability to test this hypothesis directly.

Method

Participants

Participants (N = 138) were recruited from the introduction to psychology subject pool at SUNY Buffalo. Data reported here come from a prescreening procedure conducted each semester (first described by Pfordresher & Brown, 2007, Experiment 2). Seventy-six participants (55 %) were female, 61 were male,

and 1 participant elected not to report gender. The mean age was 19 years (range: 18–27). Sixty-three participants (46 %) reported a native language other than English; of these, 33 (24 %) reported first learning a tone language (Mandarin, Vietnamese, or Cantonese). We recorded musical background using reported years of experience, summed across all instruments and voice. Across participants, the mean years of summed experience reported was 2.9 years (range: 0–43; the participant reporting 43 years of training had 10 or more years of experience with three instruments and voice). A total of 37 participants (27 %) reported 5 or more years of experience and were considered musicians (mean for this group = 8.8 years). The sample was predominantly nonmusician; most participants (N = 101, 73 %) reported fewer than 5 years of musical experience, and 72 participants (52 %) reported no experience whatsoever.

Apparatus

Participants were recorded while sitting inside a Whisper Room SE 2000 recording booth. Instructions and stimuli were delivered to participants via a pair of Sennheiser HD 280 Pro headphones. Participant recordings were made using a Shure PG58 microphone. Sound levels were controlled using a Lexicon Omega I/O box. The experiment was conducted on a 3.4-GHz PC running Windows XP. The experimental procedure was run using MATLAB scripts (The Mathworks, Natick, MA).

Procedure

Participants were run in a single session that took approximately 30 min. The session was divided into the following blocks, run in the order in which they are described.

Warm-up

The procedure began with a series of warm-up trials consisting of extemporaneous speech ("describe what you had for dinner last night"), reading a page of text, and singing a familiar tune. Participants were then asked to produce vocal sweeps: a continuous change in pitch from the lowest note an individual can comfortably sing up to the highest note he or she can comfortably sing and then back down. Finally, participants produced a single sustained pitch that was comfortable for them to produce (described as a note the participant may use to start singing a song). The experimenter then identified the nearest pitch on the C-major scale (these pitch classes were used to simplify the procedure), and this *comfort pitch* was used as the basis for experimental trials.

Vocal imitation

Following warm-up trials, six experimental vocal imitation trials were completed. On each trial, the participant first listened to a sequence of four identical pitches (a monotone sequences) and then reproduced this sequence by singing on the syllable "la." Pitches on the first and last trials were equal to the participant's comfort pitch. Trials 2 and 3 included pitches that were two and four scale steps, respectively, higher than the comfort note, using pitches from the C-major scale. Trials 4 and 5 comprised pitches that were two and four scale steps below the comfort pitch, respectively, also drawn from C-major pitches.

Following each trial, the results of a MATLAB pitch-tracking algorithm were displayed on the screen, showing the participant's produced $f0$, as well as boundaries representing deviations of ±100 cents surrounding the target $f0$. This criterion was based on categorizations of poor-pitch singers reported elsewhere (cf. Dalla Bella, Giguere, & Peretz, 2007; Pfordresher & Brown, 2007; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010). The experimenter coded each trial as *in tune* if the majority of sampled $f0$ values were within these boundaries and out of tune if not. This procedure was initially adopted to allow easy categorization of participants for inclusion in other experiments as accurate or poor-pitch singers; here, we adopt these categorizations as an additional measure to acoustic analyses of performances (also reported).

Pitch discrimination

On each pitch discrimination trial, participants heard two pure tones and reported which tone was higher in pitch. Pitches were arranged around a standard frequency of 524 Hz (C5), and could differ from this standard by plus or minus 13, 25, 50, or 100 cents or did not differ. Participants were presented with four trials from each condition, two each of ascending and descending changes (plus four no-change conditions), in a random order. Due to temporal constraints and equipment failures, pitch discrimination data were not obtained for 9 participants.

Auditory imagery

Finally, participants completed the BAIS.[1] This instrument comprises 28 items divided equally into subscales for the vividness of auditory imagery (e.g., the vividness of hearing a trumpet play "Happy Birthday") or control of auditory imagery (e.g., the ease with which one can change the auditory image of a choir of children into a choir of adults). Items probe musical, verbal, and environmental sounds. All

---

[1] This questionnaire is available on request from the second author.

responses were made on a scale of 1–7, with 7 indicating *more vivid* or *easier to change*. Because the BAIS was administered last, time constraints prevented some participants from completing it: 120 (87 %) completed the vividness subscale (which appeared first), and 114 (83 %) completed both subscales.

Results

BAIS responses

Analyses focused on whether individual differences in vocal imitation are related to self-report measures of auditory imagery, measured by the BAIS. Responses to the BAIS were highly reliable (Cronbach's alpha for all items = .910, for the vividness subscale = .833, for the control subscale = .909). The mean imagery score across both subscales (from 1 to 7) was 4.78 (SD = 1.05). Mean ratings of vividness (M = 4.90, SE = 0.09) were slightly but significantly higher than ratings of control (M = 4.66, SE = 0.11), $t(113)$ = 2.72, $p < .01$. Responses on the subscales (averaged across items) were significantly correlated, $r(112)$ = 0.50, $p < .01$.

Vocal imitation

Figure 2a shows the relationship between the percentage of trials sung in tune (based on the procedure described earlier) and the mean rating given for each participant on the vividness subscale of the BAIS. There was a significant positive correlation, $r(118)$ = .28, $p < .01$, as predicted. The corresponding relationship between performance and the imagery control subscale of the BAIS, shown in Fig. 2b, was positive but was not significant, $r(112)$ = .11, $p > .10$.

In addition to categorization of trials in tune, we also measured the mean absolute deviation of produced from target $f0$ across each trial. These production measures were highly correlated, $r(118)$ = −.85, $p < .01$, but differed with respect to the treatment of intonation as categorical (percentage of trials in tune) or continuous (pitch deviation). As with the percentage of trials in tune, mean absolute pitch deviation scores were significantly correlated with the vividness of auditory imagery, $r(118)$ = −.24, $p < .01$, but were not significantly correlated with control of imagery, $r(112)$ = −.08, $p > .10$. In order to simplify further analyses, we determined which measure most effectively predicted the vividness of auditory imagery by regressing vividness ratings on both the percentage of trials in tune and mean absolute pitch error. The multiple regression was significant ($p < .01$). A partial correlation analysis indicated that neither measure of production accounted for a significant portion of the variance independently of the other (although proportion

Fig. 2 Scatterplots displaying the relationship between the mean percentage of trials (out of six) sung in tune; *x*-axis), and mean responses on the BAIS questionnaire concerning vividness of auditory imagery (**a**, *y*-axis) or ability to control auditory imagery (**b**, *y*-axis). Each plot includes the least-squares regression line. Each element on a scatterplot represents a single participant
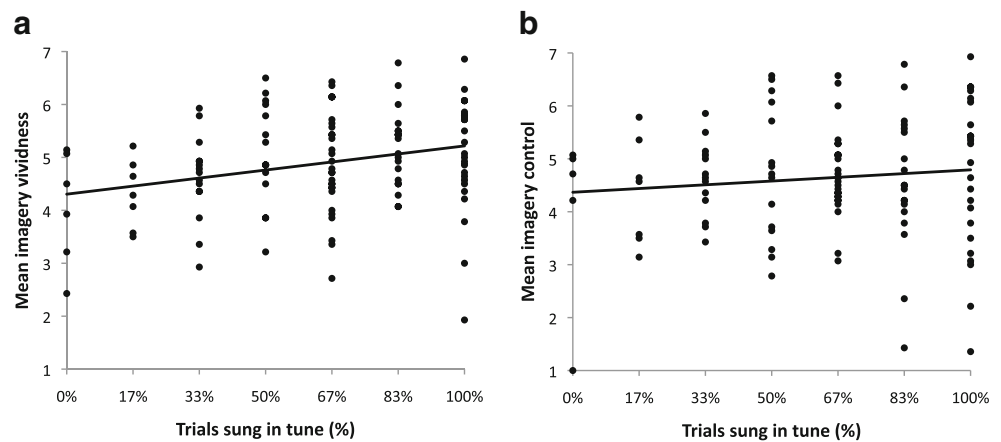
of trials in tune yielded a marginally significant effect, *p* = .07), bearing out the substantial collinearity of these predictors. At the same time, the standardized coefficient was considerably larger for the proportion of trials in tune (*ß* = .29) than for absolute pitch error (*ß* = .02). On the basis of its larger effect size, we decided to use proportion of trials in tune for all subsequent analyses.

The analyses reported above averaged across all six imitation trials for measures of production, in order to enhance statistical power. However, performance within each condition (pitch height of the modeled pitch) is of interest, as well as whether correlations between imitation and imagery vary across conditions. Figure 3a shows the percentage of trials sung in tune, aggregated across participants, for each trial. As can be seen, performance declined considerably for trials 4 and 5, which were both lower than the comfort pitch (see secondary *x*-axis). Differences across conditions were significant according to a chi-square test, $\chi(5) = 142.61$, *p* < .01. Of greater interest is whether the relationship between imitation accuracy and imagery self-reports varied across positions. Point-biserial correlation coefficients for

these relationships are shown in Fig. 3b for both BAIS subscales. Correlations with vividness were significant across all trials; by contrast, correlations with ratings of control were significant only for trial 1.

Pitch discrimination

The focus of the present research is on the role of auditory imagery in production. However, it is also possible that individual differences in pitch discrimination ability (which typically does not lead to poor pitch singing, as was discussed earlier) are correlated with individual differences in auditory imagery (although reliable activation of the primary auditory cortex during auditory imagery has typically not been found; Zatorre & Halpern, 2005). As such, we assessed whether overall performance on the pitch discrimination task correlated with responses on each BAIS subscale. First, we correlated BAIS scores with mean pitch discrimination accuracy across all conditions. Figure 4a and b plot relationships of each subscale of the BAIS (as in Fig. 2) with the percentage of trials on

Fig. 3 **a** Percentages of trials sung in tune (across participants) as a function of trial number (*upper panel*). The *lower panel* shows an example of the pitches a participant may be asked to imitate on each trial, given a comfort pitch of G3 (see also secondary *x*-axes from both upper panels). **b** Point-biserial correlations between accuracy within a single trial (dichotomous *x* variable) and mean response on each subscale of the BAIS (*y* variable). This plot shows correlation coefficients (*y*-axis of plot) separately for each imitation trial (*x*-axis of plot) and subscale
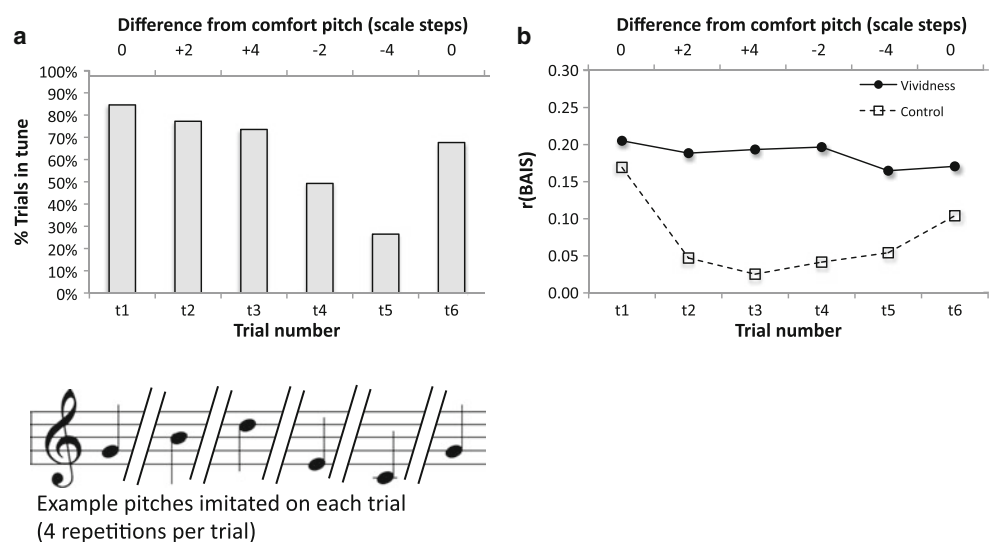
Fig. 4 Scatterplots displaying the relationship between the mean percentages of trials on which participants correctly discriminated pitch changes (*x*-axis), and mean responses on the BAIS questionnaire concerning vividness of auditory imagery (**a**, *y*-axis) or ability to control auditory imagery (**b**, *y*-axis). Each plot includes the least-squares regression line. Each element on a scatterplot represents a single participant
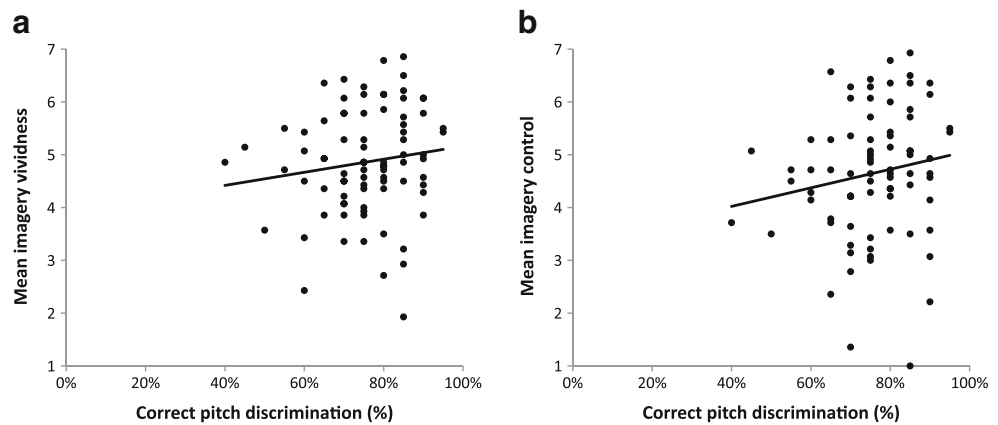
which each participant correctly discriminated the direction of a pitch change. The relationship between mean accuracy in perceptual discrimination and the vividness of auditory imagery was only marginally significant, $r(118) = .13$, $p = .07$, but did reach significance for the correlation between pitch discrimination and control of auditory imagery, $r(112) = .14$, $p < .05$. A further multiple correlation analysis was used to determine whether imitation accuracy predicted ratings of vividness independently of perceptual discrimination ability. The regression of vividness on the predictors percentage of trials sung in tune and mean accuracy of discrimination was significant ($p < .01$), but a partial correlation analysis showed a significant correlation only for the percentage of trials sung in tune ($p < .01$), whereas the partial correlation between pitch discrimination and vividness was not significant ($p = .30$).

As we did in our analyses of production, we also analyzed pitch discrimination performance, and correlations with BAIS, across different experimental conditions. Figure 5a shows mean accuracy on pitch discrimination as a function of pitch separation (note that the 0-change condition constitutes "catch trials") and direction of the pitch change. A within-subjects analysis of variance yielded a significant main

effect of pitch separation, $F(4, 516) = 114.20$, $p < .01$, a main effect of pitch direction, $F(1, 129) = 32.435$, $p < .01$, and a significant separation × direction interaction, $F(4, 516) = 4.67$, $p < .01$. The main effect of separation reflects the fact that performance increased reliably with increasing pitch separation, was at chance for the no-change condition ($M = .49$, $SE = .02$), and was highly accurate, although not at ceiling, for the 100-cent change ($M = .94$, $SE = .01$). The main effect of direction reflects a slight bias to report descending pitches (apparent in the no-change condition) that was enhanced for reasons we cannot explain in the 25-cent change condition, leading to the interaction.

We next consider whether the relationship between perception and imagery self-reports varies with the magnitude of the pitch change (the most critical factor in the pitch discrimination task). Pearson correlation coefficients as a function of pitch change condition and the two subscales of the BAIS are shown in Fig. 5b. Note that each pitch change condition for a single participant yields four data points (two ascending and two descending conditions). As can be seen, correlations were stronger for the larger pitch change conditions and were maximal for both BAIS subscales in the 50-cent change condition, possibly reflecting the moderate difficulty level of this condition. Correlations
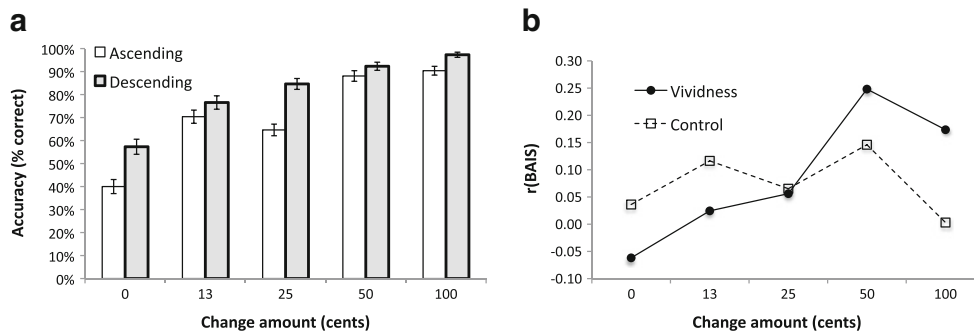

Fig. 5 a Percentages of correct discrimination as a function of the magnitude of pitch change and direction of pitch change (for the no-change condition, the *y*-axis simply reflects percentage of responses that are either ascending/descending). *Error bars* represent 1 *SE*. **b**

Pearson correlations between accuracy within a single change amount condition (*x* variable) and mean response on each subscale of the BAIS (*y* variable). This plot shows correlation coefficients (*y*-axis of plot) separately for each change amount (*x*-axis of plot) and subscale

with the vividness subscale were significant for the 50- and 100-cent change conditions, whereas no correlations with the control subscale were significant.

Third variables

Because we included both musicians and nonmusicians in our sample, we next considered the degree to which musical experience accounts for BAIS responses, as well as vocal imitation accuracy. Not surprisingly, total years of musical experience correlated significantly with the proportion of trials sung in tune, $r(118) = .29$, $p < .01$, and vividness ratings, $r(118) = .30$, $p < .01$. We were interested in whether BAIS responses predict production accuracy independently of musical experience. A multiple regression analysis of vividness ratings with predictor variables of percentage of trials sung in tune and years of musical experience was significant ($p < .01$), and a partial correlation analysis showed that each predictor accounted for independent portions of the variance ($p < .05$ for proportion of trials in tune, $p < .01$ for years of musical training). Another potential source of variability in the data has to do with linguistic background, given recent evidence suggesting some advantages for tone language speakers in vocal pitch imitation tasks (Pfordresher & Brown, 2009). However, that study found no advantage for tone language speakers in the imitation of monotone sequences like those used here. Likewise, analyses of imitation accuracy across tone and nontone language speakers here yielded no differences, nor did comparisons of groups for BAIS subscales and the perceptual discrimination task (note that evidence for a tone language advantage for simple pitch discrimination is mixed; Bidelman, Gandour, & Krishnan, 2011).

Discussion

We report evidence of an association between self-reports of the vividness of auditory imagery (measured using the BAIS) and the accuracy with which participants could imitate pitch through singing. This association was independent of musical experience, height of the imitated pitch, and pitch discrimination ability. These results support our hypothesis that vocal imitation relies on auditory imagery. As was described earlier, neuroimaging evidence suggests that auditory images prime motor planning areas and, thus, provide a mechanism for the inverse modeling of perception and action relationships that may be used in contexts like vocal imitation. Our data suggest that poor-pitch singers may fail to generate the kind of (vivid) auditory images that can be used to guide motor planning (Herholz et al., 2012; Zatorre et al., 2010). Identifying this construct as mediating the perception–action link both supports and extends the

inverse modeling approach to sensorimotor behavior. It is particularly interesting in this context because of the covert nature of singing (motor feedback is hard to observe). Auditory imagery might underlie other kinds of covert behavior, such as acquiring a foreign accent.

Although similar trends were seen in correlations with both the vividness and control subscales of the BAIS, only the vividness scale yielded significant relationships. This dissociation may reflect the kind of deficit present among poor-pitch singers. Whereas the vividness subscale concerns one's ability to generate an auditory image, the control subscale involves both the generation and manipulation of an auditory image. The fact that vividness ratings yielded the strongest correlations with vocal imitation ability suggests that poor-pitch singing involves an imagery deficiency at a very basic level.

We note that although the auditory imagery scores were generated by self-report, our confidence in the validity of this measure is enhanced by the precision with which only one of the subscales correlated with only the one hypothesized auditory–vocal ability. Combined with the prior evidence that this scale predicts neural activity on a voxel-by-voxel basis (Herholz et al., 2012; Zatorre et al., 2010), we suggest that participants can make self-judgments of the internal trait of imagery activity in a way wholly adequate for systematic investigation.

References

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

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition, 77,* 1–10.

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

Brown, R. M., & Palmer, C. (2012). Auditory-motor learning influences auditory memory for music. *Memory & Cognition, 40,* 567–578.

Clark, T., & Williamson, A. (2011). Imagining the music: Methods for assessing musical imagery ability. *Psychology of Music, 40,* 471–493.

Dalla Bella, S., Giguere, J. F., & Peretz, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America, 121,* 1182–1189.

Halpern, A. R. (2001). Cerebral substrates of musical imagery. *Annals of the New York Academy of Sciences, 930,* 179.

Herholz, S., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Article Accepted Pending Minor Revisions, 24,* 1382–1397.

Hickok, G., Houde, J. F., & Rong, F. (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron, 69,* 407–422.

Highben, Z., & Palmer, C. (2004). Effects of auditory and motor mental practice in memorized piano performance. *Bulletin of the Council for Research in Music Education, 159,* 58–65.

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.

Keller, P. E. (2012). Mental imagery in music performance: Underlying mechanisms and potential benefits. *Annals of the New York Academy of Sciences, 1252,* 206–213.

Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. (2009). Brain activation during anticipation of sound sequences. *The Journal of Neuroscience, 29,* 2477–2485.

Pfordresher, P. Q. (2011). Poor pitch singing as an inverse model deficit: Imitation and estimation. In A. Williamon, D. Edwards, & L. Bartel (Eds.), *Proceedings of the International Symposium on Performance Science (pp. 539–544)*. Utrecht: Association Européenne des Conservatories.

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.

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

Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin, 136,* 943–974.

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

Zatorre, R. J., & Halpern, A. R. (2005). Mental Concerts: Musical Imagery and Auditory Cortex. *Neuron, 47,* 9–12.

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,* 775–789.