Psychomusicology: Music, Mind, and Brain
2013, Vol. 23, No. 2, 82–99

0275-3987/13/$12.00 DOI: 10.1037/a0033299

Transfer Effects in the Vocal Imitation of Speech and Song

Matthew G. Wisniewski, James T. Mantell, and Peter Q. Pfordresher
University at Buffalo, The State University of New York

In this study, we investigated how practice imitations of speech impacted imitations of songs and vice versa. Participants were first asked to practice imitating sung or spoken sequences, and then to imitate a new sequence, which could differ with respect to domain (speaking or singing), global pitch contour (question vs. statement pattern), and/or words. Pitch accuracy during transfer was affected by changes to domain and contour, but not text. Somewhat surprisingly, best transfer was found either when both domain and contour remained the same or both changed. Transfer performance suffered when only one feature changed and the other remained consistent. Analyses of individual differences showed that poor-pitch imitators had a harder time adopting the pitch structure of new sequences, regardless of whether the sequence was speech or song. Results were not consistent with claims for either independence or complete integration of music and language, but instead argue for differences in domain possibly based on the salience of pitch structures in the signal.

Keywords: vocal imitation, imitation learning, poor-pitch singers, generalization

Debate concerning the systems underlying music and language has centered around the processing of pitch. The influential modular model of Peretz and Coltheart (2003) focuses on processing of pitch as a critical difference between the domains. Specifically, this model predicts that pitch information within a musical context (e.g., in song) is processed by a tonal encoding module independent of spoken pitch. Other accounts, here termed integrationist, propose that pitch is processed similarly in both speech and musical contexts (e.g., Koelsch, 2011; Koelsch & Seibel, 2005). For instance, Koelsch and Siebel (2005; see also Koelsch, 2011) propose that the extraction of meaning from pitch and timbre occurs in similar modules for speech and music. Similarly, several auditory scene analysis models (e.g., Bregman, 1990; Patterson, Allerhand, & Giguere, 1995; Yost, 2007) propose that the perception of speech and music is formed by the same set of computations.

The validity of independence versus integrationist accounts for pitch perception in music and language continues to be a source of debate. However, our perspective has to do with the ability to produce pitch patterns by imitation. Two points motivate this focus. First, research in both music and language cognition is dominated by perception as opposed to production, leaving open to question whether similar effects found in perception research also hold for production. Second, research has suggested that some individuals exhibit a production-specific musical pitch deficit, here termed poor-pitch singing (Pfordresher & Brown, 2007; Roberts & Davies, 1975; Welch, 1979), that may exist in the absence of any perceptual deficit such as congenital amusia (Peretz et al., 2002). Both points led us to explore whether the imitative production of speech incorporates a pitch processing system that is distinct from the imitative production of song.

Independent Versus Integrationist Views of Music and Language

Models of music and language processing can be said to fall along a continuum bracketed by the extreme views of full *independence* and complete *integration*. A fully independent view, following Fodor's description of modularity (Fodor, 1983, 2000), conceptualizes speech processing as comprising a distinct (*domain specific*) set of processing modules that are not shared with music processing, and furthermore are not influenced by information processing that occurs within the music system (*information encapsulation*). A similar set of constraints should exist for music processing. One view that approximates this approach, though focusing on domain specificity rather than encapsulation, is the aforementioned model of Peretz and Coltheart (2003). With respect to pitch processing—whether in perception or production—a fully independent view would predict that constraints or benefits associated with pitch processing in one domain (e.g., spoken pitch contour) should not transfer to the other domain (sung pitch). On the other hand, a fully integrationist view constitutes a completely unified system underlying the processing of information across both domains (Koelsch, 2011; Koelsch & Seibel, 2005). Hence, the effects in one domain (whether beneficial or interfering) should transfer to the other completely.

A great deal of research has led to conflicting support for integrationist versus independence views. On the integrationist side, language and music share many characteristics, and the two domains frequently interact. Recent evidence suggests that spoken pitch intervals conveying sadness match the minor third in music (Curtis & Bharucha, 2010); individuals exhibiting deficient imitation of sung pitch often exhibit deficient imitation of speech (Mantell & Pfordresher, 2013; Pfordresher & Mantell, 2009), and native language influences the accuracy of imitations for speech and song tokens (Pfordresher & Brown, 2009). Similarities and

Matthew G. Wisniewski, James T. Mantell, and Peter Q. Pfordresher, Department of Psychology, University at Buffalo, The State University of New York.

Correspondence concerning this article should be addressed to Peter Q. Pfordresher, Department of Psychology, Park Hall room 204, Buffalo, NY 14260. E-mail: pqp@buffalo.edu

interactions between language and music have been observed in a variety of other tasks that look at syntax (Patel, 2003; Slevc, Rosenberg, & Patel, 2009), characteristics of songs written by composers with different native languages (Patel & Daniele, 2003), and memory for lyrics and melodies (Serafine, Davidson, Crowder, & Repp, 1986). Also, electrophysiological research has shown that some event-related potential (ERP) components, traditionally obtained using linguistic stimuli, show similar dynamics using musical stimuli. For instance, a larger N400 response is elicited by the presentation of unprimed versus primed words and musical excerpts (Daltrozzo & Schön, 2009; for review see Koelsch, 2011). Several theorists have used the above-mentioned findings to argue that music and language have a common evolutionary origin (Brown, 2000), that music and language are processed in overlapping brain systems (Griffiths, Johnsrude, Dean, & Green, 1999; Patel, 2003), and that memories for linguistic and musical experiences are integrated (Serafine et al., 1986).

There is other evidence, however, that the processing of language and music may be independent. People with amusia have problems recognizing previously heard melodies but may show no significant impairments in their ability to recognize previously heard lyrics (Ayotte, Peretz, & Hyde, 2002; Peretz & Coltheart, 2003). Similarly, there are aphasic populations showing deficits in language tasks but normal performance in melodic tasks (Marin & Perry, 1999). Neuroimaging studies of cortical areas associated with processing stimuli from either domain suggest a hemispheric asymmetry with the left hemisphere being dominant for language and the right hemisphere dominant for music (Peretz & Zatorre, 2005; Riecker, Ackerman, Wildgruber, Dogil, & Grodd, 2000; Wong, Parsons, Martinez, & Diehl, 2004; Zatorre, Evans, Meyer, & Gjedde, 1992). Also, in an fMRI study looking at neural correlates of singing and speaking, several brain areas (right inferior frontal gyrus, right premotor cortex, and right anterior insula) were activated when singing lyrics, but not when speaking them (Saito, Ishii, Yagi, Tatsumi, & Mizusawa, 2006). These studies, and several others (for review see Peretz, 2009; Peretz & Zatorre, 2005), suggest that it is possible to damage processing in one domain without affecting the other, and that processing of language and music in normal individuals may engage separate systems.

In short, the degree to which music and language draw on shared or distinct systems is far from settled. Given the diversity of results summarized above, it seems likely that an intermediate approach combining elements of independence and integration is likely to provide the best account for all of the data (e.g., Patel, 2008). Along these lines, the study we report here addresses the relative contributions of domain-specific constraints (whether a sequence is representative of music or language) as well as more stimulus-driven constraints such as pitch contour (which may be similar across domains). To observe direct carry-over effects from one domain to the other (or lack thereof), we adopted a transfer-of-training paradigm.

Evidence for Transfer Within and Across Domains

Transfer in general refers to a carry-over effect from learning in one task to learning or performance of a subsequent task (Schmidt & Lee, 1999). Most important to the present research is that the presence of transfer suggests that prior learning *primes* representations used for the subsequent task (Bock, 1986). Thus, the presence of beneficial transfer suggests that the two tasks involve a common representation that is activated during learning. By contrast, the absence of transfer suggests full independence, and negative transfer (diminished performance in the subsequent task) suggests inhibitory effects, possibly based on some kind of processing bottleneck. Recent research on music and language, discussed below, has looked to cross-domain transfer effects for evidence of integration versus independence. To our knowledge, this research has relied on long-term transfer effects with cross-sectional designs. By contrast, we address the effects of short-term transfer to look at immediate causal relationships across domains.

Some studies have suggested that musical training can enhance the processing of pitch in early stages of the auditory system, therefore having a beneficial effect on the processing of speech (for review see Kraus & Chandrasekaran, 2010). For instance, Kraus and colleagues (e.g., Parbery-Clark, Strait, & Krauss, 2011) have found that auditory-evoked responses of the brainstem are more strongly correlated with acoustic signals in musicians than nonmusicians, with positive correlations indicating better speech comprehension. Such beneficial transfer from music training to language processing has led to the recently proposed OPERA (overlap, precision, emotion, repetition, & attention) hypothesis (Patel, 2011, 2012). Part of this proposal has to do with the assumption that some amount of overlap (the “O” in “OPERA”) exists across music and language systems. Evidence also suggests that certain kinds of language background may facilitate aspects of music processing. Tone language speakers have been found to imitate sung pitch more accurately on average than native English speakers (Pfordresher & Brown, 2009), specifically for patterns of varying pitches as opposed to single pitches. Some evidence for a tone-language advantage in the perception of musical pitch has also been found (Bidelman, Gandour, & Krishnan, 2011; Giuliano, Pfordresher, Stanley, Narayana, & Wicha, 2011; Pfordresher & Brown, 2009); however, a recent study calls into question whether this perceptual advantage is language based or culture based, by showing an advantage among speakers of Mandarin tone language but not among speakers of the Hmong tone language (Hove, Sutherland, & Krumhansl, 2010).

Long-term transfer effects support integrationist views, yet there are limitations to this approach. Aside from concerns inherent in any cross-sectional design (as in the studies referred to above), the focus on long-term effects prevents one from delving deeper into stimulus-specific factors that might influence processing across domains. For instance, even if experience speaking a language facilitates the imitation of musical pitch (as in Pfordresher & Brown, 2009), it is unclear from this finding whether the imitation of a specific spoken pitch pattern will facilitate the imitation of a similar musical pitch pattern. Such a result is predicted by integrationist approaches; however, one needs to examine the evolution of immediate transfer effects over time to address this question properly.

In single-session transfer paradigms, experimenters can manipulate training experiences rather than rely on the grouping of individuals based on their reports of prior music and language experiences (Peretz, 2009). However, to date such studies have only addressed transfer within a single domain (music or language) and not across domains. In one transfer study, bilingual speakers of German and English practiced producing a target sentence as fast as possible (MacKay & Bowman, 1969). Shorter production durations over the course of practice indicated better learning, with positive transfer occurring when durations in transfer benefitted from increases in reading speed during training. Positive transfer of learning was found when participants switched from sentences in one language to translations of the sentence in a different language, even when translations resulted in changes to word order. Transfer was not found if sentences in the new language were not translations. Thus, transfer of learning was determined by the abstract message being communicated, independently of specific word order. A later music performance study demonstrated similar constraints on transfer in the domain of music. Palmer and Meyer (2000; cf. Meyer & Palmer, 2003) asked pianists to practice keyboard sequences so that they could play them as fast as possible. Total duration of performances diminished with learning, as in MacKay and Bowman (1969), and transfer of learning was found when the transfer melody was structurally identical to the practiced melody, even when pianists switched hands during transfer. Taken together, these results suggest that transfer may be independent of motor movements, driven instead by the cognitive systems used to plan movements. Beneficial transfer effects across domains likewise would imply the use of shared systems during both training and transfer.

The Present Study

In this study, we adapt the transfer task of MacKay & Bowman (1969; cf. Palmer & Meyer, 2000) to the vocal imitation of spoken versus sung pitch. At issue was whether pitch imitation relies on a single integrated system, separate independent systems, or some kind of hybrid system. Participants listened to a stimulus (the target) and then imitated it as accurately as possible (the imitation). We adopted an intentional imitation task, as opposed to an incidental imitation task, for two reasons. First, our concern was primarily with the specific functioning of the pitch imitation system, rather than the conditions under which participants may engage in imitative or nonimitative behaviors (which deals more with response strategies). Second, according to a truly modular view, one would not expect the intention to imitate to cause a participant to lapse into a kind of “song” mode. There is nothing specifically musical about vocal imitation. Adults unintentionally imitate speech in social situations (e.g., Pardo, 2006), and intentional vocal imitation of speech is done regularly when children learn to speak, when adults learn a new language, or when adults use imitation for demonstrative purposes (e.g., telling a story).

In our task, participants repeatedly imitated the same target eight times during a training phase, followed by another target for four transfer trials. Transfer targets could differ from the training target with respect to the domain they represent (speech or song), global pitch contour (statement or question), and the text that was spoken or sung. Our data analyses focused both on group and individual comparisons. Previous imitation studies have revealed that there are large individual differences with regard to pitch imitation ability and that these differences impact how well one can imitate under different conditions (Mantell & Pfordresher, 2010, 2013; Pfordresher & Brown, 2007; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010). Independent accounts can predict different trends in individual differences across domains, whereas domain general accounts predict that the same trend should apply across domains.

According to a fully independent view, transfer from one domain to another forces the participants to switch between two autonomous and encapsulated systems. As such, no benefit should be found when transferring across domains, regardless of how structurally similar the target patterns are. By contrast, according to a fully integrationist account, transfer should be determined entirely by the acoustic similarity across targets, with domain playing only an incidental role insofar as domain differences are associated with acoustical differences. An integrationist view would thus predict large transfer effects based on contour, and smaller effects of domain. A third, less well determined possibility comes from a hybrid approach, which would follow if the predictions of neither extreme view are found.

Methods

Participants

Sixteen students (eight female) from The University at Buffalo, The State University of New York, participated in exchange for credit in an introductory psychology course. Participants were randomly assigned to either imitate speech during practice trials or to imitate song during practice trials ($n = 8$ each). Age in the sample ranged from 18 to 21 years ($M = 18.86$, two participants declined to report age).

The sample was dominated by nonmusicians. Thirteen participants reported no musical training whatsoever; among the three remaining participants, two reported 3 years of training (flute in one case, trombone in the other). The final participant had 9 years of vocal training and 7 years of piano training. Performance of the musician, though highly accurate, did not differ qualitatively from the other participants, and vocal imitation accuracy in this participant (who imitated song during practice) was not the most accurate within the sample. There were five participants whose mean signed pitch error in the experiment was greater than 100 cents (one semitone) off pitch (either flat or sharp). Three of these poor-pitch imitators (Pfordresher & Brown, 2007) were in a condition that practiced imitating song targets. The other two practiced speech.

Twelve participants were monolingual English speakers. Of the remaining four participants, two reported first learning English in tandem with Hindi (one imitated song during practice, the other imitated speech), and the remaining two first learned an Asian language before learning English (one learned Mandarin, and imitated song during practice, the other learned Korean and imitated speech during practice). Importantly, all participants were fluent English speakers, and all but one participant (the Korean speaker) reported English as their most comfortable language.

Stimuli

Stimuli comprised 12 sequences of three to five syllables that were sung or spoken with a global pitch contour denoting a question or statement. The entire set of 48 sequences was produced by both a male and a female model, for use with male or female participants, respectively.[1]

The stages in construction of stimuli were as follows. First, the male and female models, independently of each other, spoke each word sequence (see Appendix A for text) as a statement and as a question. Spoken statements featured a global contour that descended toward the last syllable, whereas questions culminated in a rising contour. The statement/question difference defines contour in this study. Thus, the contour of a sequence is not defined as pitches that only go up or down, but is defined in a global sense.

Song targets were created by first composing melodies that had the same melodic contour as sentences (one note per syllable), based on changes in the mean F0 within each spoken syllable across the entire sentence. The global contour, as well as local contour, was matched across spoken and sung targets, and both featured the same text. Figure 1 displays the mean contours for statements and questions in pitch-time. The *y*-axis depicts midi values. Midi units correspond to 100 cents (= 1 semitone) in pitch space, and midi values range from 0 (diatonic note C0 at 8.18 Hz) to 127 (diatonic note G10 at 12,543.85 Hz). Interpolation was required to depict sequences of differing lengths together, so each sequence was interpolated to 60 steps (shown on the *x*-axis). For each sequence, the midi value of each note was determined before calculating the linear slope between each successive note in the sequence. Finally, the value of each point on the line was determined by its interpolation step number. As can be seen, statements and questions differ in shape. Statements tend to rise slightly and then fall. Questions tend to decline slightly before rising. The reader can refer to Appendix A to see musical notation corresponding to individual statement and question melodies.

Several differences across spoken and sung targets were in line with typical differences between song and speech. In contrast to the variability of syllable timing and variability of pitch within spoken syllables, sung melodies featured pitches that were produced isochronously and discretely within each syllable. In addition, sung pitches were designed to suggest a clear tonal center within a major key. The same vocalists who spoke the sentences also sang the melodies.

We then used Praat (Boersma & Weenink, 2009) to equate the overall duration of speech and song stimuli that were matched with respect to text and contour. The original relative timing of individual syllables was preserved for each sequence (e.g., the longest syllable before duration equating was also the longest syllable after). Our duration-equating procedure ensured that the timing of speech sequences did not change enough to make the sentences sound unusually unnatural. A rating study carried out on these stimuli (reported in Mantell & Pfordresher, 2013) demonstrated that the duration-equated stimuli could clearly be identified as either speech or song, despite the fact that songs were originally produced at a slower rate than speech. Also, even though the stimuli for each domain do not capture all aspects of their domain (e.g., only isochronous and diatonic melodies were used), people still perceive them as belonging to their respective domains.

Apparatus

Participants sat on a stool in a sound-attenuated booth (Whisper Room Inc., *SE* 2000 Series, Morristown, TN) for recordings. They heard targets, auditory feedback, and instructions over Sennheiser HD 280 Pro headphones at a comfortable loudness level. A Shure PG58 microphone connected to a Lexicon Omega preamplifier collected recordings at a sample rate of 22,050 Hz. Recordings were stored as .wav files for analysis. The same apparatus was used to make recordings of the stimuli.

Design and Procedure

A mixed-model 2 (practice domain) × 2 (text change) × 2 (contour change) × 2 (domain change) design was used to structure experimental sessions and for analyses of transfer. Practice domain was a between-participants variable; participants practiced either speech or song targets based on random assignment. Change to text, contour, and domain during transfer trials (each is a binary variable indicating that the feature was changed or held constant) were within-participants factors. Crossing within-participants factors led to eight conditions that were presented in different blocks of trials. For instance, if the practice text was *he ate it all*, for the text to stay the same in the transfer trials, a participant would imitate a target with the text *he ate it all*. Given a text change manipulation, the participant would imitate a different text such as *she bought apples*. Figure 2 illustrates examples of conditions representing transfer of domain (song/speech) and contour (statement/question).

Before they started imitating targets in the experimental trials, participants completed several warm-up tasks that involved reading a short passage of prose, singing the "happy birthday" song, producing vocal sweeps, and producing steady tone pitches in a comfortable pitch range. Warm-ups served to acclimatize the participant to the recording environment, set recording levels, and prepare the participant's voice.

Experimental imitation trials began at the conclusion of the warm-up period. Male participants imitated target stimuli from male productions and female participants imitated target stimuli from female productions. Each trial started with the presentation of a target stimulus followed by a short noise burst. Participants were instructed to imitate the pitch and timing of the target stimulus as accurately as possible after the noise burst.

Trials were arranged into eight blocks, each block comprising 12 imitation trials, for a total of 96 trials in the experiment. Within a block, participants first imitated one target for eight trials in succession. These were practice trials and provided a baseline measure of imitation accuracy for assessment of transfer. Immediately after the eighth practice trial in a block, participants went on to complete four more transfer imitation trials. On one of the eight blocks, the transfer target was identical to the practice target, whereas the transfer target differed from the practice target with respect to domain, contour, and/or text on the remaining seven blocks. The order of blocks was randomized across participants. The exact sequence used as a target for the practice and transfer trials in a block was randomized without replacement. For example, if a female participant was assigned to the condition practicing question contoured songs, the practice target for a block could be any one of the question contours for females shown in Appendix A.

[1] Stimuli can be obtained from Peter Pfordresher at pqp@buffalo.edu.

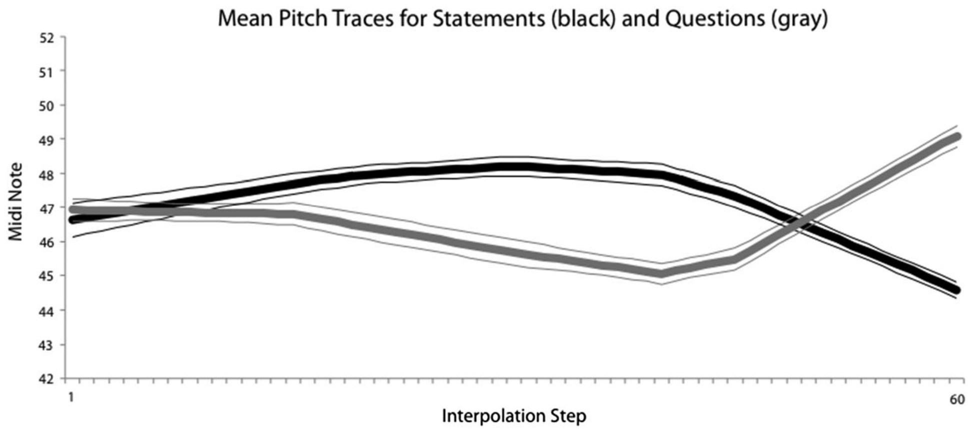

Figure 1. Mean pitch traces for questions and statements averaged across gender (after shifting female traces down 1 octave) and interpolated to 60 points. The thick lines show the mean traces, and thin lines show standard error. Black lines correspond to statements, and gray lines correspond to questions.

Data Analysis

Praat (Boersma & Weenink, 2009) was used to extract samples of vocal frequency over entire sequences to create F0 vectors from recordings of each stimulus target and production. An analysis-by-synthesis algorithm facilitated with Praat scripts (Boersma & Weenink, 2009) was used to convert the pitch-time trajectories in the raw recordings into numerical vectors. This technique allowed the experimenter to review aurally each vocalization and the corresponding pitch-time vector to determine the accuracy of the pitch extraction and correct the pitch-time vector if necessary. Importantly, this procedure afforded the experimenter an opportunity to fix pitch extraction errors from Praat. All of the errors corrected by the experimenter were errors that resulted from converting the original recording to a set of F0 values. These corrections did not involve judging accuracy of imitations relative to targets (which were conducted using automated scripts, described next). Experimenter bias was not a concern, as little could be done in analysis to alter results. After an appropriate pitch vector was selected, it was written as a text file and saved for future use.

The F0 vectors of imitations were compared with the F0 vectors of the targets to assess accuracy. In contrast to other research that has used the transfer paradigm, we focused on the accuracy and precision of pitch imitation rather than the rate at which sequences were produced. Our focus is based on the goal of the task, which was to imitate pitch accurately rather than to reproduce the sequence as rapidly as possible. Custom MATLAB (The MathWorks, Inc., Natick, MA) scripts were used to temporally align the beginning of production within each imitation–target trial pair so that our sequence-length pitch accuracy measures could be applied (detailed below). The scripts also equated the imitation–target pair durations by resampling the imitation duration to match the target stimulus. After temporally aligning the pairs, we adjusted pitch outliers (defined as sampled F0 values that fall into different octave bands than the samples adjacent to them), by shifting the octave of the sampled F0 the octave nearest to adjacent values (linear interpolation was used when necessary). Outliers were identified on fewer than 3% of all sampled F0 values. Pitch outliers of this sort typically reflect artifacts in extracted F0 (often based on difficulty assigning F0 to consonants) and can thus have an unwarranted influence on data analyses.

For pitch accuracy, we used *mean absolute pitch error:* the mean absolute value of difference scores between sampled F0 for target and imitation across the entire sequence. Analyzing mean absolute pitch error allowed for both sharp and flat productions within the same imitation to contribute to the overall error. We also analyzed *pitch correlations:* the correlation between the pitch contour of the stimulus target and the imitation. Pitch correlations measure how well the imitation tracks the pattern of change in pitch over time with respect to relative pitch. For instance, it is theoretically possible for an imitator to be consistently flat, but have a perfect correlation with the stimulus target's contour. Both measures are sensitive to the accuracy and precision with which pitch is imitated (cf. Pfordresher et al., 2010), although for the sake of brevity, we will refer to these analyses as measuring accuracy in the rest of this article. It is important to note that our pitch analyses measure accuracy across the entire production sequence, rather than converting dynamic F0 within each syllable to a single point estimate (e.g., the mean pitch of a syllable). Our measures of imitation accuracy are therefore sensitive to fine-grained fluctuations in F0 that may be important for the imitation of speech in particular.

Results

Transfer Effects

We first consider the evolution of accuracy across practice and transfer trials, to gain a descriptive understanding of performance within each type of block. Figure 3 contains plots of mean absolute pitch error and pitch correlation results for practice and transfer trials, and for participants that practiced either sung (Figure 3A and 3C) or spoken (Figure 3B and 3D) targets. Transfer trials (blocks 9–12) are plotted according to whether or not participants switched domain and/or contour. The factor text is omitted from plots because it did not contribute significantly in any analyses.

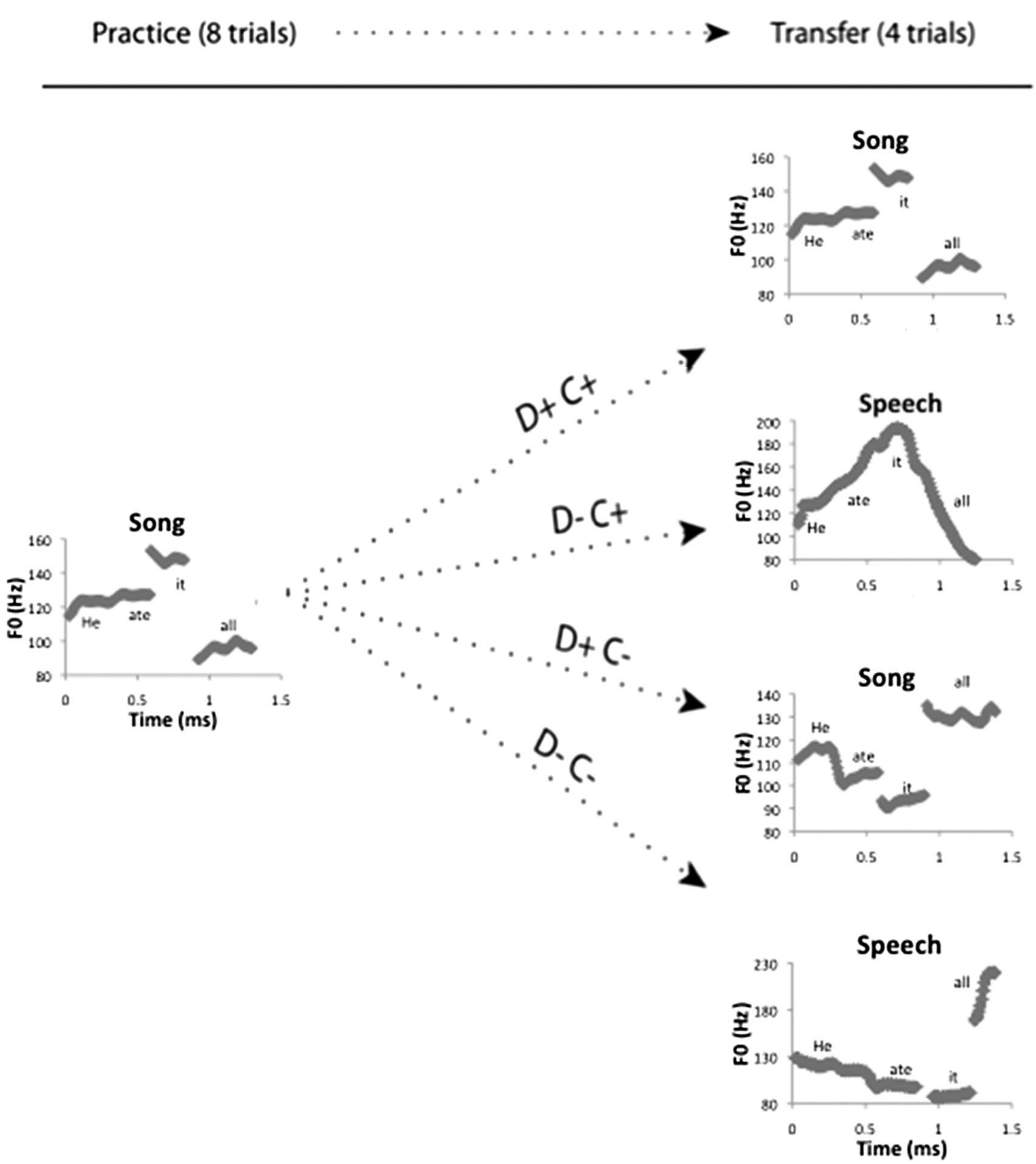

Figure 2. Block structure (upper panel). Illustration of transfer trial conditions (lower) that share the same text with the practice target. Target stimuli are shown as time by F0 plots. In this example, eight practice blocks comprise the sung target "he ate it all." Transfer trials can represent either the same (D+) or different domain (D−) and the same (C+) or different (C−) contour as the practice trials. Statements have a falling contour, and questions have a rising contour.

A somewhat surprising outcome shown in Figure 2 is that participants did not improve during "practice" trials. We assessed this through mixed 2 (practice domain) × 8 (practice trial) ANOVAs on measures of accuracy (mean absolute pitch error and pitch correlation). Neither ANOVA yielded a main effect of practice trial ($F < 2$, $p > .20$ in each case). The ANOVA on absolute error during practice trials did show a significant effect of practice domain, $F(1, 14) = 4.74$, $p < .05$, $\eta_p^2 = .25$, owing to absolute error being lower for song practice than for speech practice. However, there was no main effect of practice domain on pitch correlations, and no Practice Domain × Practice Trial interaction for either ANOVA, $F < 2.00$, $p > .05$.

Though somewhat surprising, given performance on practice trials in other research using the transfer paradigm (e.g., Palmer & Meyer, 2000), the lack of improvement is in line with recent research concerning pitch matching that has failed to show performance improvements over multiple imitations (Hutchins & Peretz, 2012). More importantly, our primary focus here is the degree to which imitation of a sequence during practice

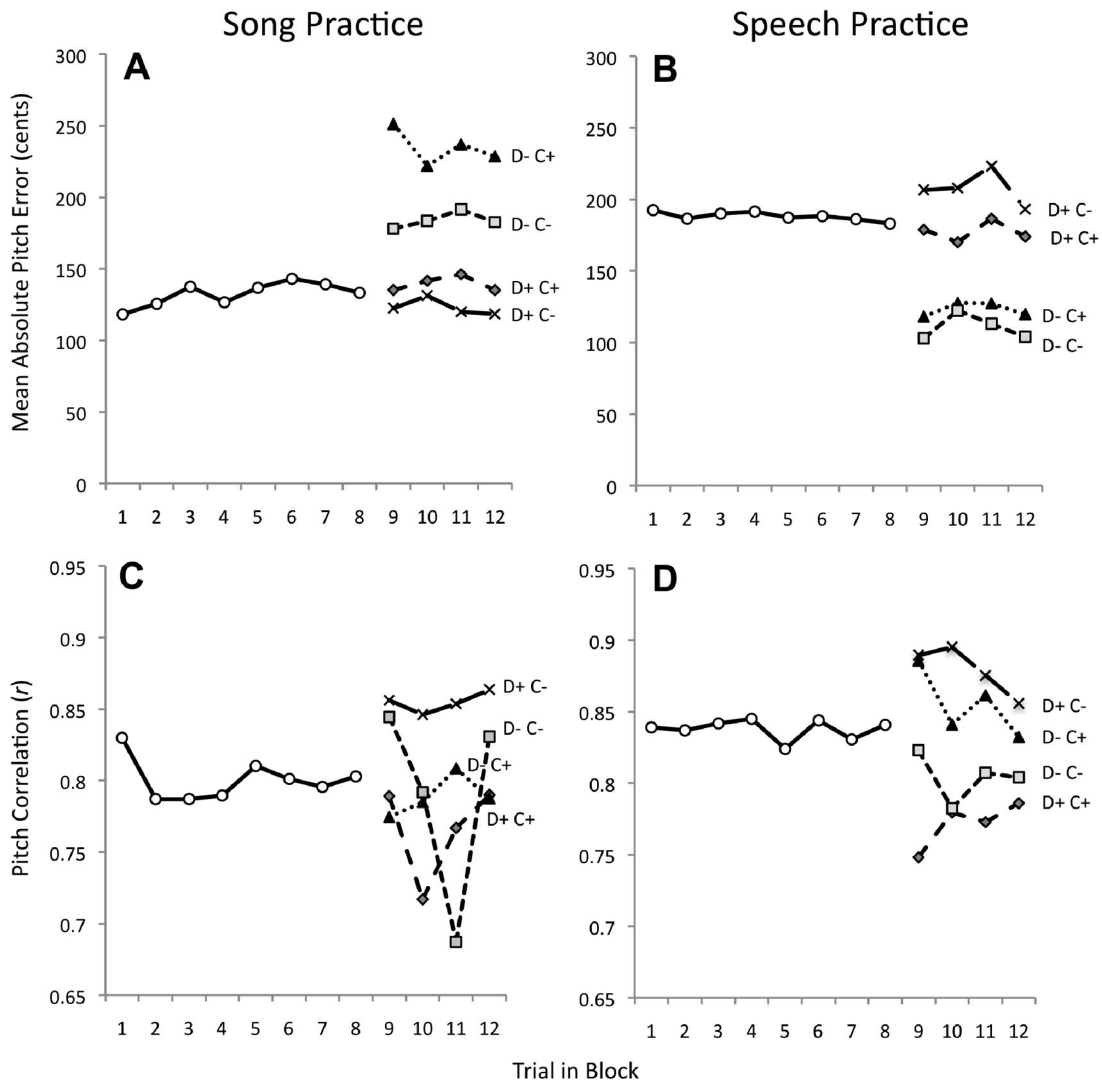

Figure 3. (A–D) The mean absolute pitch error (Panels A and B) and pitch correlations (Panels C and D) for each trial within blocks and under the conditions for which domain (D) and contour (C) were the same (+) or different (−) from practice are shown. The text manipulation is not shown.

influences performance on transfer trials. Practice influenced performance on transfer trials in several respects. Interestingly, transfer effects did not clearly support either a fully independent or a fully integrated model, thus arguing for a hybrid architecture. We qualitatively describe transfer effects here and report statistical analyses in the next section.

The strongest effects of practice on transfer trials were borne out in analyses of mean absolute pitch error (Figure 3A and 3B). When people practiced imitating a song and transferred to song, imitation in transfer remained consistently accurate regardless of whether or not the contour changed. However, when participants switched to speech during transfer, performance deteriorated and was influenced by contour. Somewhat surprisingly, performance was worst when the domain switched from song to speech (e.g., a sung statement switched to a spoken statement), but the contour remained the same (e.g., sung and spoken examples were both statements or were both questions). When participants imitated speech in practice, imitation remained consistently accurate during transfer to speech, though performance deteriorated somewhat if spoken contour shifted. Unlike practice with song, performance during transfer improved when people switched from speech to song. Participants achieved levels of performance that were even better than when they sang throughout all trials (cf. the D + C+ transfer condition in the top left panel). Although accuracy did seem to be dependent on domain, which could be interpreted as consistent with independent pitch processing, the effects of contour transferring across domains is harder to explain with such an account. In contrast to transfer effects on mean absolute pitch error, results for the pitch correlation measure were less reliable but still suggest complex interactions between the factors of domain and contour.

Mean absolute pitch error. Difference scores were calculated by subtracting the mean performance across all four transfer trials in each block from the mean performance on the last four trials of practice in that block.[2] For the mean absolute pitch error measure, lower error in transfer trials would lead to positive values, and greater error would lead to negative values. Thus, positive transfer leads to higher values of the difference score; the absence of transfer leads to a difference score of 0; and negative transfer leads to negatively signed difference scores. A 2 (practice domain) × 2 (text change) × 2 (contour change) × 2 (domain change) mixed ANOVA was performed on the difference scores. A main effect of practice domain was found, $F(1, 14) = 7.53$, $p < .05$, $\eta_p^2 = .35$, such that those who practiced speech tended to perform better in transfer than in practice ($M = 75.78$, $SE = 29.07$), and those who practiced song tended to perform worse in transfer than in practice ($M = -56.46$, $SE = 29.07$). The ANOVA also revealed a significant Practice Domain × Domain Change interaction (see Figure 4), $F(1, 14) = 12.52$, $p < .01$, $\eta_p^2 = .47$. This interaction simply reflects an overall advantage for imitating song over speech with respect to mean absolute pitch error. When domain does not change from training to transfer (although other sequence features might), performance remains unchanged. However, if one has practiced imitating a song and switches to speech, transfer performance suffers (white bar to the left) whereas the reverse order leads to improvement during transfer (white bar to the right). In other words, switching to song leads to better performance, but switching away from song deteriorates performance. However, not all effects were reducible to a simple song advantage—which could be taken to support an independent view (such as Peretz & Coltheart, 2003)—as we discuss next.

Change to the domain and contour during transfer trials also influenced performance, in a way that was independent of an overall advantage for imitating song. There was a significant Contour Change × Domain Change interaction, $F(1, 14) = 10.40$, $p < .01$, $\eta_p^2 = .43$,[3,4] which is plotted in Figure 5. Difference scores suggest nonadditive effects of contour change and domain change. Whereas negative transfer was found when either the contour was changed or the domain was changed in isolation, performance during transfer actually improved when both features of the transfer sequence were changed. Thus, it appears that performance in transfer was facilitated when the transfer sequence was maximally distinctive from the practiced sequence, but suffered when the transfer sequence was similar in some respects but not others. A Tukey's HSD test ($p < .05$) was performed on means shown in Figure 5, revealing that both a contour change (M difference $= 74.07$, $SE = 33.78$) and domain change alone (M difference $= 69.90$, $SE = 30.01$), led to worse performance than when contour and domain both changed from practice to transfer trials. The finding that having the same contour in transfer as in practice had an affect on transfer performance when domain changed (facilitated if contour was different, detrimental if contour was the same) suggests that some aspect of pitch processing was shared for both imitations. In order for the dynamics of pitch in the practice target to matter, there would need to be something learned about pitch that carried over to transfer imitations across domain. At the same time, the considerable effect of simply altering the

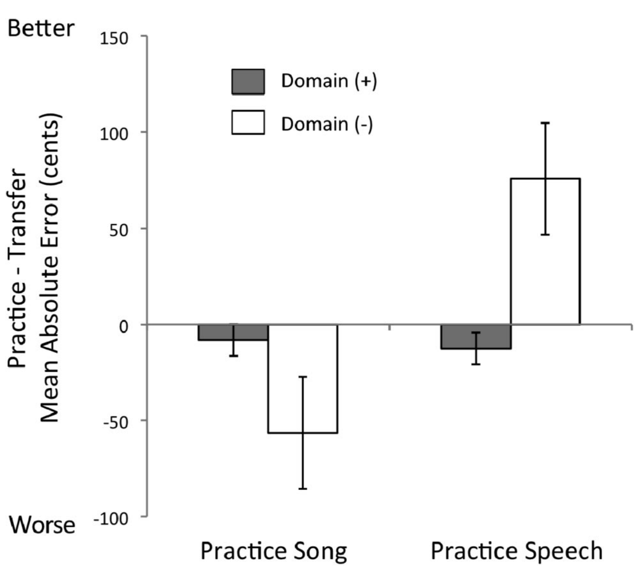

Figure 4. Plot of the Practice Domain × Domain Change interaction. Error bars depict the standard error of the mean.

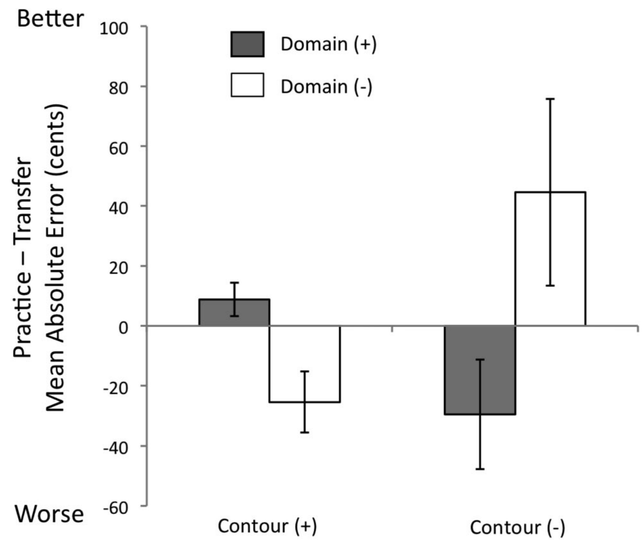

Figure 5. Plot of the Contour Change × Domain Change interaction. Error bars depict one standard error of the mean.

[2] Alternative analyses using difference scores calculated by subtracting the first trial of transfer from the last trial of practice, the four trials of transfer from the last trial of practice, and the four trials of transfer from mean performance on all practice trials yielded similar results.

[3] Further analyses revealed that the Domain Change × Contour Change interaction was still significant for trials in which text did not change from practice to transfer, $p < .05$. The same pattern of means was seen.

[4] Contour complexity was randomly distributed throughout conditions because sequences of a particular contour were selected randomly from Appendix A. In the stimulus set, however, contour was more complex for questions than for statements. We ran a post hoc ANOVA on blocks in which the number of changes in contour direction were the same for practice and transfer (using imputation procedures for missing data). The Domain Change × Contour Change interaction was statistically significant, $F(1, 15) = 10.21$, $p = .006$, and that means were qualitatively similar to the analysis reported in the ANOVA run on difference scores for all blocks. This suggests that effects of contour in the current study were due to changing contour rather than changing complexity. There was also a significant main effect of contour change, $F(1, 15) = 5.78$, $p = .03$.

domain associated with transfer trials goes against the assumptions of a fully integrated architecture, as described in the introduction.

Pitch correlations. Difference scores were also calculated for the pitch correlation data in the same manner as for mean absolute pitch error, and were analyzed in the same way. The ANOVA revealed no significant differences between the difference scores for pitch correlations, yielding no evidence that practice trials had differential effects on transfer trials in regards to pitch correlation across conditions. However, it is noteworthy that the advantage for accuracy in imitations of song over speech was not seen in the pitch correlation data as it was in the mean absolute pitch error data. Thus, the considerable song advantage found for imitation of absolute pitch may not exist in the imitation of relative pitch (Mantell & Pfordresher, 2013).

Song Advantage for Imitating Absolute Pitch

One of the questions that emerges from our analyses of difference scores is to what degree differences in transfer are simply related to an overall advantage for imitating song than imitating speech. This advantage is consistent with the view that the imitation of song is facilitated by tonal encoding, whereas imitation of speech is not (Peretz, 2009; Peretz & Coltheart, 2003). We focused first on difference scores as a measure of performance in transfer relative to practice. However, difference scores are inherently ambiguous with respect to which term dominates the effects that are observed (minuend or subtrahend). Thus, we now turn to results of an analysis designed to address a possible song advantage during transfer by focusing specifically on performance during transfer trials (averaged across all four trials). In this context, a song advantage yields a significant main effect of the domain used during transfer, and qualifications of this advantage (of greatest interest here) lead to other effects. Because no advantage for song was found for measures of pitch correlation, we focus on mean absolute pitch error.

Mean absolute pitch error scores during transfer trials were submitted to a mixed-model ANOVA, with the between-participants factor being domain during practice (song or speech) and the within-participants factors being domain during transfer, text change, and contour change. The ANOVA led to a significant main effect of domain during transfer, $F(1, 14) = 73.33$, $p < .01$, $\eta_p^2 = .84$, which supports the overall advantage for imitating song (M error $= 126$ cents, $SE = 8$) as opposed to speech ($M = 271$ cents, $SE = 13$). In addition, there was a significant Practice Domain × Transfer Domain × Contour Change interaction, $F(1, 14) = 8.64$, $p < .05$, $\eta_p^2 = .38$, which is plotted in Figure 6. As can be seen, imitations of song during transfer, in addition to yielding lower overall error scores than when participants imitated speech, were also unaffected by characteristics of trials during practice (domain or contour). Error scores for the imitation of speech were higher overall than when people sang during transfer. More importantly, error rose considerably when practice trials contained a sung contour that matched the contour being spoken during transfer. Thus, there is something about the perseveration of a sung contour that interferes with transfer to the imitation of speech. The asymmetry in transfer effects from music to language and from language to music mirrors findings from investigations of long-term music and language learning on perception where music training seems to have a larger effect on language processing than language has on music processing (Kraus & Chandrasekaran, 2010; Krishnan, Gandour, & Cariani, 2009). This result also speaks to the possibility that the salience of sung contours during practice led to perseverations of these contours during transfer, an issue we address directly in the next section.

Global Contour Perseveration

Earlier analyses showed that perseveration of contour interferes with transfer from imitation of song to the imitation of speech. This result suggests that the salience of a sung contour makes it difficult to make subtle adjustments to that contour when transferring to speech, whereas more dramatic changes to the pitch contour may, somewhat surprisingly, be easier to negotiate. Based on these observations, we hypothesized that participants may exhibit a tendency to perseverate a sung pitch contour when transferring to a spoken contour of the same type. We developed a new measure to operationalize the degree to which participants perseverated the pitch content of practice during transfer trials, which we

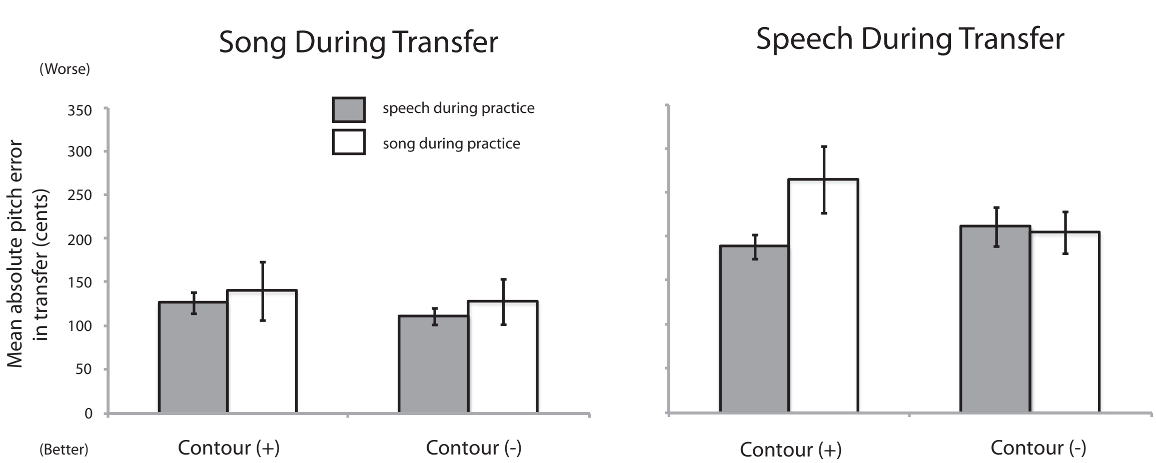

Figure 6. Plots of mean absolute error in transfer for instances when contour did or did not change and song or speech was imitated in transfer. Error bars show the standard error of the mean.

refer to as the *normalized distinctiveness* (*Ndistinct*) of an imitative performance during transfer from the target stimulus during practice (Eq. (1)).

$$Ndistinct = \frac{1/n\sum\left|I_{transfer} - T_{practice}\right| - 1/n\sum\left|I_{transfer} - T_{transfer}\right|}{1/n\sum\left|T_{transfer} - T_{practice}\right|} \quad (1)$$

In Eq. (1), *I(transfer)* refers to the vector of produced F0 values from imitative performance during transfer trials, *T(transfer)* refers to the vector of F0 values in the target stimulus during transfer trials, and *T(practice)* refers to the vector of F0 values in the target stimulus during practice trials. Each term in the equation constitutes the mean absolute difference between two vectors of F0 values, standardized with respect to total duration. The first term of the numerator measures how closely produced pitch patterns during transfer trials match F0 values from the target that was imitated during prior practice trials. The second term of the numerator measures mean absolute error during transfer. If the participant has successfully transferred to the new stimulus, the first term should exceed the value of the second term, leading to a positive score. Lower scores can suggest that the participant has still "held on" to the previous target, leading to a lower error score in the first term than the second term. The denominator standardizes values of the numerator according to differences between the two targets, which can vary in distinctiveness from each other. A detailed example of how this normalized distinctiveness was calculated is described in Appendix B.

Because this measure is undefined for blocks in which the target during transfer was identical to the target during practice (which have a denominator of 0.0), such conditions were discarded. Likewise, we did not include transfer of text in the ANOVA because that factor would not be balanced. Normalized distinctiveness scores were analyzed with a 2 (practice domain) × 2 (transfer domain) × 2 (contour change) mixed-model ANOVA (practice domain being the lone between-participants factor). There was a significant Practice Domain × Contour Change interaction, $F(1, 14) = 5.304$, $p < .05$, $\eta_p^2 = .27$, and a significant Transfer Domain × Contour Change interaction, $F(1, 14) = 7.069$, $p < .05$, $\eta_p^2 = .34$.

Figure 7A displays the Practice Domain × Contour Change interaction. When participants sang during practice trials and contour remained the same, distinctiveness from practice to transfer was lower than in other conditions. This effect is consistent with the view that participants tended to perseverate sung contours more so than spoken contours (leading to lower normalized distinctiveness). The Transfer Domain × Contour Change interaction, plotted in Figure 7B, demonstrates the obverse of this effect by showing greater distinctiveness for sung imitations during transfer, particularly when sung imitations featured a contour that differed from the one presented during practice. Both effects speak to the salience of contour for song as opposed to speech, leading to a tendency to perseverate sung contours that had been present during practice, and facilitation of switching to a new sung contour during transfer.

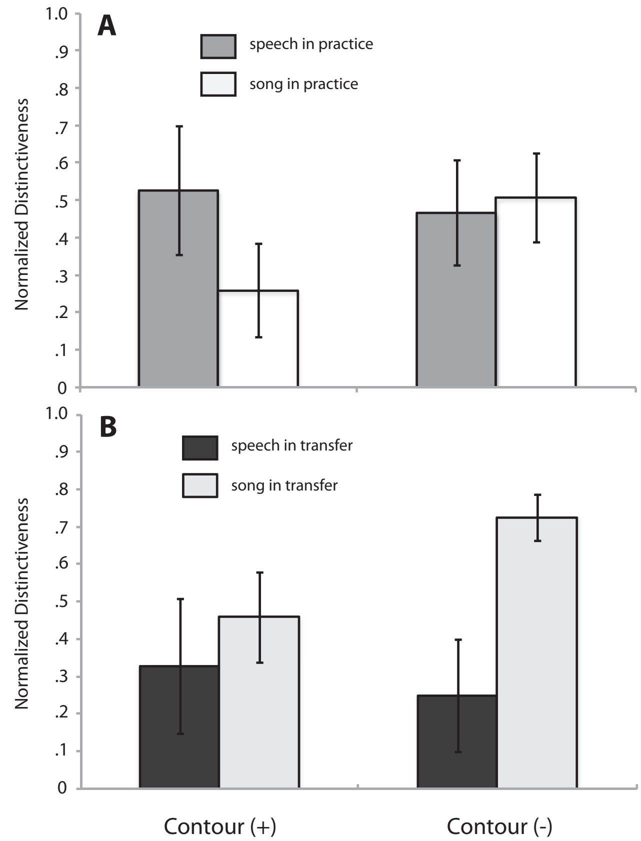

Figure 7. Plots the (A) Practice Domain × Contour Change interaction and (B) Transfer Domain × Contour Change interaction for normalized distinctiveness. Error bars show the standard error of the mean.

Individual Differences

As mentioned in the introduction, we were interested in transfer performance both at the group and individual levels. Previous work suggests individuals vary considerably with respect to the vocal imitation of sung pitch (Berkowska & Dalla Bella, 2009; Dalla Bella, Berkowska, & Sowinski, 2011; Pfordresher, 2011; Pfordresher & Brown, 2007; Pfordresher & Mantell, 2009; Pfordresher et al., 2010; Welch, 1979), and further evidence suggests that such "poor-pitch" singers may have difficulty transferring from one sequence to anther, given that they exhibit a tendency (not seen in accurate singers) to make pitch errors that drift in the direction of their own comfortable pitch range (Pfordresher & Brown, 2007). Poor-pitch singers may have similar difficulty imitating speech (Mantell & Pfordresher, 2013), in line with integrationist views of music and language. To see whether individual differences in imitation abilities were related to transfer effects and the song advantage reported above, we calculated Pearson correlation coefficients between average measures of imitation accuracy across all 96 trials (mean absolute error), designed to measure overall pitch imitation accuracy, with measures of the effectiveness of transfer, described below.

First, we addressed how overall accuracy covaried with the degree to which participants exhibited a song advantage during transfer trials, given that an advantage for imitating song played a large role in the Practice Domain × Domain Change interaction found for mean absolute pitch error difference scores. For each participant, we subtracted the mean absolute pitch error for all imitations of song during transfer from the mean absolute pitch error across all imitations of speech during transfer. These difference scores are positive when participants exhibit a song advantage. No individual had a negative value for this variable, although there were substantial differences between individuals in the strength of this advantage. Overall mean absolute pitch error was significantly correlated with the magnitude of this difference score across participants, $r(15) = .747$, $p < .001$ (Figure 8A). The song advantage increased with increasing overall error; poor-pitch singers thus exhibit a larger song advantage than accurate singers. The fact that this relationship varies across a continuum of singing accuracy (meaured by *y*-axis values) suggests that this effect cannot simply be based on a ceiling effect among accurate imitators. It is also worth mentioning that mean absolute pitch error scores in general were positively correlated across music and speech imitation trials during transfer trials, $r(15) = .928$, $p < .001$, as has been found previously (Mantell & Pfordresher, 2013; Pfordresher & Mantell, 2009).

To see whether there was also a relationship between individuals' pitch imitation ability and the observed Contour Change × Domain Change interaction for mean absolute error difference scores, correlations were calculated for individuals' difference scores for the conditions shown in Figure 5 (D+ C+, D+ C−, D− C+, and D− C−). Only one of the four relationships was significant. Overall mean absolute error was significantly correlated with the difference scores for the D+ C− condition, $r(15) = -.648$, $p = .007$ (Figure 8B). The direction of correlation was such that less accuracy in imitation ability meant a greater decrement in performance when contour was switched alone.

These correlations suggest that worse imitators are affected more by changes to the stimulus, whereas good imitators tend to remain equally accurate from practice to transfer trials. Specifically, when the domain or contour changes, the accuracy of imitations for poor singers is affected more than for accurate singers. To see whether individual imitation ability was related to the ability to deal with change to the target stimulus, we calculated correlations between mean absolute pitch error in transfer with absolute pitch errors relative to the target imitated during practice trials. This is the first term in the numerator of Eq. (1). We computed correlations using this term, rather than full normalized distinctiveness, for two reasons. First, because the X variable in correlations comes from the second term of the numerator (i.e., error in imitation), individual differences with respect to normalized distinctiveness scores are inherently confounded with individual differences in the X variable. Second, because this analysis averages across conditions, the concerns leading to the denominator of Eq. (1) are not problematic. Note that error scores relative to practice targets should be interpreted as showing how distinct one's imitation in transfer was from the previously practiced target, with lower scores indicating less change from the practiced target.

Figure 9 shows the relationship between mean absolute pitch error during transfer (X) and error in transfer relative to the practice target (Y), along with the best fitting regression line (light diagonal line). In addition, the dark diagonal line shows unity; values on this line indicate performance that is no more distinct from transfer targets than to practice targets. Values to the left of this line suggest good transfer; that is, performances that are more similar to the transfer target than to the practice target.[5]

Overall error scores during transfer (X) were positively correlated with errors relative to practice target, $r(15) = .72$. However, this relationship is underadditive; the best-fitting regression line yielded a slope of $b = .48$, which was significantly lower than a slope of 1.00, $t(15) = 2.67$, $p < .05$. The significance of this underadditivity is that poor-pitch singers (those with higher *x*-axis values) are influenced by a tendency to perseverate pitch patterns from practice trials, leading to a compressed relationship between overall accuracy and distinctiveness from practice targets.

Discussion

This study, to our knowledge, is the first to explore how imitation practice affects the ability to imitate the pitch of sequences that differ from practice sequences in regard to domain (speech or song). Transfer performance was best when both contour and domain changed or when neither changed in transfer trials. When only one factor changed performance during transfer suffered. This effect was mainly driven by contour perseveration from practiced song to speech imitation in transfer. It is also noteworthy that changes to the text of sequences (related to finer-grained frequency information) did not influence transfer.

Analysis of individual differences showed that individuals who were poorest at imitating pitch had greater difficulty switching between practice and transfer. Worse imitators tended to show a larger difference in the ability to imitate speech and song. They also showed a greater detriment to performance when contour changed within a domain, and less deviation from the practiced target when the stimulus differed in any respect (domain, contour, or text) from what was practiced.

Independence Versus Integration of Pitch Processing

This study was designed to investigate independence versus integration of pitch processing with the assumptions that (1) if the mechanisms involved in imitating the pitch of speech and song were independent, we would find a large main effect of domain that does not interact with other factors; and (2) if such mechanisms were integrated, then the more dissimilar the transfer target relative to the practice target (the more variables that change) the less similar performance should be in transfer.

In contrast to predictions based on independence, no difference score analyses yielded a main effect attributable to change of domain. This factor was associated with more complex interactions that are difficult to reconcile within an independent-mechanisms framework. At the same time, there were effects of domain that run against predictions based solely on integration. Like the prediction of independence, an integration view would predict additive effects based on the number of factors changing from practice to transfer, with a smaller role (effect size) related to change of domain. The interactions found here likewise are problematic for a complete integration view. Instead, the current data suggest two factors at play: enhanced salience of sung pitch contour as opposed to spoken pitch contour, and a greater tendency to perseverate the contour of sung pitch during transfer.

[5] The mean absolute difference between F0 vectors of targets during transfer and targets during practice was 282 cents. This value represents the error relative to the practice target that would occur, on average, if one imitated transfer targets perfectly.

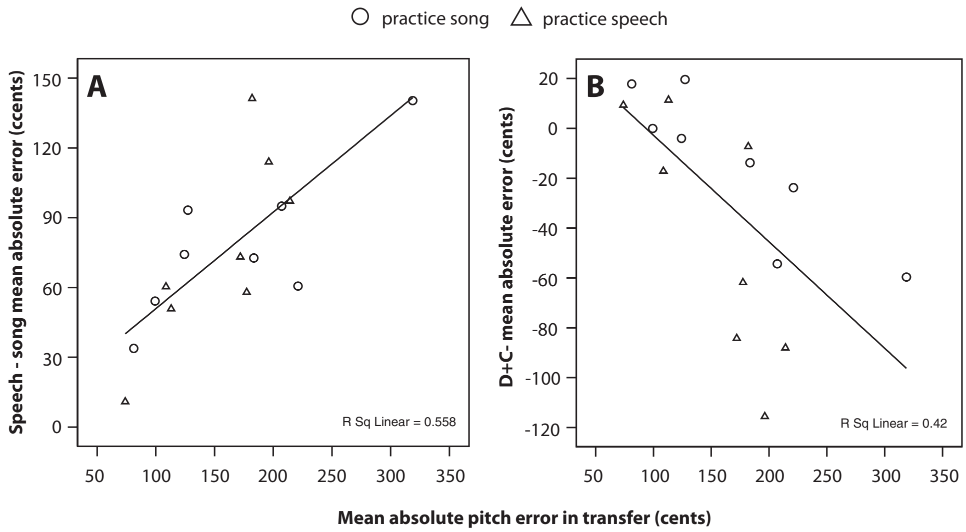

Figure 8. Scatter plots of individuals' mean absolute error for all trials in the experiment (*x*-axes) related to the difference between mean absolute error of song and speech imitations in transfer (A), and the D+ C− condition (B). Individuals are labeled by the domain in which they practiced. The best linear fits of the data are shown.

Various results argue for overall greater salience of sung pitch than spoken pitch. First, there was an overall tendency for greater accuracy in production both during practice and transfer for singing than speaking. Furthermore, transfer to song led to better imitation regardless of what was imitated during practice. These findings are consistent with the Peretz and Coltheart (2003) model suggesting that tonal encoding of pitch in music enhances pitch representations over those associated with language. We controlled for rate between domains, but allowed speech and song to differ in ways characteristic of naturalistic stimuli. As such, our findings could be related to the fact that song contour naturally fluctuates less than speech contour (Stegemöller, Skoe, Nicol, Warrier, & Kraus, 2008), or that the songs we used were isochronous, diatonic, tonal, and in a major key. Future work is needed to reveal which characteristics of song make it easier to imitate.

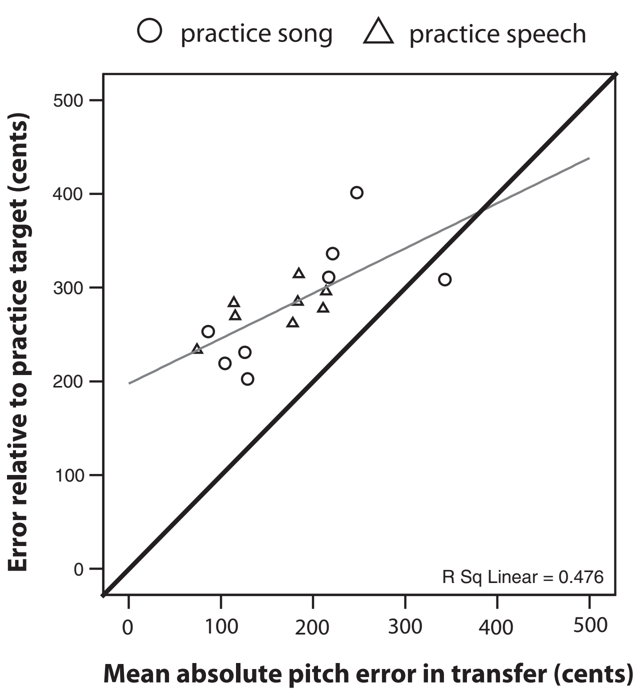

Figure 9. Scatterplot of error relative to the practice target related to mean absolute error in transfer. Individuals are labeled with respect to the domain that was practiced. The best linear fit of the data is shown by the light line. The dark diagonal line displays unity. Values to the left of this line indicate imitations in transfer that were more similar to the transfer than the practice target.

A "song advantage" explanation cannot account for all findings. A simple yet important result is that imitation in transfer across all trials was more accurate when people imitated speech during practice. Thus, the contrast of any transfer trial with previous imitation of speech had a generally facilitating effect regardless of domain, a result that cannot be explained with a benefit for singing over speaking. Practice imitations of song had a negative effect on speech imitations during transfer in some cases. This short-term transfer effect is the opposite of what has been reported over the long term, namely beneficial transfer of prior musical training to speech perception (Kraus & Chandrasekaran, 2010). Furthermore, interactions qualify the benefit of speech in important ways. In particular, the fact that transfer from song to speech while preserving contours led to worse performance than conditions where the contour changed is difficult to reconcile with a view in which pitch processing is carried out independently for domains.

Improvements within practice trials were not found. Thus, the role of practice trials in the current study was in their influence on subsequent trials, within and across domains. Various priming paradigms have pointed to the role of persistence in speech and music production (Bock, 1986; Jungers, Palmer, & Speer, 2002; Zurbriggen, Fontenot, & Meyer, 2006). For instance, Jungers et al. (2002) found that tempos can carry over from one produced keyboard melody to the next. In a vocal production task, Zurbriggen et al. (2006) observed that when expert singers were primed with a practice melody, their performance on a following melody was facilitated (faster and less errorful) if that melody was an exact transposition of the prime. Thus, contour in that study seemed to have a beneficial impact on production within the song domain. The current data suggest asymmetric persistence effects across domains. These effects may arise because pitch patterns from tonal song have a greater tendency to persist in memory, whereas spoken pitch patterns are difficult to encode (but see Deutsch, Henthorn, & Dolson, 2004, for salience of pitch in tone languages). As a result, when presented with the more complex spoken pitch patterns, participants are more likely to revert to the more salient sung pitch pattern.

Ultimately we think the results suggest a view that takes into account the overall salience of pitch patterns, with sung pitch patterns being more salient by virtue of their resonance with stored pitch categories (tonal encoding, cf. Peretz & Coltheart, 2003), and stability of pitch within canonical rhythmic units (notes vs. syllables). Such a modified integrationist approach, we think, can better explain the second major factor at play in the data: tendency to perseverate sung pitch more so than spoken pitch.

Poor-Pitch Singing

Analyses of individual differences yielded some important findings for theories of vocal imitation as well. Individuals differ with respect to their ability to replicate a melody by singing, and an estimated 15% of adults consistently mistune pitches by singing more than a semitone off pitch. These individuals have been labeled "poor-pitch" singers (cf. Dalla Bella, Giguere, & Peretz, 2007; Pfordresher & Brown, 2007; Welch, 1979). This deficit appears to be specific to the imitation of pitch, in that poor-pitch singers do not usually exhibit deficits in pitch perception or vocal motor control (Bradshaw & McHenry, 2005; Dalla Bella et al., 2007; Pfordresher & Brown, 2007) yet show similar deficits when imitating pitch in speech (Mantell & Pfordresher, 2013; Pfordresher & Mantell, 2009). Thus, it has been proposed that poor-pitch singing may be a problem of the vocal imitation system, reflecting a deficit in the internal modeling of sensorimotor relationships in the voice (Pfordresher, 2011; Pfordresher & Halpern, in press). As such, the current data—which bear on one's ability to flexibly shift vocal imitation across sequences—are relevant to our understanding of this vocal imitation deficit.

Recent research suggests that poor-pitch singers are characterized by inflexibility in vocal imitation of pitch. First, whereas poor-pitch singers typically mistune pitch while singing, they do not mistune pitches scaled to the most comfortable range of their voice, and errors in singing among poor-pitch singers tend to "drift" back toward their most comfortable pitch range (Pfordresher & Brown, 2007). Second, poor-pitch singers tend to compress the size of pitch intervals while singing imitatively, even though they do not exhibit restricted range of pitch during nonimitative vocal tasks (Dalla Bella, Giguere, & Peretz, 2009; Pfordresher & Brown, 2007, 2009). Finally, poor-pitch singing is greatly alleviated when poor-pitch singers imitate recordings of themselves as opposed to recordings of other singers or of idealized targets (Pfordresher & Mantell, 2012).

The current design addresses flexibility in a new way, by measuring how well individuals can switch from one sequence to another across practice and transfer trials. Rather than dichotomize participants into two groups ("accurate" vs. "poor-pitch" imitators), we examined flexibility in transfer along a continuum of overall imitation accuracy. There was the tendency for less accurate imitators to exhibit a greater advantage for imitating song rather than speech. This accompanied an overall positive correlation between the domains of speech and song. Thus, poor-pitch singers may be able to incorporate the enhanced salience of sung pitch in a way that alleviates the difficulty they experience during vocal imitation. Second, across all trials, singers who exhibited greater inaccuracy while imitating also appeared less able to transfer, given that error scores relative to the transfer targets scaled underadditively with errors relative to previously imitated practice targets. This result suggests that poor-pitch singers find it difficult to adapt to new sensorimotor constraints during imitation.

Future Directions

All our melodies were diatonic, tonal, in a major key, isochronous, semantically neutral, and emotionally neutral. Therefore, there are limitations of our study that stem from a failure of the stimulus set to explore the speech to song continuum and capture characteristics that can be similar for speech and song other than contour and text. One similarity between speech and music is the ability for the minor third to communicate sadness in both (Curtis & Bharucha, 2010). It could be that emotional similarity of practice and transfer trials could impact transfer accuracy (see Patel, 2011, 2012), but we have no way of exploring such a hypothesis with the current dataset. Koelsch (2011) reviews several studies suggesting concepts primed by speech influence the extraction of meaning from music and vice versa. A conceptual transfer effect across domains might also be seen (e.g., MacKay & Bowman, 1969) using stimuli designed to assess such questions.

Several other possibly important dimensions for transfer varied randomly across our stimuli. For instance, some target melodies contained more than one instance of the tonic, giving a clearer sense of tonality than other targets, and some targets had more directional changes in contour than others, leading to greater contour complexity. It would be a daunting task to adequately examine transfer effects pertaining to all of the above mentioned variables in one study. Therefore, we opted to simplify our stimuli to focus on dimensions previously posited to be important in models of speech and music. However, other dimensions are suitable for the transfer paradigm and should be explored in the future to get a fuller picture of transfer effects within and across domains.

It has recently become popular to study transfer effects of long-term speech and/or music learning across domains, but that approach is limited with respect to how well researchers can investigate transfer of learning for finer grained information. For instance, tonality and contour complexity questions cannot be addressed with cross-sectional designs, as researchers cannot be sure of the exact nature of a participant's background. Future studies may find it useful to use the transfer paradigm in conjunction with cross-sectional studies. Such research may further inform model development and test predictions of models that find a balance between full independence and full integration.

References

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain, 125,* 238–251. doi:10.1093/brain/awf028

Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology, 5,* 69–83. doi:10.2478/v10053-008-0068-2

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. doi:10.1016/j.bandc.2011.07.006

Bock, K. J. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18,* 355–387. doi:10.1016/0010-0285(86)90004-6

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1) [Computer software]. Retrieved from http://www.praat.org/

Bradshaw, E., & McHenry, M. A. (2005). Pitch discrimination and pitch matching ability of adults who sing inaccurately. *Journal of Voice, 19,* 431–439. doi:10.1016/j.jvoice.2004.07.010

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: Bradford Books, MIT Press.

Brown, S. (2000). The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271–300). Cambridge, MA: The MIT Press.

Curtis, M. E., & Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion, 10,* 335–348. doi:10.1037/a0017928

Dalla Bella, S., Berkowska, M., & Sowinski, J. (2011). Disorders of pitch production in tone deafness. *Frontiers in Psychology, 2,* 164.

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. doi:10.1121/1.2427111

Dalla Bella, S., Giguere, J. F., & Peretz, I. (2009). Singing in congenital amusia. *Journal of the Acoustical Society of America, 126,* 414–424. doi:10.1121/1.3132504

Daltrozzo, J., & Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience, 21,* 1882–1892. doi:10.1162/jocn.2009.21113

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception, 21,* 339–356. doi:10.1525/mp.2004.21.3.339

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.

Giuliano, R. J., Pfordresher, P. Q., Stanley, E., Narayana, S., & Wicha, N. (2011). Native experience with a tone language enhances pitch discrimination and the speed of neural responses to pitch change. *Frontiers in Psychology, 2,* 146. doi:10.3389/fpsyg.2011.00146

Griffiths, T. D., Johnsrude, I., Dean, J. L., & Green, G. G. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *Neuroreport, 10,* 3825–3830. doi:10.1097/00001756-199912160-00019

Hove, M. J., Sutherland, M. E., & Krumhansl, C. L. (2010). Ethnicity effects in relative pitch. *Psychonomic Bulletin & Review, 17,* 310–316. doi:10.3758/PBR.17.3.310

Hutchins, S. M., & Peretz, I. (2012). A frog in your throat of in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology: General, 141,* 76–97. doi:10.1037/a0025064

Jungers, M. K., Palmer, C., & Speer, S. R. (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing,* 1–2, 21–35.

Koelsch, S. (2011). Toward a neural basis of music perception-a review and updated model. *Frontiers in Psychology, 2,* 110. doi:10.3389/fpsyg.2011.00110

Koelsch, S., & Siebel, W. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences, 9,* 578–584. doi:10.1016/j.tics.2005.10.001

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11,* 599–605. doi:10.1038/nrn2882

Krishnan, A., Gandour, J. T., & Cariani, P. A. (2009). Experience-dependent neural representation of dynamic pitch in the brainstem. *Neuroreport, 20,* 408–413. doi:10.1097/WNR.0b013e3283263000

MacKay, D. G., & Bowman, R. W. (1969). On producing the meaning in sentences. *The American Journal of Psychology, 82,* 23–39. doi:10.2307/1420605

Mantell, J. T., & Pfordresher, P. Q. (2010). Modular processing? Phonetic information facilitates speech and song imitation. In S. M. Demorest, S. J. Morrison, & P. S. Campbell (Eds.) *Proceedings of the 11th International Conference on Music Perception and Cognition* (pp. 338–339). Seattle, WA: University of Washington.

Mantell, J. T., & Pfordresher, P. Q. (2013). Vocal imitation of song and speech. *Cognition, 127*:177–202. doi:10.1016/j.cognition.2012.12.008

Marin, O. S. M., & Perry, D. W. (1999). Neurological aspects of music perception and performance. In D. Deutsch (Ed.), *The psychology of music* (pp. 653–724). Oxford: Elsevier. doi:10.1016/B978-012213564-4/50018-4

Meyer, R. K., & Palmer, C. (2003). Temporal and motor transfer in music performance. *Music Perception, 21,* 81–104. doi:10.1525/mp.2003.21.1.81

Palmer, C., & Meyer, R. K. (2000). Conceptual and motor learning in music performance. *Psychological Science, 11,* 63–68. doi:10.1111/1467-9280.00216

Parbery-Clark, A., Strait, D. L., & Kraus, N. (2011). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologica, 49,* 3338–3345.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119,* 2382–2393. doi:10.1121/1.2178720

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6,* 674–681. doi:10.1038/nn1082

Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology, 2,* 142. doi:10.3389/fpsyg.2011.00142

Patel, A. D. (2012). The OPERA hypothesis: Assumptions and clarifications. *Annals of the New York Academy of Sciences, 1252,* 124–128. doi:10.1111/j.1749-6632.2011.06426.x

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition, 87,* B35–B45. doi:10.1016/S0010-0277(02)00187-7

Patterson, R. D., Allerhand, M. H., & Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America, 98,* 1890–1894. doi:10.1121/1.414456

Peretz, I. (2009). Music, language and modularity framed in action. *Psychologica Belgica, 49,* 157–175.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33,* 185. doi:10.1016/S0896-6273(01)00580-3

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6,* 688–691. doi:10.1038/nn1083

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology, 56,* 89–114. doi:10.1146/annurev.psych.56.091103.070225

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, The Netherlands: Association Européenne des Conservatories.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of “tone deafness”. *Music Perception, 25,* 95–115. doi:10.1525/mp.2007.25.2.95

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics, 71,* 1385–1398. doi:10.3758/APP.71.6.1385

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *Journal of the Acoustical Society of America, 128,* 2182–2190. doi:10.1121/1.3478782

Pfordresher, P. Q., & Halpern, A. R. (in press). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review*.

Pfordresher, P. Q., & Mantell, J. T. (2009). Singing as a form of vocal imitation: Mechanisms and deficits. In J. Louhivuori, 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). Finland: Jyväskylä.

Pfordresher, P. Q., & Mantell, J. T. (2012). Self-imitation and the role of inverse models in poor-pitch singing. *Abstracts of the Psychonomic Society, 17,* 34.

Riecker, A., Ackerman, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport, 11,* 1997–2000. doi:10.1097/00001756-200006260-00038

Roberts, E., & Davies, A. D. (1975). Poor-pitch singing: Response of monotone singers to a program of remedial training. *Journal of Research in Music Education, 23,* 227–239. doi:10.2307/3344852

Saito, Y., Ishii, K., Yagi, K., Tatsumi, I. F., & Mizusawa, H. (2006). Cerebral networks for spontaneous and synchronized singing and speaking. *Neuroreport, 17,* 1893–1897. doi:10.1097/WNR.0b013e328011519c

Schmidt, R. A., & Lee, T. D. (1999). *Motor control and learning: A behavioral emphasis* (3rd ed.). Champaign, IL: Human Kinetics.

Serafine, M. L., Davidson, J., Crowder, R. G., & Repp, B. H. (1986). On the nature of melody-text integration in memory for songs. *Journal of Memory and Language, 25,* 123–135. doi:10.1016/0749-596X(86)90025-2

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review, 16,* 374–381. doi:10.3758/16.2.374

Stegemöller, E. L., Skoe, E., Nicol, T., Warrier, C. M., & Kraus, N. (2008). Music training and vocal production of speech and song. *Music Perception, 25,* 419–428. doi:10.1525/mp.2008.25.5.419

Welch, G. F. (1979). Poor pitch singing: A review of the literature. *Psychology of Music, 7,* 50–58. doi:10.1177/030573567971006

Wong, P. C. M., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *The Journal of Neuroscience, 24,* 9153–9160. doi:10.1523/JNEUROSCI.2225-04.2004

Yost, W. A. (2007). Perceiving sounds in the real world: An introduction to human complex sound perception. *Frontiers in Bioscience, 12,* 3461–3467. doi:10.2741/2326

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch processing in speech perception. *Science, 256,* 846–849. doi:10.1126/science.1589767

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. doi:10.1037/0096-1523.32.4.944

Appendix A

Notation for Melodies

Figure A1. Notation for melodies corresponding to both genders, contours, and all texts. All sequences were used for at least one participant in the study except for the female question contour for the text "she parked the car," which, due to our target randomization and random assignment of participants to conditions, was never used.

(*Appendices continue*)

Appendix B

Computing Normalized Distinctiveness

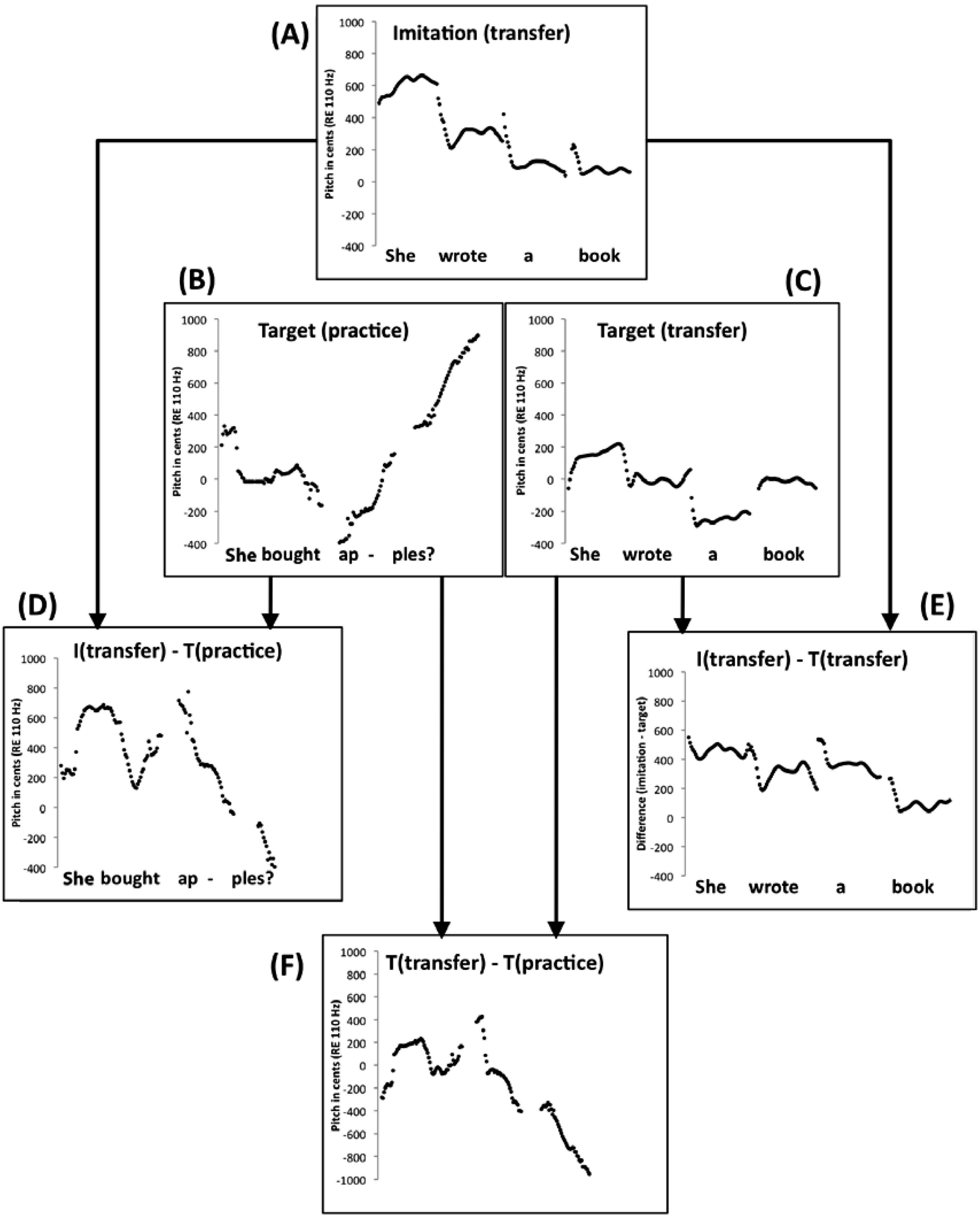

Figure B1. An illustration of the worked example in Appendix B for calculating normalized distinctiveness.

(*Appendices continue*)

In this appendix, we describe the normalized distinctiveness measure (Eq. 1) in greater detail, with reference to an example. Figure B1 shows plots with data adapted from targets and a male participant's imitation. All plots show F0 in cents relative to a 100 Hz reference note.

The top center panel (panel A) shows the pitch/time trajectory for an imitation of the sung melody, "She wrote a book," with a question contour during a transfer trial. Below this panel are plots of the trajectories associated with the target this participant previously imitated during practice (panel B), when the participant had imitated the sentence "She bought apples?" that had a question contour, as well as the target the participant currently imitates during transfer (panel C). Clearly, the imitative performance in panel (A) is not a perfect performance. There is a tendency for the performance to go "sharp" relative to the target, and the upwards pitch change between the last two notes (meant to convey a "question" contour in the target) is imitated as if in unison. At the same time, the imitation appears much more similar to the current target (panel C) than the previously imitated target during practice (panel B).

Below these pitch trajectories are three more panels that relate to the three terms shown in Eq. (1). All of these show signed differences in pitch across time, the absolute values of which are used in Eq. (1). Arrows are drawn to indicate which of the trajectories from panels (A–C) contribute to the difference scores. Panel (D) shows difference scores contrasting produced pitch in imitation to the pitch values of the target that was used during previous practice trials. The absolute values of these differences scores are summed and then averaged, constituting the minuend of the numerator in Eq. (1) $= 1/n\sum|I_{transfer} - T_{practice}|$. In this example, the mean absolute value for this term is 435 cents. Absolute values were used so that negative and positive differences for $I_{transfer} - T_{practice}$ would not cancel out in summation. Panel (E) shows a similar trajectory of difference scores based on contrasting the pitch produced during imitation with pitch in the target currently being imitated, *T(transfer*). The mean absolute value of these scores constitutes the subtrahend of the numerator in Eq. (1) $= 1/n\sum|I_{transfer} - T_{transfer}|$. In almost every case, the mean absolute value of these scores should be lower than the minuend, leading to a positive signed difference in the numerator. That is the case here; the mean absolute value for the scores in panel (E) is 311 cents, and so the numerator of Eq. (1) would be 124 cents.

The numerator on its own provides useful information in that it shows in absolute terms how "close" the present imitation is to the current target versus the target previously imitated during practice. Based on this measure, the current performance seems much more similar to the current (transfer) target than to the earlier (practice) target. However, targets varied considerably with respect to their similarity to each other. Whereas the present example used targets that were highly distinct (a spoken versus a sung target, with different text settings), others pairs were more similar to each other. Thus, the denominator of Eq. (1) plays a critical role in "normalizing" the difference expressed in the numerator. Panel (F) shows differences between the transfer target and the pervious practice target (note that the scaling of the ordinate for this panel differs from other panels). Despite the fact that these targets differed from each other considerably, the mean of their differences from each other is lower than either term in the denominator, 288 cents.

Having derived all the terms for Eq. (1), we can compute that the normalized distinctiveness of this imitation from the previous practice trial is 0.43. This is a high value given the mean performance (cf. Figure 6), thus confirming the qualitative observations given above. A smaller difference between the targets, of course, would lead to a higher score, as would a greater contrast between the minuend and subtrahend of the numerator. In the interest of providing benchmarks, an absolutely perfect imitation of the current target would lead to a score of 1.00, whereas a performance that perfectly matches the previous target but not the current target would lead to −1.00.

Received May 9, 2012
Revision received February 28, 2013
Accepted March 22, 2013 ■