View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Bucknell University

Bucknell University

Bucknell Digital Commons

Faculty Journal Articles

Faculty Scholarship

2017

Pitch Imitation Ability in Mental Transformations of Melodies

Emma B. Greenspon

Peter Q. Pfordresher

Andrea R. Halpern
Bucknell University, ahalpern@bucknell.edu

Follow this and additional works at: https://digitalcommons.bucknell.edu/fac_journ

Part of the Cognitive Psychology Commons, and the Music Commons

Recommended Citation

Greenspon, Emma B.; Pfordresher, Peter Q.; and Halpern, Andrea R.. "Pitch Imitation Ability in Mental Transformations of Melodies." *Music Perception* (2017) : 585-604.

This Article is brought to you for free and open access by the Faculty Scholarship at Bucknell Digital Commons. It has been accepted for inclusion in Faculty Journal Articles by an authorized administrator of Bucknell Digital Commons. For more information, please contact dcadmin@bucknell.edu.

Pitch Imitation Ability in Mental Transformations of Melodies

Emma B. Greenspon & Peter Q. Pfordresher
University at Buffalo, State University of New York

Andrea R. Halpern
Bucknell University

Previous research suggests that individuals with a Vocal Pitch Imitation Deficit (VPID, a.k.a. "poor-pitch singers") experience less vivid auditory images than accurate imitators (Pfordresher & Halpern, 2013), based on self-report. In the present research we sought to test this proposal directly by having accurate and VPID imitators produce or recognize short melodies based on their original form (untransformed), or after mentally transforming the auditory image of the melody. For the production task, group differences were largest during the untransformed imitation task. Importantly, producing mental transformations of the auditory image degraded performance for all participants, but were relatively more disruptive to accurate than to VPID imitators. These findings suggest that VPID is due partly to poor initial imagery formation, as manifested by production of untransformed melodies. By contrast, producing a transformed mental image may rely on working memory ability, which is more equally matched across participants. This interpretation was further supported by correlations with self-reports of auditory imagery and measures of working memory.

Received: December 29, 2015, accepted January 25, 2017.

Key words: singing, sensorimotor integration, imagery, working memory, pitch perception

Singing is a ubiquitous form of musical communication, and is an early developing and universal form of music performance. At the same time, many (possibly a majority of) adults in Western cultures claim to be unable to sing, and for these people the primary difficulty is in matching pitch (Pfordresher & Brown, 2007; cf. Cuddy, Balkwill, Peretz, & Holden, 2005). Although most of these individuals may be underestimating their ability, nevertheless, a sizable minority (possibly 19%) misproduce pitch by more than a semitone on average (Pfordresher & Larrouy-Maestri, 2015, but see also Berkowska & Dalla-Bella, 2013; Hutchins & Peretz, 2012). The present research is part of an attempt to understand why certain people exhibit this vocal pitch imitation deficit (VPID), and the mechanisms that contribute to vocal imitation.

The present research follows from recent evidence suggesting that VPID may be linked to a deficit in auditory imagery. Auditory imagery has been linked to multimodal associations in the brain that lead to multimodal mental images. These multimodal images may form the basis for shared representations that underlie perception/action associations (Hommel, 2009; Hommel, Müsseler, Aschersleben, & Prinz 2001). Therefore, *auditory imagery* may guide the process of *sensorimotor translation* that is critical for vocal imitation: the ability to convert perceived pitches from a melody into a motor plan. Previous research showed that VPID individuals report having less vivid auditory images than accurate imitators (Pfordresher & Halpern, 2013). If so, these impoverished images may fail to yield associations with motor planning necessary for sensorimotor translation.

We report an experiment that builds on this previous correlational finding by having participants form and also manipulate auditory images. Participants (screened to represent accurate and VPID groups) produced (sang) melodies or mental transformations of melodies, and in other trials attempted to recognize melodies or their transformations. If imitators differ with respect to underlying imagery abilities, we reasoned that effects of transformation on performance should differ across groups. Furthermore, differences in production and recognition tasks can shed light on whether VPID deficits are either limited to tasks that require multimodal associations or are also found in tasks that can rely simply on auditory imagery.

What is a Vocal Pitch Imitation Deficit (VPID)?

VPID is a deficit of pitch matching most commonly associated with singing but that also relates to the imitation of spoken prosody (Mantell & Pfordresher, 2013; Wisniewski, Mantell, & Pfordresher, 2013), which is why we adopt the term VPID rather than "poor-pitch singing" (Welch, 1979). VPID singers tend to produce erroneous pitches that drift toward the direction of their

Music Perception, volume 34, issue 5, pp. 585–604, issn 0730-7829, electronic issn 1533-8312. © 2017 by the regents of the university of california all rights reserved. please direct all requests for permission to photocopy or reproduce article content through the university of california press's Reprints and Permissions web page, http://www.ucpress.edu/journals.php?p=reprints. DOI: https://doi.org/10.1525/mp.2017.34.5.585

comfort pitch (an estimate of the center of one's vocal range) during vocal pitch matching (Pfordresher & Brown, 2007). Following Pfordresher and Brown (2007), we operationally define VPID based on a consistent tendency to sing more than a semitone off, either sharp or flat, when attempting to match pitch via imitation. This tendency is found in a minority of the population (as noted above) and is correlated with several other deficiencies of vocal pitch imitation. When imitating pitch sequences, VPID singers also tend to compress the size of pitch intervals (Dalla Bella, Giguére, & Peretz, 2009; Pfordresher & Brown, 2007; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010) and perseverate on pitch patterns that were produced on previous trials (Wisniewski et al., 2013). Second, VPID singers exhibit both a tendency to shift pitch in a consistent direction (thus exhibiting a response bias, or "inaccuracy" in the statistical sense) and also inconsistency in repeated imitations of a pitch class ("imprecision"). Third, VPID singers are also inconsistent (a.k.a. imprecise), in that they will produce a given pitch differently across repeated attempts[1] (Pfordresher et al., 2010). In sum, VPID singing is characterized by both biased and inconsistent production of pitch.

A simple account of VPID can emerge from deficits in either auditory perception or motor control of pitch, as these processes are critical to pitch imitation and complex in their own right (for reviews see Sundberg, 1987; Zarate, 2013). However, for the most part, VPID singers do not exhibit deficits in pitch discrimination or melodic processing (Dalla Bella, Giguére, & Peretz, 2007; Hutchins & Peretz, 2012; Pfordresher & Brown, 2007). Similarly, limitations of pitch range and motor control found during imitation tasks are not necessarily observed when VPID individuals engage in spontaneous vocalization (Pfordresher & Brown, 2007). Finally, although VPID could reflect poor pitch memory, deficits still appear in tasks that involve minimal memory demands, such as matching a single pitch, and are in fact more pronounced in these tasks than in imitation of complex multi-pitch melodies (Pfordresher & Brown, 2007).

Thus, recent research on VPID has focused primarily on the role of sensorimotor translation in this deficit. We have recently proposed a model, termed the Multi-Modal Imagery Association (MMIA) model (Pfordresher, Halpern, & Greenspon, 2015), which suggests that perceptual imagery (here auditory pitch events) is mapped onto motor imagery (here a plan for the control of phonation) through probabilistic sensorimotor associations derived from a lifetime of imitating vocal sounds. Consistent with this theory, past evidence suggests that VPID singers are particularly sensitive to the familiarity of the timbre of the target that they imitate (Hutchins & Peretz, 2012). Critically, whereas all individuals are more successful at imitating recordings of themselves than recordings of another singer, this advantage is enhanced for VPID singers (Moore, Estis, Gordon-Hickey & Watts, 2008; Pfordresher & Mantell, 2014).

Mental Imagery and Its Role in Sensorimotor Translation

Research on both visual and auditory imagery has focused on the distinction between the vividness and the ease with which one can control (manipulate) a mental image (Halpern, 2015; Lequerica, Rapport, Axelrod, Telmet, & Whitman, 2002). Whereas vividness relates to the clarity of the mental image, control relates to how easily one can alter an existing mental image (Lequerica et al., 2002). Although these are two distinct aspects of imagery, these processes are typically correlated (Marks, 1999). Thus, one cannot successfully manipulate (control) a mental image unless a vivid image of the percept has already been formed. Interestingly, self-reported vividness predicts performance on tasks that require veridical mental representations, such as imitating monotone melodies, whereas self-reported control predicts performance on tasks that require maintenance and manipulation of mental representations, such as mentally moving up and down a musical scale (Halpern, 2015).

We propose that a critical component of sensorimotor translation is the formation of a multi-modal mental image (cf. McNorgan, 2012). This assertion is based in part on an earlier study in which VPID participants tended to report less vivid auditory imagery than accurate imitators (Pfordresher & Halpern, 2013), whereas self-reports of imagery control did not predict singing accuracy. This representation links to motor planning; thus the multi-modal image constitutes a mental representation that facilitates associations between perceptual and motor systems.

Evidence for a link between imagery and sensorimotor translation has been established in behavioral and neuroimaging research. Smith and colleagues (Reisberg, Smith, Baxter, & Sonenshine, 1989; Smith, Wilson, & Reisberg, 1995) found that suppressing subvocalization interfered with the formation of auditory images, suggesting that motor planning and production influence auditory imagery processes. Additionally, fMRI research

[1] It should be noted that this imprecision is also fairly frequent among singers who are not inaccurate on average and thus not defined as VPID according to our criterion.

has shown that auditory imagery activates some auditory and motor areas of the brain (Hubbard, 2010; Zatorre & Halpern, 2005). Most important, these areas are associated with higher-order aspects of perception and action; namely, the secondary (rather than primary) auditory cortex and supplementary (rather than primary) motor area. In particular, the supplementary motor area is thought to be involved in motor planning (Zatorre, Chen, & Penhune, 2007) and activation of this area may reflect sensorimotor priming (Bangert et al., 2006; Engel et al., 2012). Furthermore, activity in these areas has been elicited during diverse auditory imagery tasks such as imagining musical timbre (Halpern, Zatorre, Bouffard, & Johnson, 2004), imagining familiar melodies (Halpern & Zatorre, 1999), and manipulating auditory images of familiar melodies (Zatorre, Halpern, & Bouffard, 2010). The role of mental imagery suggested by these findings is one in which the formation of an auditory image primes motor plans associated with the production of the event being imagined. This is in line with William James' ideomotor theory, which states that imagining a particular outcome automatically initiates the corresponding action (James, 1890; Shin, Proctor & Capaldi, 2010; cf. Phillips-Silver & Keller, 2012).

According to the MMIA model (Pfordresher, Halpern, & Greenspon, 2015), past action-perception associations between laryngeal movements and specific vocal outcomes are used to estimate the motor plan needed to produce a desired perceptual event (e.g., a sung pitch). Statistical learning over an individual's lifetime allows these established associations to be extended to form an abstract schema. The schematic mapping of multi-modal images allows individuals to generalize these associations to novel situations. Distortions in this mapping come from two sources that are motivated by empirical findings: Noise (imprecision, cf. Pfordrdesher et al., 2010), and response bias (the tendency to favor one's own "comfort pitch," cf. Pfordresher & Brown, 2007). Further model details are described in Pfordresher, Halpern, and Greenspon (2015) and Appendix B.

The Present Study

In the current study we were interested in measuring imagery quality and control in accurate and VPID singers. We asked participants to carry out auditory imagery tasks, in order to test the relationship between VPID and mental imagery that had been supported via correlation by Pfordresher and Halpern (2013). In the present study, participants were asked to sing and recognize mental transformations of melodies; a design inspired by object rotation tasks in visual imagery (Shepard & Metzler, 1971, see Figure 1).

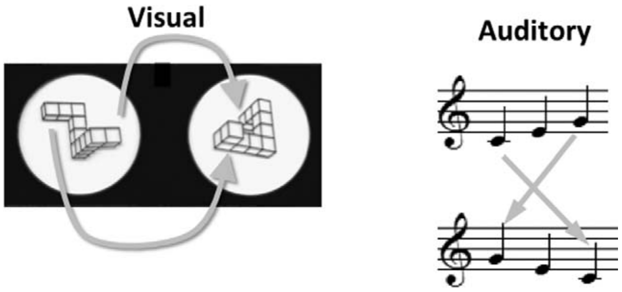

FIGURE 1. Example of a stimulus pair used in Shepard and Metzler's (1971) study of mental rotation for visual imagery (left). Example of a melody transformation from the current study involving reversal of pitch order (right).

In Shepard and Metzler's (1971) mental rotation task, participants recognized three-dimensional objects that could be rotated at angles ranging from 0° to 180° around a vertical axis. Because this task places considerable demand on both maintaining and manipulating a mental image, over and above working memory as captured in traditional verbal tasks, it stands as a classic paradigm to study mental imagery. An auditory analogue to this task was developed by Zatorre and colleagues (2010), who had participants recognize pitch-order reversals of familiar melodies. Zatorre and colleagues (2010) found that participants responded to incorrect reversals more quickly if the altered pitch was positioned earlier rather than later in the reversed melody, implicating a coherent serial process of imagery transformation. Furthermore, the accuracy with which participants recognized correct versus altered reversals scaled positively with self-reports of auditory imagery from the Bucknell Auditory Imagery Scale (BAIS; Halpern, 2015).

The current study is one of the first to measure how well people can produce (i.e., sing) mental transformations of melodies. Other studies have incorporated somewhat comparable tasks, such as the recognition of transformed melodies (Zatorre et al., 2010), sung production based on notated transformations (Zurbriggen, Fontenont, & Meyer, 2006), sung repetitions and reversals of atonal tone sequences (Benassi-Werke, Quieroz, Araujo, Bueno, & Oliveira, 2012), and probe tone ratings based on imagined transformations (Vuvan & Schmuckler, 2011). Yet none of these previous paradigms address the ability to produce melodies based strictly on a transformed auditory image of that melody. We expanded on Pfordresher and Halpern (2013) and

Zatorre and colleagues' (2010) imagery measures in order to assess whether additional imagery demands (i.e., mental transformations) lead to differing effects on performance for VPID versus accurate imitators. Furthermore, while the production of mental transformations relies on multimodal imagery, recognition of mental transformations is thought to strongly rely on unimodal auditory imagery. Therefore, we included a recognition task to address how imagery demands may differentially influence production and recognition performance for these two groups. Participants, preselected to represent accurate or VPID imitation ability, produced melodies either as an exact repetition or based on a mental transformation. Participants also completed recognition tasks based on exact repetitions or transformations of the target melody.

We included three types of transformations: transpositions, reversals, and serial order shifts. Transpositions and reversals were inspired by previous studies (Foster & Zatorre, 2010; Foster, Halpern, & Zatorre, 2013; Zatorre et al., 2010). Transpositions offered a transformation that preserved relative pitch, but not absolute pitch. Reversals created a transformation that contained all of the same absolute pitches of the original melody, but the pitches' temporal positions were altered. Finally, serial order shifts, similar to reversals, contained all of the same absolute pitches of the original melody but shifted the position of the starting and ending tones.

We consider performance in the untransformed condition to reflect the quality or vividness of one's mental image of the target (i.e., how closely an image matches sensory experience). Performance on the transformation trials, on the other hand, reflects imagery control. Poor image quality and poor image control can both lead to inaccurate transformations.

We used correlational analyses as a test of construct validity for the processes contributing to both tasks. The BAIS was used to address the contribution of imagery vividness and image control via its two subscales. These were predicted to correlate with all tasks. In addition, transformation tasks require working memory. Following Christiner and Reiterer (2013), we used forward and backward digit span tasks to measure short term and working memory capacity, respectively. These authors found that performance on both span tasks correlated with accuracy on imitation of song and speech, such that more accurate imitators exhibited larger memory capacity.

As this is the first study to compare performance across different mental transformations of melodic sequences, it was largely exploratory with respect to the relative difficulty of each type of transformation, although we expected the transposition to be the easiest transformation, given the ubiquity of key changes in musical practice and the preservation of contour in this transformation condition.

Method

As mentioned before, the participant categories were derived from a screening procedure. We therefore discuss separately the aspects of the method relating to the initial screening procedure, and those relating to the experiment.

PARTICIPANTS

Screening. An initial group of 233 participants were screened to select VPID and accurate imitators. The majority of the screened sample (90%) was recruited from an introductory psychology course at the University at Buffalo in exchange for course credit. The other 23 participants were recruited through campus flyers and were paid $10/hour for participation in the experiment. All participants reported normal hearing and vocal production.

The screening process followed two stages in order to focus on participants who exhibit VPID without a concomitant pitch perception deficit. In the first screening stage, 40 participants (17%) were selected out of the original 233 participants based on performance in a vocal pitch-matching task (see Procedure). Accurate participants were those who correctly matched five or six out of six target sequences, whereas VPID participants correctly matched no more than one sequence.

In the second stage of screening, we removed any VPID participants who exhibited a potential deficit of pitch perception. We retained nine VPID singers whose performance was above 70 percent correct on a pitch discrimination task (described below) and one participant who did not complete the pitch discrimination task but whose available data in the recognition task led us to believe that no perceptual deficit was present.[2]

In order to preserve the same sample sizes across groups, we then excluded the 10 accurate singers who scored the lowest on the pitch discrimination task (all scored above 70 percent correct). This left a final sample

[2] The rationale for including this participant was based on constraints associated with the infrequency of VPID. Although this participant did not complete the pitch discrimination task in screening, this participant's performance in the untransformed condition of the recognition task from the main experiment (18% errors) fell within the range of error rates of accurate singers ($M = .09\%$ errors, range: 0-34) and thus suggested intact pitch perception.

of 10 accurate singers and 10 VPID singers who did not differ significantly in pitch discrimination accuracy.[3]

Experiment. In the final sample, accurate singers (eight male, two female) had a mean age of 20.4 years (range = 18-29 years) and an average of 3.1 years of music training (range: 0-11 years). VPID individuals (six male, four female) had a mean age of 21.2 years (range: 18-30 years) and an average of 3 years of music training (range: 0-8 years). Ten accurate and seven VPID singers from the final sample were recruited from an introductory psychology course; three VPID singers were recruited through campus flyers as described above. Two of the three VPID participants recruited through flyers were university students; the third was a professional engineer.

APPARATUS

Participants completed all vocal production trials, for the screening task and the experiment, in a sound-attenuated recording booth (Whisper Room SE 2000). Stimuli were presented through Sennheiser HD 280 Pro headphones. Participants were recorded using a Shure PG58 microphone while sound levels were adjusted using a Lexicon Omega I/O box. Recognition trials were completed on a 3.4 Ghz PC running Windows XP and using the same Sennheiser headphones. The screening and experiment were run using Matlab (MathWorks, Natick, MA).

STIMULI

Screening. Stimuli were voice synthesized sung sequences generated using the software package Vocaloid: Leon (Zero-G Limited, Okehampton, UK). All notes were produced on the syllable "dah." The male stimuli reflected a normal male singing range. The female stimuli were shifted an octave higher and the formants were altered to model female vocal timbre.

For the screening, participants completed six imitation trials. Each trial consisted of one pitch repeated four times. All pitches were from the C major scale and there were no pauses between pitches. The pitch used for the first trial was each participant's self-selected comfort pitch (see Procedure). The second and third trials were pitches two and four semitones above the participant's comfort pitch, respectively. The fourth and fifth trials were two and four semitones below the participant's comfort pitch, respectively. The sixth trial was the same as the first trial.

Experiment. The experiment included both vocal imitation tasks and recognition tasks that involved mental transformations.

Production stimuli. Stimuli in the experiment were produced on the same syllable and with the same timbre as the stimuli described in the screening. Participants completed a set of practice trials prior to each experimental condition. Practice stimuli were melodies not used in the actual experiment, but had the same features as the experimental stimuli that are described below. Practice stimuli included short sequences ranging from 2-4 notes.

The experimental stimuli consisted of 3 or 4-note target sequences. Pitches for each sequence were selected from the C-major scale: C3 for male participants (B2, C3, D3, E3, F3, G3, A3) and C4 for female participants (B3, C4, D4, E4, F4, G4, A4). There were eight sequences for each note length resulting in 16 target sequences in all. Each sequence included no repeated pitches and was matched for contour complexity with all other sequences (number of changes in pitch direction). All sequences started on a C or G and ended on the tonic, dominant, mediant, or subdominant; i.e., C, G, E, or F respectively. Note durations were 1 s with no pause between notes in the target sequence.

Each target sequence was followed by a cue note that was designed to facilitate performance on the various tasks. Each cue note was the correct starting pitch for the respective condition: *untransformed*, *reverse*, *serial order shift,* and *transposition*. In the untransformed condition, participants repeated the melody as they heard it; thus the cue note matched the starting note of the melody. For reverse transformation trials, participants produced the melody in reverse order of pitches; thus the cue note matched the final note of the original melody. For serial order shifts, participants started the melody at the second to last note and cycled around to end on the first note for three note melodies and second note for four note melodies; the cue note thus matched the second to last note of the melody. Finally, for the transposition condition, participants sang the melody in a new key and the starting note was thus a transposition of the first note of the melody (always E$^\flat$). We chose E$^\flat$ as the starting note in order to cue a distant key and help diminish the likelihood that participants would sing a tonal transposition: shifting the notes while remaining in the original key (Bartlett & Dowling, 1980). For melodies beginning on C, the transposed key becomes E$^\flat$ (same as the cue

[3] When including participants from the first stage of screening (20 accurate and 20 VPID singers), we found that VPID singers exhibited significantly lower pitch discrimination scores (M = 72% correct) than accurate singers (M = 81% correct), $t(38) = 2.57$, $p < .05$. An alternate procedure to the one we followed is to remove variability associated with pitch discrimination via detrending from this first-stage sample. After detrending, mean data by group and condition differed only negligibly from the unadjusted scores and from the final sample (10 accurate and 10 VPID): all significant effects were preserved and the pattern of results was nearly identical.

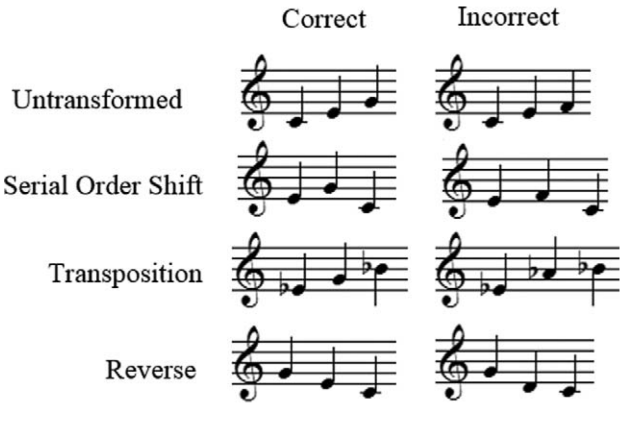

FIGURE 2. Left) Notations of correct production responses for each condition. Right) Notations of incorrect recognition melodies for each condition.

note); for melodies that began on G, the starting note of E♭ cued a transposition to the key of A♭ (in which E♭ is the dominant). See the left panel of Figure 2 for a notated example of the four conditions.

Recognition stimuli. The core set of melodies for recognition trials was identical to production trials. During recognition trials, one of the original melodies was followed by an exact repetition or a transformation, thus forming a standard/comparison pair. On half the trials one pitch in the comparison deviated from an accurate repetition or transformation, and participants had to detect these changes.

Recognition trials were preceded by practice trials that consisted of 3 and 4-note melodies modeled after the experimental stimuli. There were four practice trials in total, one for each condition. Half of the trials comprised the target followed by an exact transformation or repetition whereas half comprised the target followed by an incorrect transformation or repetition. Participants received feedback on whether their response was correct or incorrect for practice trials, but not for experimental trials.

The experimental trials consisted of a target sequence followed by a pause of 2.5 s and then a correct or incorrect repetition or transformation of the target (the comparison melody). Incorrect comparison melodies contained a single altered pitch that was never the first pitch in the sequence. The majority of the melodies (56%) had a pitch that was 2 semitones higher or lower than the original pitch, 30% had pitches that were altered to the nearest semitone and less than 15% had pitches that were altered 3, 4, or 5 semitones from the original pitch. Altered pitches were chosen based on the following constraints: altered pitches were different from the other pitches in the sequence, were tonal, and did not change the contour of the sequence. See the right panel of Figure 2 for the notation of an incorrect melody for all 4 conditions.

The transformations of the target were either reversals, shifts in serial order or transpositions, as in the production task. For the reversal condition, participants heard the target melody in reversed order. The serial order shift condition played the melody starting from the second to last note and then the melody cycled back to the beginning. In the transposition condition, participants heard the melody transposed to the key of "E♭" for sequences that began on "C", and transposed to the key of "A♭" for sequences that began on "G."

PROCEDURE

Screening. We adopted a protocol from Pfordresher and Brown (2007, Experiment 2) to screen for accurate and VPID singers. Experimenters instructed participants on the use of appropriate posture and breathing during singing trials. Participants first completed a vocal warm-up series that consisted of singing "Happy Birthday" and reading "The Rainbow Passage" (Fairbanks, 1960) out loud. Participants also completed vocal sweeps: continuous changes in pitch from the lowest note in their vocal range to the highest note in their vocal range. Finally, participants were asked to provide a single pitch approximately in the middle of their vocal range that they felt comfortable singing and this was labeled as their *comfort pitch*.

Following the warm-up series, all participants completed six imitation trials. Each trial consisted of one pitch repeated four times from the C major scale. Pitches were coded for accuracy after each trial. SHRP, a Matlab pitch-tracking algorithm (Sun, 2002) was used to display the participants' produced F_0 along with boundaries of ± 100 cents around the target F_0 on the computer screen. For each sung note, the experimenter who conducted the screening task coded pitch traces that were outside the designated boundaries as errors whereas pitch traces inside the designated boundaries were coded as accurate. Within a single trial, participants tend to produce repeated pitches consistently; thus each trial was coded as accurate or in error as a whole. Participants with five or six accurate trials were categorized as accurate singers and participants with five or six error trials were categorized as VPID singers.

Participants also completed a pitch discrimination task. For this task, participants heard two pure tones presented sequentially and were asked to tell the experimenter which tone they perceived as being higher in pitch. Each tone was presented for 1 s and there were 2 s of silence between tones. The first tone was fixed at 524 Hz (C5). Pitch differences between the two tones could be 0, 13, 25, 50, or 100 cents, equally distributed between ascending and descending pitch differences. There were 20 trials total: four of the trials had no difference between tones; for the other 16 trials, all possible combinations of ascending and descending tones were presented twice.

Following the pitch discrimination task, participants completed a forward and backward digit span task. For the digit span tasks the experimenter read a sequence of numbers at a spoken rate of approximately one digit per second. Participants then repeated the series they heard either in the original order or in reverse order. There were two different sequences for each length. The task ended when a participant got two sequences of the same length wrong or after they completed all sequences. The maximum sequence length in the forward digit span task was nine digits and the maximum sequence length in the backward digit span task was eight digits. None of the participants correctly repeated the maximum sequence length in either order.

Experiment. Participants completed the experiment within two weeks of the screening (range: 0-14 days, mean = 6.95 days). As in the screening, participants were seated in the recording booth and experimenters instructed participants on appropriate body posture and breathing for singing trials. Participants completed the same warm-up series from the screening inside the recording booth.

Production. Production trials followed the warm-up phase. We blocked production trials by transformation condition, in order to ensure that participants fully understood the task. The first block was the untransformed condition (predicted to be easiest) and the order of the following three blocks was counterbalanced across participants in a Latin square design. Before each condition in the production task, participants completed the practice trials relevant for that block, followed by experimental production trials. There were 6 trials in each condition resulting in 24 trials total. On each trial, participants heard the target melody, followed by a pause, then the cue note, and finally a pink noise burst that served as a cue to start singing. Participants attempted to produce the melody as accurately as possible, given the instructions pertaining to that block.

Recognition. Following the production task, participants exited the recording booth and completed a battery of surveys that included measures of music and language background and the BAIS. After completing the surveys, participants were seated at a computer to complete the recognition trials. As in production trials, participants completed a series of practice trials and were provided feedback on each trial before continuing on to experimental trials. There was one trial for each condition. In the practice trials, participants were instructed to listen for a correct or incorrect repetition or transformation of a target melody. Experimental trials in the recognition task were blocked by condition, as in the production task. Participants were informed about the type of transformation they should be listening for prior to hearing the transformed melody.

We used a 2-alternative forced-choice response design. After the melodies were played over the headphones, participants provided their response by pressing a button on the computer screen. They pressed either a "yes" button for correct transformations/repetitions, or a "no" button for incorrect transformations/repetitions. There were 6 trials in each condition resulting in 24 trials total. After the experiment, the participants were debriefed and were provided either course credit or monetary compensation for their participation.

DATA ANALYSES

In order to measure production and recognition comparably, both were analyzed using error rates.[4] Errors in production were defined as deviations of pitch that were greater than +/- 50 cents surrounding the target pitch (one semitone boundary tolerance in total). The first step in this computation involved extracting the F_0 the participant produced on each sung note from each trial, using the autocorrelation algorithm in Praat (Boersma & Weenink, 2013). The produced F_0 was then converted to cents using C4 (262 Hz) and C3 (131 Hz) as the referent pitches for female and male singers, respectively. The median F_0 from the middle portion of the sung tone was used to represent the sung pitch. This value was compared to the target F_0 (also in cents), and was considered an error if the absolute difference exceeded 50 cents.

In recognition tasks, we measured the proportion of all trials on which participants made an error, which could be a false alarm (incorrectly responding that the

[4] In addition to using error rates in the production task, we also calculated absolute pitch deviations in cents by taking the absolute difference between the produced pitch in cents and the target pitch in cents (see Table A1). We found comparable results to our analyses on error rates when using absolute pitch deviations, though our results were less robust. Furthermore, absolute pitch deviations were strongly correlated with error rates in the production task, $r(18) = .87$, $p <$ .0001. We use error rate as our dependent measure for the production task because this measure complements our error rate measure for the recognition task.

comparison included a changed pitch when it did not) or missing a change.

As a composite measure of the disruptive effect of transformations we looked at how participants performed on the transformation tasks relative to their performance in the untransformed task. We used the untransformed task to serve as a baseline of performance and looked at performance with the transformations as a proportion of change from performance in the untransformed task. This was interpreted as each participant's transformation effect and was calculated using the formula in equation 1.

$$\textit{Transformation Effect} = \frac{avg(T) - B}{B} \qquad (1)$$

In this equation, *avg(T)* represents performance averaged across the transformation trials and *B* represents each participant's baseline of performance in the untransformed condition. For instance, a participant who produced an average of 10% errors across all transformation conditions, but only produced 5% errors in the normal baseline condition, would yield a TE score of (10-5)/5 = 2. We averaged across all three transformation conditions in order to create a comparable analysis to the group x condition interaction from the ANOVA described below. The transformation effect is a useful statistic because it allows us to evaluate individual differences in performance of mental transformations across a continuum of imitation ability. The transformation effect represents the magnitude of differences relative to the standard of performance established by baseline conditions, similar to a Weber fraction. Although this ratio can technically extend from negative to positive infinity, we expected transformation effect values to be restricted to small positive values.

Results

EFFECT OF MENTAL TRANSFORMATIONS ON PRODUCTION PERFORMANCE

We first examined how mental transformations influenced error rates during production. As described above, errors in the production task were defined as pitch deviations greater than 50 cents above or below the target pitch. Transformations were difficult for all participants; however, the effect of transformation was larger for accurate than for VPID singers.

We performed a 2 (group) x 4 (transformation) x 2 (length) mixed-model ANOVA on mean error rates; the only between-subjects factor was group, the rest were varied within subjects, see Figure 3A. Mauchley's test indicated violations of sphericity; all effects reported as significant were also significant after applying the Greenhouse-Geisser correction. There was a significant main effect of group, $F(1, 18) = 35.62$, $p < .001$, $\eta_p^2 = 0.67$, reflecting higher errors for VPID than accurate singers, consistent with the screening procedure. The main effect of transformation was also significant, $F(3, 54) = 23.71$ $p < .001$, $\eta_p^2 = 0.57$. Transformation conditions yielded higher error rates than the untransformed condition. Finally, there was a significant

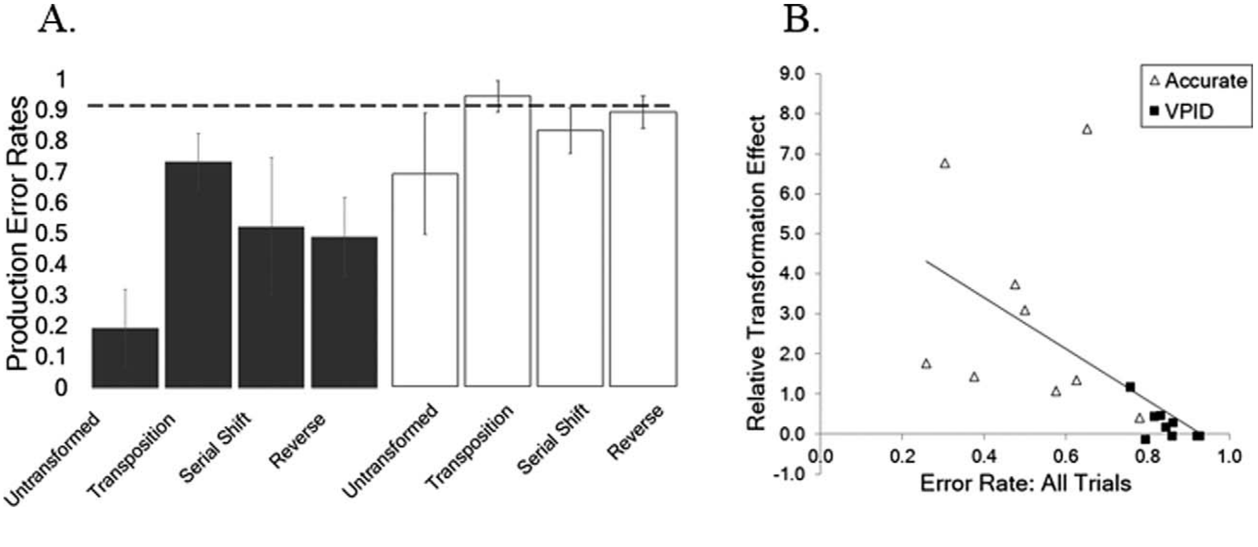

FIGURE 3. A) Accurate singers' (black bars) and VPID singers' (white bars) performance in each condition of the production task plotted with 95% CI error bars. The horizontal dashed line represents chance performance based on 12 pitch chromas. B) Scatterplot showing the relationship between individual performance across all trials in the production task and individual transformation effects in the production task.

group x transformation interaction, $F(3, 54) = 3.50$, $p < .01$, $\eta_p{}^2 = 0.16$. VPID singers had significantly more errors in the untransformed condition of the production task compared to accurate singers, $t(18) = 4.97$, $p < .05$, $d = 2.16$. Although both groups produced more errors when transforming melodies, accurate singers were relatively more disrupted by mental transformation than VPID singers, which led to the significant interaction. Furthermore, this interaction was not a result of VPID singers exhibiting ceiling effects. First, pitch errors were below the ceiling (a 100% error rate) on average, as described below. Second, we found that accurate singers tended to be more disrupted by the transformation conditions than VPID singers when we analyzed errors of melodic contour, which fell well below the chance level.[5] Interestingly, transforming melodies resulted in accurate singers performing more similarly to VPID singers than when singing untransformed melodies, see Figure 3A.

A series of complex planned comparisons were designed to test whether the magnitude of the transformation effect varied across groups. Within each group, the mean across all transformation conditions was contrasted with the mean for the untransformed condition (using coefficients of +1 for each transformed condition and -3 for the untransformed). This transformation contrast was significant both for accurate, $t(36) = 5.07$, $p < .05$, $d = 1.69$, and for VPID, $t(36) = 3.50$, $p < .05$, $d = 1.17$, imitators. More important, an analysis of whether this contrast varied with the Factor of group (Keppel & Wickens, 2004) was significant, $F(3, 54) = 3.21$, $p < .05$.

In order to evaluate differences across the three transformation conditions across both groups (i.e., the main effect of transformation) we ran post hoc comparisons using paired *t*-tests with a Bonferroni correction of 0.017. Transpositions had larger error rates than serial shifts, $t(19) = 4.04$, $p < .02$, $d = 1.10$, and reversals, $t(19) = 4.16$, $p < .02$, $d = 1.02$. Error rates for reversals were not significantly different from error rates for serial shifts, $t(19) = 0.30$, $p = .77$, $d = 0.045$.

Transformation effect across individuals. One concern in the analysis above is that the group means for VPID participants in transformation conditions approach chance levels, which were defined as the probability of producing any one of the 12 pitch chromas in the C major scale. For two conditions, 95% confidence intervals crossed chance levels (see dashed horizontal line in Figure 3A). This brings up the question of whether the smaller transformation effect in VPID participants is a byproduct of performance being compressed by task difficulty.

One way to address this issue is to examine how the transformation effect varies across a continuum of performance, including relatively poor-performing accurate singers as well as VPID participants who are more accurate (cf. Pfordresher & Mantell, 2014, for a similar procedure and explanation). If the results in Figure 3A are a byproduct of participants performing at chance, one should not see a gradual change in the transformation effect with overall performance. This analysis further serves the purposes of addressing the fact that variability within groups may reflect a meaningful continuum of performance (Pfordresher & Larrouy-Maestri, 2015).

Figure 3B shows the relationship between the transformation effect in production trials and overall pitch error rates. Two extreme values (values of more than two standard deviations from the mean) were removed. One outlier was from the accurate group and one from the VPID group. As can be seen, there was a significant negative correlation, such that singers with lower overall pitch error rates in the production task tended to show larger production transformation effects, $r(16) = -.60$, $p < .05$. In particular, note that the transformation effect diminishes for poorer-performing singers in the accurate group whose overall mean error rate was well below chance.

In sum, our main finding from the production task was that whereas both groups were disrupted by transformation trials, accurate singers were more disrupted relative to their performance in the untransformed condition as compared to VPID singers. This finding was also replicated with a correlational analysis: Singers who produced fewer pitch errors showed larger transformation effects than singers who produced more pitch errors.

EFFECT OF MENTAL TRANSFORMATIONS ON RECOGNITION PERFORMANCE

In addition to how well individuals could produce transformations of melodies, we were also interested in how participants performed when asked to recognize but not produce repetitions and transformations of these pitch sequences. As with the production task, VPID participants performed more poorly than accurate singers in

[5] We defined contour errors as the following: If the sung contour differed from the target contour the interval was coded as 1. If the sung contour matched the target contour the interval was coded as 0. Planned comparisons indicated that accurate singers produced fewer contour errors in the untransformed condition compared to the three transformations conditions, $t(36) = 4.57$, $p < .05$, $d = 1.52$. VPID singers showed the same pattern, $t(36) = 2.13$, $p < .05$, $d = 0.71$, however, the effect size was much smaller. The 95% CIs for each condition did not include chance performance (50% error rate) for

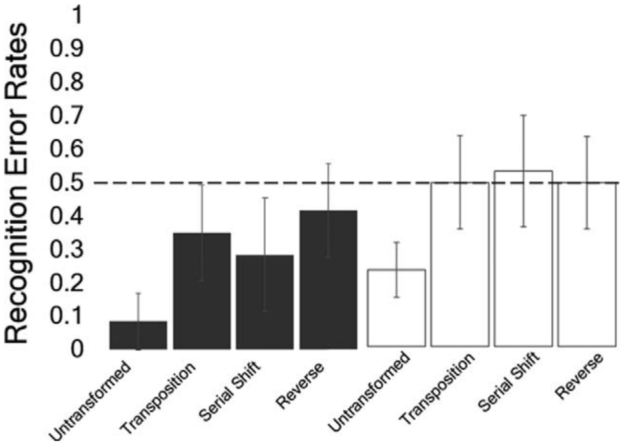

FIGURE 4. Accurate singers' (black bars) and VPID singers' (white bars) performance in each condition of the recognition task plotted with 95% confidence intervals. The horizontal dashed line represents chance performance (0.5).

the recognition task. However, unlike the production task, transformations disrupted recognition to a similar extent for accurate and VPID participants.

Figure 4 shows mean error rates in recognition trials by group and transformation condition. We ran a 2 (group) x 4 (transformation) x 2 (length) mixed ANOVA on these error proportions (group was the sole between-subjects factor). Mauchly's test indicated no violations of sphericity. There was a significant main effect of group, $F(1, 18) = 9.28$, $p < .01$, $\eta_p^2 = 0.16$. VPID singers had worse overall performance on the recognition task than accurate singers. The main effect of transformation was also significant, $F(3, 54) = 9.04$, $p < .001$, $\eta_p^2 = 0.33$, and reflected the same pattern of results found for production errors. Looking specifically at the untransformed condition, VPID singers (Mean VPID = 0.24 errors) performed less accurately (more than double the average error rates) than accurate singers (Mean Accurate = 0.09 errors), however, this difference was not significant; $t(18) = 1.99$, $p = .06$, $d = 0.73$. Comparing this result to the significant group difference in the untransformed condition of the production task indicates that imagery quality differentially affects production and recognition performance. No other effects or interactions were significant, including the critical group x transformation interaction, $F(3, 54) = 0.30$, $p = .80$, $\eta_p^2 = 0.02$.

We next ran complex planned comparisons for each group to evaluate whether error rates in the untransformed condition differed from the three transformation conditions. Accurate singers produced fewer errors in the untransformed condition compared to the three transformation conditions, $t(36) = 3.64$, $p < .05$, $d = 1.21$. VPID singers also followed this pattern of results, $t(36) = 3.38$, $p < .05$, $d = 1.13$. However, unlike the production task, recognizing transformed melodies did not disrupt accurate singers more than VPID singers, $F(3, 54) < 0.01$, $p = .99$. This occurred despite the fact that VPID performance in transformation conditions—as was the case in production—reflected chance levels of performance (see dashed line in Figure 4).

We ran paired *t*-tests with a Bonferroni correction of 0.017 as post hoc comparisons to evaluate whether there were any significant differences across transformation conditions in the recognition task. None of these comparisons were significant.

We also conducted a regression analysis on recognition data, modeled after the correlational analysis used in production. The regression for recognition data was not significant, consistent with the non-significant group x transformation interaction in the ANOVA and the non-significant difference in contrast coefficients.

RELATIONSHIPS OF EXPERIMENTAL TASKS TO PREDICTOR VARIABLES

BAIS vividness and control. We tested the validity of our claim that perception and recognition tasks used here involve auditory imagery. The BAIS, described earlier, has previously been shown to correlate with neural and behavioral measures from a diverse set of imagery tasks (Halpern, 2015) and thus provides a good measure of construct validity for the role of auditory imagery in our experimental tasks. Error rates for untransformed sequences in the production task correlated negatively with self-reports of imagery vividness, $r(18) = -.41$, $p < .05$, and imagery control $r(18) = -.58$, $p < .01$. Error rates for transformed trials (averaged across the three conditions) in the production task were also negatively correlated with imagery vividness, $r(18) = -.42$, $p < .05$, and control, $r(18) = -.55$, $p < .01$. See Figure 5 for all four scatterplots. Somewhat surprisingly, no correlations between BAIS and error rates during recognition were found. BAIS correlations thus suggests that imagery plays a strong role in production but not recognition of auditory sequences. Accurate ($M = 5.2$, $SD = 1.07$) and VPID ($M = 4.6$, $SD = 0.98$) singers did not differ in scores on the vividness subscale, $t(18) = 1.43$, $p = .17$, $d = 0.44$. However, accurate ($M = 5.7$, $SD = 0.79$) singers reported higher scores for the control subscale than VPID ($M = 4.9$, $SD = 0.80$) singers, $t(18) = 2.75$, $p = .01$, $d = 1.08$.

Short-term and working memory capacity. We evaluated the use of short term and working memory capacity

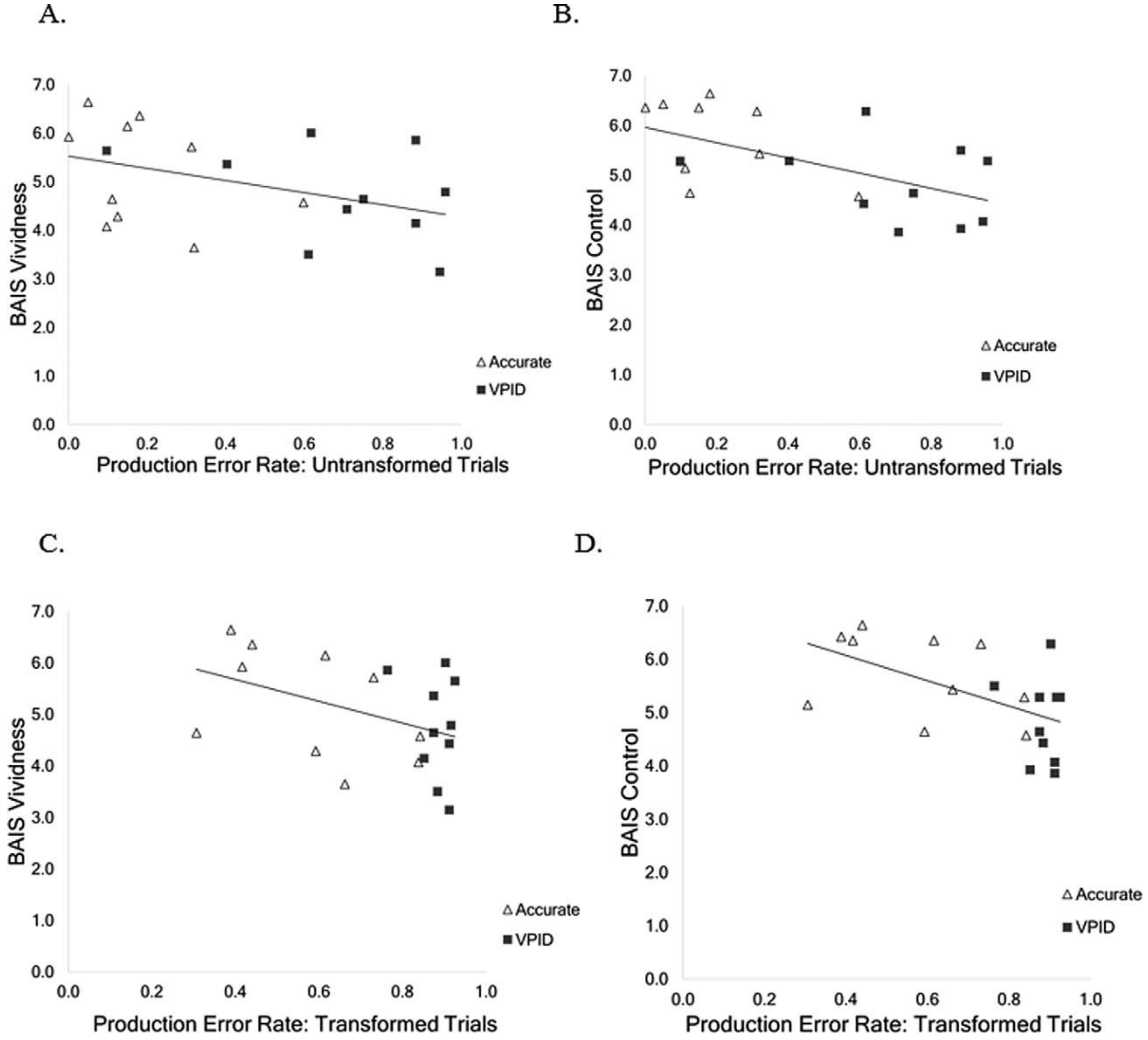

FIGURE 5. Scatterplots displaying the relationship between BAIS subscales (vividness: panels A and C, control: panels B and D) and pitch error rates for untransformed (panels A and B) or transformed (panels C and D) production trials

through forward and backwards digit span measures, respectively. Working memory capacity, as assessed by performance on the backward digit span task, was negatively correlated with how well participants produced transformations of the target sequence, $r(18) = -.49$, $p = .01$ and also with recognition of transformed targets, $r(18) = -.43$, $p < .05$, see Figure 6. Backwards digit span did not correlate with any other measures of production or recognition performance, and there were no significant correlations with forward digit span. These correlations thus suggest that working memory is involved in recognizing or producing mental transformations of auditory sequences. The digit span tasks also did not correlate with either of the BAIS subscales (Vividness or Control). Accurate (M forward $= 6.1$ digits, M backward $= 4.3$ digits) singers did not differ from VPID (M forward $= 6.0$ digits, M backward $= 3.8$ digits) singers in the forward, $t(18) = 0.90$, $p = .38$, $d = 0.14$, or backward digit span, $t(18) = 1.50$, $p = .15$, $d = 0.48$.

Comparing predictor variables. As noted before, mental imagery and working memory are closely linked. As such, we next used multiple regression to assess whether

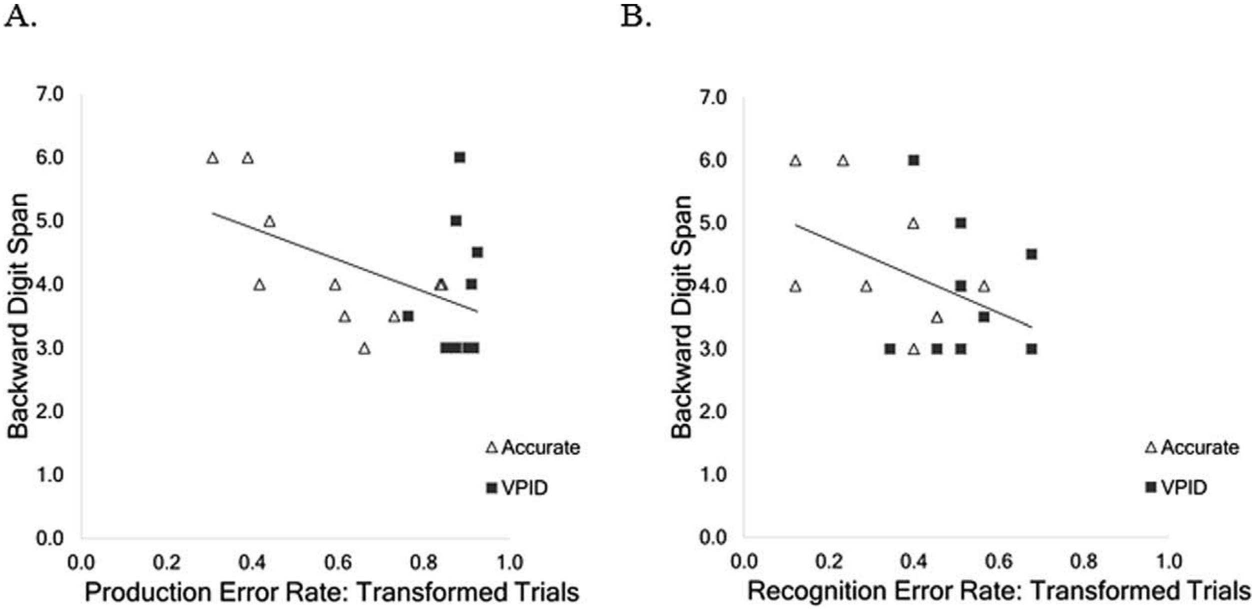

FIGURE 6. A) Scatterplot showing the relationship between pitch error rates across transformed trials in the production task and backward digit span performance. B) Scatterplot showing the relationship between error rates across transformed trials in the recognition task and backward digit span performance.

each of these predictors account for performance independent of the other. Error rates during the production tasks were regressed on the three predictor variables that yielded significant bivariate relationships: BAIS vividness, BAIS control, backwards digit span. Separate regression analyses were performed for the production of untransformed versus transformed targets given that the bivariate correlations suggested different processes contributing to these tasks. We did not perform a similar analysis for recognition data because only one bivariate relationship was significant.

Figure 7 illustrates the results of this multiple regression analysis. X variables are shown on the left of this figure, and Y variables to the right. First, we assessed predictors of error rates during the production of untransformed sequences, shown at the top of Figure 7. The multiple regression equation accounted for a significant amount of variance in production error rates; $R^2 = 0.44$, adjusted $R^2 = .34$, $F(3, 16) = 4.37$, $p < .05$. More important, the only significant partial correlation came from the BAIS Control subscale, $r(18) = -.47$, $p < .05$, represented by a dark, solid line in Figure 7. We next performed a similar regression analysis for error rates during the production of transformed sequences. The multiple regression equation accounted for a significant amount of variance in error rates; $R^2 = 0.46$, adjusted $R^2 = .36$, $F(3, 16) = 4.54$, $p < .05$. For this analysis, the only significant partial correlation was backward digit span, $r(18) = -.48$, $p < .05$, also represented by a dark solid line in Figure 7. In sum, partial correlations suggest that imitation of untransformed melodies is related to imagery processes, specifically imagery control, whereas transforming melodies is related to working memory capacity.

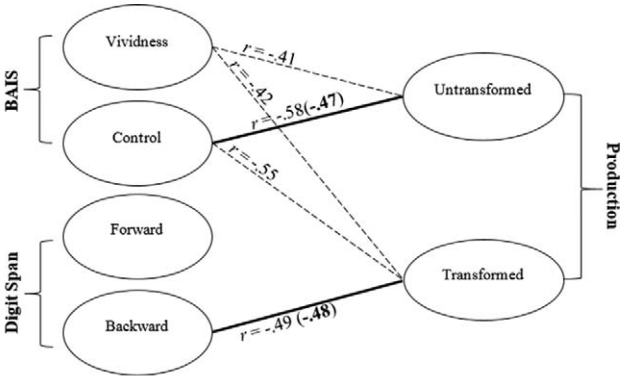

FIGURE 7. Results of multiple regression analyses for error rates in production. Significant bivariate correlations between predictor variables (left) and error rates (right) are shown above each connecting line; absent lines indicate non-significant bivariate correlations. Significant partial correlations from the regression analyses are symbolized by dark solid lines, and the value of significant partial correlations are shown bolded in parentheses. Dashed lines indicate nonsignificant partial correlations. All significant correlations are based on $\alpha = .05$.

SIMULATION

As described in the Introduction, this study was motivated in part by the MMIA model of sensorimotor translation that was designed to account for VPID (Pfordresher, Halpern, & Greenspon, 2015). We therefore tested whether the model could simulate the present results. A limitation of the original model was that it does not account for potential interference effects based on mental transformations in working memory. As such, we modeled the disruptive effect of mental transformations by incorporating a proactive interference parameter for transformation trials based on the auditory-motor mapping associated with a given serial position during a previous iteration of the sequence. As described in Appendix B, this new parameter was used to determine how mental transformations would affect vocal imitation across a continuum of performance. Individual differences in vocal imitation were based on original model parameters (noise and bias in sensorimotor translation). As such, our simulations follow from the assumption that proactive interference effects are constant across both groups (and all levels of performance accuracy). Therefore differences in the size of the transformation effect have to do with how this constant source of interference interacts with basic deficits of sensorimotor translation.

Figure 8 shows the results of the simulation (see Appendix B for details of its implementation). Error rates and the transformation effect arise from simulated model output, and were computed in the same way as we treated the data. Figure 8A shows the simulation of the correlation shown in Figure 3B. Each point in this plot reflects a unique combination of the two variance parameters from the original model, and thus reflects the average across levels of other model parameters (including proactive interference) as well as 10 repetitions of the simulation. As can be seen, the model successfully simulates the negative relationship between overall error rates in production and the magnitude of the transformation effect. The reason for this effect in the model is that when one initially forms a distorted auditory image, proactive interference effects have a relatively smaller effect than when one initially forms a more accurate image. In other words, the VPID image is already poorly formed, so effects of transformation cannot make the image much worse. Note that the correlation is not due simply to the extreme points in the upper left. For instance, if the two values with transformation effects larger than 4 are removed the correlation strength increases (r = -.88).

We also explored whether the MMIA model simulates the main effect of transformation, which is shown in Figure 8B. When we first examined the data, we were surprised that the most disruptive condition was the transformation that involved a transposition of key, which we had predicted would be easier than the rest. As can be seen, the MMIA model simulates this difference across conditions. This happens in the model because pitches in the transposition condition are consistently offset from the original mapping, leading to strong proactive interference. By contrast, pitches in the other transformation conditions are on average more

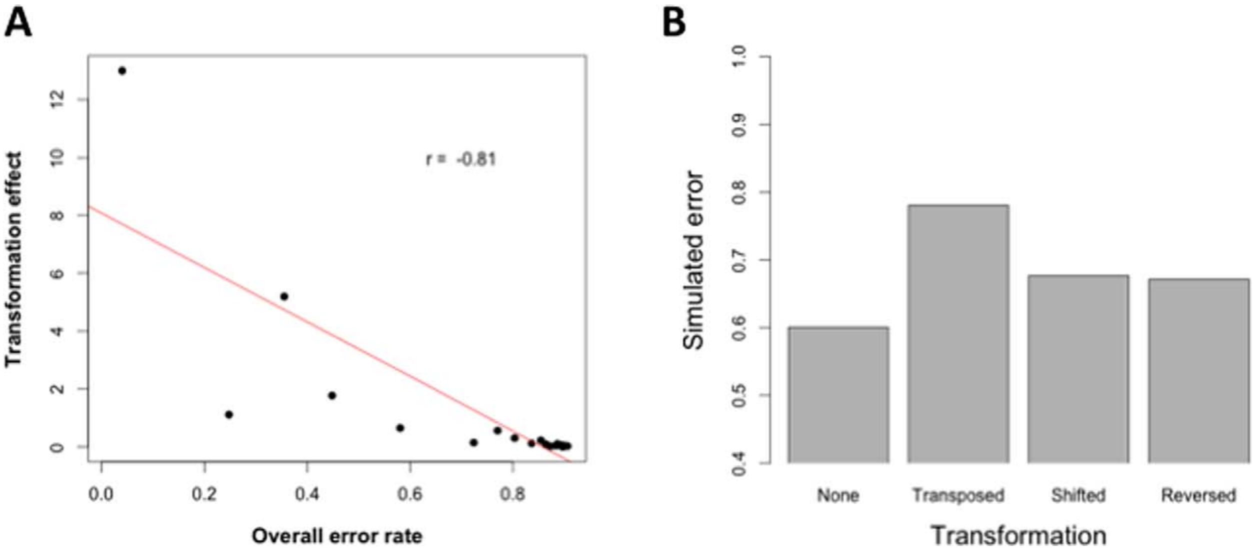

FIGURE 8. A) Simulation of the relationship between total error rates and the relative transformation effect. B) Simulated errors by condition.

proximal to pitches at the same position in the untransformed sequence.

Discussion

In the current study we were interested in whether VPID is related to a deficit of mental imagery. For this reason, we were interested in how varying demands on imagery processes affect the accuracy of vocal pitch imitation. We found that increased imagery demands differentially affected production performance for accurate and VPID singers, a task that relies on multimodal imagery. However, both groups' performance was disrupted to a similar degree in the recognition task, a task we interpret as relying on unimodal imagery. Therefore, we suggest that VPID is related to a deficit in multimodal imagery, such that VPID singers have an inaccurate mapping between their perceptual and motor representations.

The most important finding from the present study was the fact that producing mental transformations of short melodies led to larger effects for accurate singers than VPID singers. This result at first may seem counterintuitive. If VPID relates to a deficit of mental imagery, why is the effect of manipulating the image smaller for this group than for accurate singers in the production task? The answer lies in the effect of transformations on an accurate versus poorly formed initial image, as simulated in the extension of the MMIA model (Pfordresher, Halpern, & Greenspon, 2015; Appendix B). Critically, in these simulations differences in production accuracy of the untransformed condition were due entirely to sensorimotor translation between individual pitches, as opposed to simulations of the transformed conditions that were additionally influenced by proactive interference. The model produced the same pattern of results observed in our data: The effect of transformation scaled inversely with overall production error rates because the effect of interference was greater when the initial sensorimotor associations were computed more accurately.

Thus, as suggested by our results and simulations from the MMIA model, VPID deficits may be based on the inability to form an accurate multimodal image. However, both groups appear to have similar difficulty manipulating these images. This interpretation suggests that different mechanisms may underlie imagery formation versus manipulation (cf. Dror & Kosslyn, 1994; Lequerica et al., 2002). Not surprisingly, group differences were large and significant for the production of untransformed melodies—the condition that most likely measures imagery formation. Correlations between performance on this task and both subscales of the Bucknell Auditory Imagery Scale (BAIS; Halpern, 2015) confirmed the role of auditory imagery in the production of melodies that have not been transformed.[6] This interpretation was further supported by our multiple regression analysis, in which imagery control had the only significant partial correlation with production errors for untransformed trials. With respect to the production of mentally transformed melodies, multiple regression analyses suggested that working memory capacity, rather than auditory imagery, plays a dominant role in distinguishing individual performance. Individuals with larger working memory capacity (as measured by backwards digit span) were more accurate at producing and recognizing correct transformations than individuals with lower capacity.

We also assessed how well accurate and VPID imitators recognize melodies and their transformations. These conditions were included to assess whether VPID deficits are specific to tasks that require sensorimotor translation. In the recognition task, VPID singers tended to have more errors in the untransformed condition than accurate singers; however, performance was not significantly different between the two groups. This is in contrast to the significant group difference in untransformed performance found in the production task. We suggest that one reason the production and recognition tasks differentially affected group performance in the untransformed condition is because the recognition task relies more strongly on unimodal imagery than the production task, which relies on multimodal imagery. Therefore, VPID appears to be more specifically related to a deficit in multimodal imagery. This is in line with our findings from the pitch discrimination task: VPID singers do not differ from accurate singers in a task that recruits only the auditory domain.

An alternative explanation for our recognition task is that it is predominantly a pitch memory task rather than an imagery task (Dewar, Cuddy, & Mewhort, 1977). We do not think this is the case. First, if participants were using a pitch memory strategy then they should be most disrupted in the transposition condition in which all pitches in the comparison melody differ

[6] In the present results, both BAIS subscales predicted performance, whereas Pfordresher and Halpern (2013) only found a significant correlation of the Vividness subscale with vocal pitch matching. A critical difference between the two studies is the type of stimuli used in the two experiments. Whereas the current study used sequences that contained three or four different pitches, Pfordresher and Halpern (2013) used monotone sequences. Therefore, imitating more complex stimuli may increase demands on imagery control compared to imitating less complex stimuli.

from the target melody. However, participants are most accurate in the untransformed condition but are similarly disrupted across all three transformation conditions. Second, altered pitches in our "incorrect" comparison melodies were designed to make a pitch memory strategy highly difficult. As described in our *Method*, altered pitches followed the same contour and key as the "correct" comparison melodies. For this reason, altered pitches were unlikely to be more perceptually salient to the listener relative to the other pitches in the melody. For these reasons, we interpret the recognition task as an auditory imagery task.

The difficulty of the present mental transformation tasks suggests a somewhat surprising degree of inflexibility in mental images of pitch. Consider the participants' ability in the backwards digit span in comparison to producing a melody in reverse order. In the backward digit span task all participants were able to correctly reverse sequences comprising three digits. However, both accurate and VPID singers were poor at reversing sequences composed of three pitches. In fact, accurate singers' performance when transforming target melodies led to error rates that approached VPID baseline performance. In other words, introducing mental transformations into a production task reduces an otherwise accurate performer to the status of a VPID individual. The discrepancy between performance in the singing task and the digit span task is a noteworthy finding in that it suggests that manipulating pitch information may be more cognitively taxing than manipulating verbal information (Deutsch, 1970).

In conclusion, the current study informs our understanding of VPID. A critical finding from the current study is that effects of imagery manipulation are smaller for VPID singers than accurate singers. We interpret this finding as evidence that VPID singers have a deficit in imagery formation and thus exhibit relatively small effects of transformation. Interestingly, we also found that auditory images tend to be inflexible such that both accurate and VPID singers exhibit difficulty controlling these images. Our results contribute to our understanding of imagery as a multi-component process involving both image formation and the ability to manipulate images. We found that both processes are related to vocal imitation accuracy. The contribution of imagery control to singing performance is a novel finding. Together, these findings expand our understanding of how auditory images relate to vocal imitation.

Author Note

This research was supported by NSF Grant BCS-1246964.

Correspondence concerning this article should be addressed to Emma Greenspon, Department of Psychology, University at Buffalo, SUNY, Buffalo, NY 14260. E-mail: ebgreens@buffalo.edu

References

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., et al. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*, *30*, 917-926.

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 501-515.

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. *The Quarterly Journal of Experimental Psychology, 65*, 1161-1171.

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

Boersma, P., & Weenink, D. (2013). *Praat software.* Amsterdam, Netherlands: University of Amsterdam.

Christiner, M., & Reiterer, S. M. (2013). Song and speech: Examining the link between singing talent and speech imitation ability. *Frontiers in Psychology, 4,* 1-11.

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.

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.

Dalla Bella, S., Giguère, J. F., & Peretz, I. (2009). Singing in congenital amusia. *Journal of the Acoustical Society of America*, *126*, 414-424.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science*, *168*, 1604-1605.

Dewar, K. M., Cuddy, L. L., & Mewhort, D. J. (1977). Recognition memory for single tones with and without context. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 60-67.

Dror, I. E., & Kosslyn, S. M. (1994). Mental imagery and aging. *Psychology and Aging, 9*, 90-102.

Engel, A., Bangert, M., Horbank, D., Hijmans, B. S., Wilkens, K., Keller, P. E., & Keysers, C. (2012). Learning piano melodies in visuo-motor or audio-motor training conditions and the neural correlates of their cross-modal transfer. *NeuroImage*, *63*, 966-978.

Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed.). New York: Harper and Row.

Foster, N. E., & Zatorre, R. J. (2010). Cortical structure predicts success in performing musical transformation judgments. *Neuroimage*, *53*, 26-36.

Foster, N. E., Halpern, A. R., & Zatorre, R. J. (2013). Common parietal activation in musical mental transformations across pitch and time. *Neuroimage*, *75*, 27-35.

Halpern, A. R. (2015). Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain*, *25*, 37-47.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, *9*, 697-704.

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, *42*, 1281-1292.

Hommel, B. (2009). Action control according to TEC (Theory of Event Coding). *Psychological Research, 73*, 512-526.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*, 849-937.

Hubbard, T. L. (2010). Auditory imagery: Empirical findings. *Psychological Bulletin*, *136*, 302-329.

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.

James, W. (1890). *The principles of psychology.* Cambridge, MA: Harvard University Press.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Lequerica, A., Rapport, L., Axelrod, B. N., Telmet, K., & Whitman, R. D. (2002). Subjective and objective assessment methods of mental imagery control: Construct validations of self-report measures. *Journal of Clinical and Experimental Neuropsychology*, *24*, 1103-1116.

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

Marks, D. F. (1999). Consciousness, mental imagery and action. *British Journal of Psychology*, *90*, 567-585.

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.

Moore, R., Estis, J., Gordon-Hickey, S., & Watts, C. (2008). Pitch discrimination and pitch matching abilities with vocal and nonvocal stimuli. *Journal of Voice, 22*, 399-407.

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., 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 and Review*, *20*, 747-753.

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*, 242-253.

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.

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.

Phillips-Silver, J., & Keller, P. E. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Frontiers in Human Neuroscience*, *6*, 26.

Reisberg, D., Smith, J. D., Baxter, D. A., & Sonenshine, M. (1989). 'Enacted' auditory images are ambiguous; 'Pure' auditory images are not. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *41*(3-A), 619-641.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701-703.

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

Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, *33*, 1433-1454.

Sun, X. (2002). *Pitch determination algorithm* [Matlab Central File Exchange]. Natick, MA: MathWorks.

Sundberg, J. (1987). *The science of the singing voice*. Dekalb, IL: Northern Illinois University Press.

Vuvan, D. T., & Schmuckler, M. A. (2011). Tonal hierarchy representations in auditory imagery. *Memory and Cognition*, *39*, 477-490.

Welch, G. F. (1979). Vocal range and poor pitch singing. *Psychology of Music*, *7*, 13-31.

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.

Zarate, J. M. (2013). The neural control of singing. *Frontiers in Human Neuroscience*, *7, 1-12.*

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8*, 547-558.

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.

Zurbriggen, E. L., Fontenot, D. L., & Meyer, D. E. (2006). Representation and execution of vocal motor programs for expert singing of tonal melodies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 944-963.

Appendix A

TABLE A1 *Mean Absolute Pitch Deviations*

	Condition				
Group	*Normal*	*Transposed*	*Shift*	*Reversed*	TE
Accurate	49.07 (14.73)	253.88 (23.32)	176.47 (34.46)	167.13 (21.42)	3.06
VPID	170.72 (36.86)	350.89 (25.53)	263.3 (29.41)	290.34 (20.85)	0.77

Note: Standard errors are in parentheses. TE = transformation effect.

TABLE A2. *Mean Proportion of Contour Errors*

	Condition				
Group	*Normal*	*Transposed*	*Shift*	*Reversed*	TE
Accurate	.02 (.02)	.23 (.01)	.26 (.03)	.19 (.02)	12.6
VPID	.12 (.07)	.24 (.05)	.27 (.05)	.29 (.05)	1.16

Note: Standard errors are in parentheses. TE = transformation effect.

Appendix B

DETAILS OF SIMULATION

As discussed in the Introduction, this research follows from theoretical assumptions that have recently been articulated in the MMIA model (Pfordresher, Halpern, & Greenspon, 2015). We now describe an extension of this model that simulates the present results. This simulation was motivated by a specific interpretation of the results and thus functions as a test of this interpretation. We include a brief summary of the model; a more complete summary of the original model can be found in the original article.

In order to simulate the challenge associated with transformation effects, we added to the model a source of proactive interference based on previous mapping relationships between perception and action. The assumption underlying this new model component is that people encode sequential associations between perception and action and are influenced by these associations during future performances. In the context of mental transformations, then, the association one forms during initial encoding lead to difficulty when transforming the mental image.

We implemented this new model component as follows. We first ran the model through a single iteration of the target sequence, with every pitch value (y) being mapped probabilistically to an associated motor target (x), as in the original model. In the original model, this mapping is influenced by two joint probability distributions. The first distribution, *Zmap*, accounts for the degree of precision in the mapping between y and x. Mapping becomes less precise (more variable) as a variance parameter for this distribution increases. The second probability distribution, *Zbias*, maps every value of y to a single value of x, called *x-bias*, that reflects a participants "comfort pitch" (cf. Pfordresher & Brown, 2007). A variance parameter for *Zbias* determines the strength of this biasing effect, which is stronger when the variance parameter is smaller. Thus, three parameters influence mapping in the original model: The variance parameters for each distribution, plus the value of *x-bias*.

We then ran the model through a second iteration designed to simulate the process of mental transformations. During this second run, we incorporated the extended model. In addition to the two error sources in the original model, the extended model included a vector of past target values called $x\text{-}prior_t$, with t indexing sequence position. Each value of x in this vector comes from the sensorimotor associations from the first run. As with the original sources of error, *x-prior* was modeled as a probability distribution:

$$Zprior_{i,j,t} = \exp\left[\frac{(y_i - xprior_t)^2}{2\sigma^2_{prior}}\right] \qquad \text{(B1)}$$

This equation generates a matrix of probabilities (Z values) for mapping between values of y (perceived pitch, indexed by the subscript i) and x (vocal pitch, indexed by the subscript j). Probabilities are based on differences between values of *y* and the *x* value that was associated with the present serial position during the previous run (*xprior)*. The variance component determines how influential this source of error is: The influence of *Zprior* is greater when its variance parameter is low. When constructing a transformed mental image, we assume that biases from prior mappings influences production jointly with the two components from the original model. We modeled this interaction by multiplying all three distributions.

Previous simulations reported by Pfordresher, Halpern, and Greenspon (2015) suggest that VPID in general can be modeled based on the ratio of variance associated with mapping (*Zmap*) to variance associated with bias (*Zbias*). VPID-like behavior occurs when variance of mapping outweighs variance of bias (as in *Zprior,* low variance of response bias leads to poorer sensorimotor mapping). We reasoned that the greater effects of transformation for accurate singers, observed here, may be found because for VPID singers, any mapping from perception to action is influenced by this constant bias parameter, whereas for accurate singers, effects of prior mappings may lead to substantially larger deviations of produced from target pitches for productions of transformed as opposed to original sequences.

We tested these ideas in simulations based on pairs of runs with each pair constituting a trial. On the first run of each trial the model mapped pitches from a single target sequence with pitches (in cents) encoded as [0, 200, 400, 700], or [C D E G]. Across trials, the value of

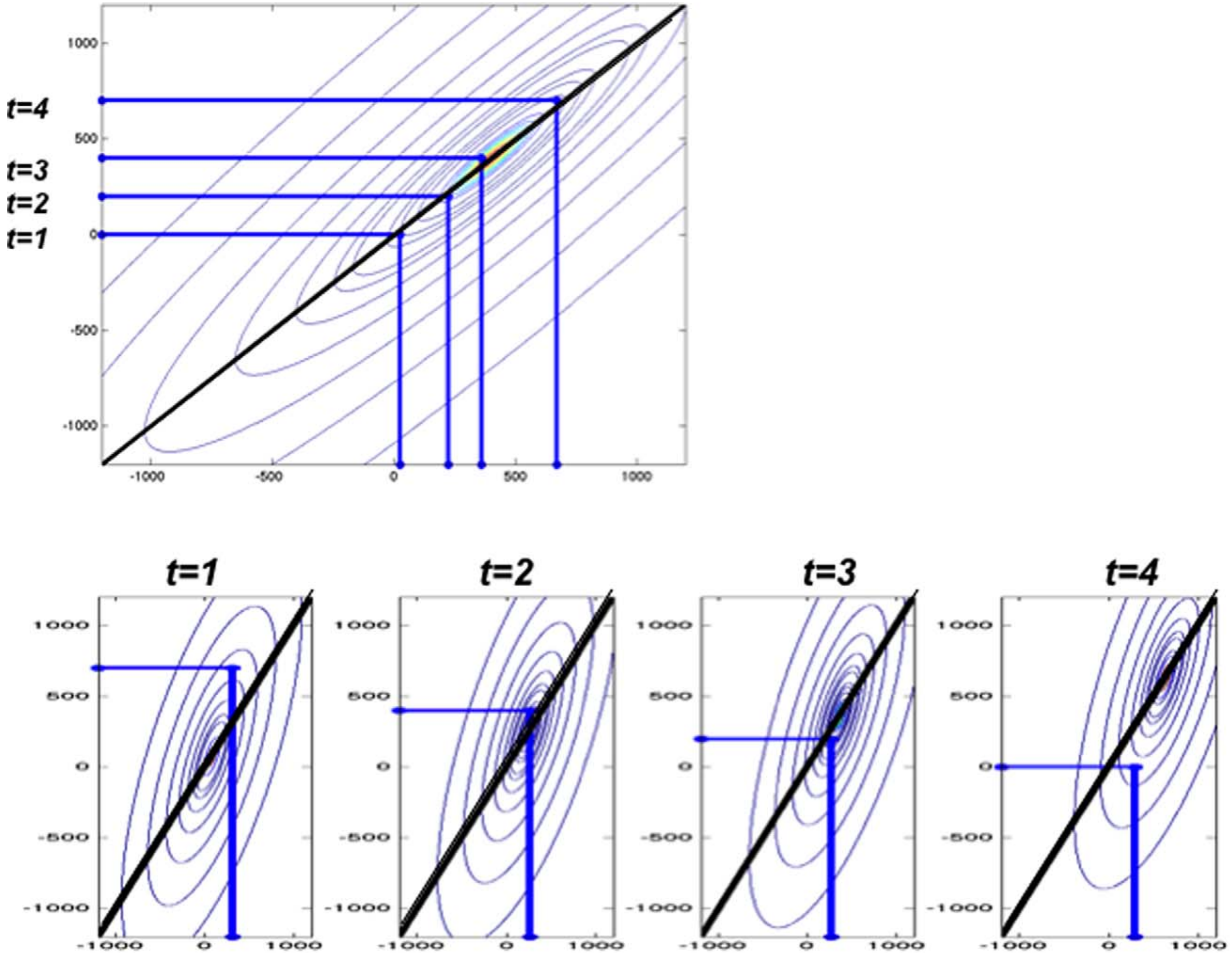

FIGURE B1. Example of a simulated trial for an accurate imitator (low variance for mapping relative to bias). In each panel, the ordinate represents the auditory image of pitch height, whereas the abscissa indicates the motor image to which that image is mapped. The top panel shows mapping relationships in the first run (no influence of *Zprior*), and the four lower panels represent mapping of each successive pitch in the second run, in

x-bias was varied from -400 cents to +400 cents in 200-cent steps but this parameter remained constant within a trial. Standard deviations for all components were likewise varied across trials (but not within a trial) from 40 to 100 cents in 20-cent intervals; the standard deviation of the mapping component included all of these values plus 10 cents and 20 cents (lower values of mapping variance are necessary to simulate accurate performances). For each combination of parameter values (480 in all), ten trials were simulated, with the first run in each trial establishing original mapping of y to x, and the next three runs run showing the influence of *x-prior* for different transformation conditions. Thus, 9,600 4-note trials were simulated in all (38,400 simulated mappings).

Two illustrative examples highlight the critical components of the manipulation. Figure B1 shows the simulation for a precise and relatively unbiased mapping of y to x, as can be seen by the proximity of intersections between x and y values to the major diagonal in the upper plot (run 1). None of the imitated pitches would be considered errors based on the relationship between *y* and *x* values. However, the biasing effect of these initial mappings leads to substantially higher error on run 2, which is a reversal transformation. Note how the locus of bias shifts in these lower plots, based on the original mapping. The effect is particularly noticeable on the first and last notes, which are biased in the direction of the first run, in which these notes represented opposite extremes of the pitch range. This happens

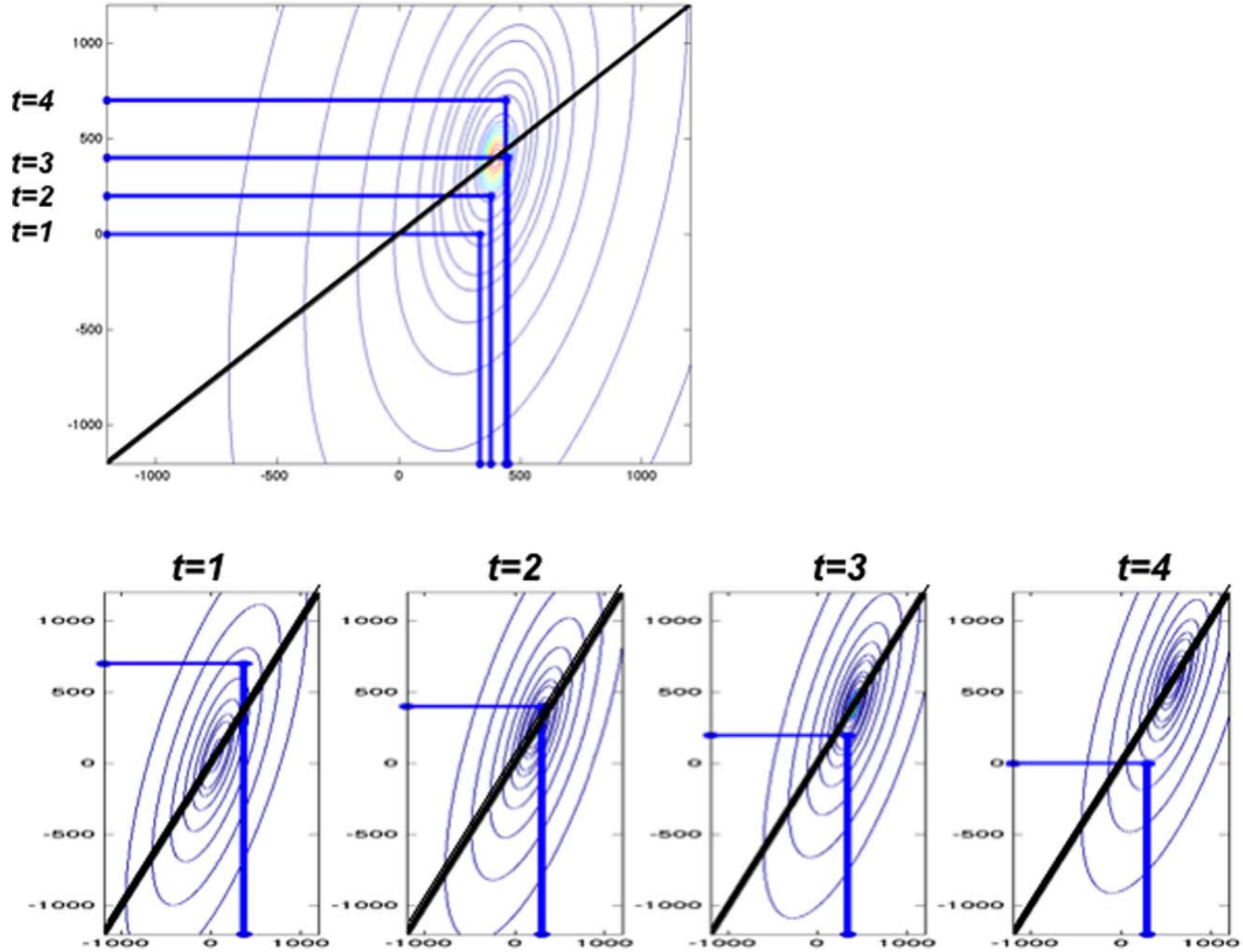

FIGURE B2. Example of a simulated trial for an inaccurate imitator (high variance for mapping, relative to bias). In each panel, the ordinate represents the auditory image of pitch height, whereas the abscissa indicates the motor image to which that image is mapped. The top panel shows mapping relationships in the first run (no influence of *Zprior*), and the four lower panels represent mapping of each successive pitch in the second run, in which *Zprior* varies for each position (*t*). The unity line in each panel represents accurate mapping; contour plots indicate regions of probability decreasing from the center.

because the variance associated with prior mapping is lower (40 cents) than the variance associated with overall bias (100 cents). The relative transformation effect from this example is 6.72, based on absolute deviations of produced pitches from target pitches (a better measure for examples than error rates, given that only 4 pitches are used).

Alternatively, Figure B2 shows an example in which a small transformation effect may occur for a generally poor singer. In this simulation, variance associated with overall bias is low relative to the variance of mapping, with the same variance associated with prior mappings as in the previous example. Imitation in run 1 is strongly influenced by overall bias, leading to poor performance (only one correct pitch). Although performance of the reverse transformed sequence is also poor, production does not get noticeably worse than in the first run. The relative transformation effect from this example is 0.11.

In order to simulate the empirical correlation between overall error rates and the relative transformation effect (Figure 3B), we converted each simulated produced pitch (x-values) to an error score based on intended pitches, and computed overall error rates and relative transformation effect scores as in the original study. The resulting correlation is shown in Figure 8A. Each dot reflects a unique combination of variances for *Zbias* and *Zmap*, and were averaged across all other parameters. It is important to note that in averaging across values used for *Zprior* we do not let that model component contribute to the correlation. The relationship between overall error rate and the transformation effect is negative, as in the obtained data, with a correlation coefficient of similar magnitude.