Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyväskylä, Finland
Jukka Louhivuori, Tuomas Eerola, Suvi Saarikallio, Tommi Himberg, Päivi-Sisko Eerola (Editors)

Singing as a Form of Vocal Imitation: Mechanisms and Deficits

Peter Q. Pfordresher,[*1] James T. Mantell [*2]

[*]*Department of Psychology, University at Buffalo, The State University of New York, U.S.A.*
[1]pqp@buffalo.edu, [2]jtm29@buffalo.edu

ABSTRACT

There has been a recent upsurge of interest in the neural and cognitive bases of inaccurate singing, commonly referred to as "tone deafness." Explanations of this deficit have commonly focused on perceptual and motor functions. It is clear, however, that neither of these mechanisms can fully account for deficits in singing. We summarize the results of several studies concerning inaccurate, or "poor pitch" singing. Taken together, the results of these studies argue that the basis for singing-related deficits lies in the link between perception and action, rather than strictly motoric or perceptual factors. Moreover, singing deficits may involve general purpose vocal imitation mechanisms, rather than mechanisms that are specific to music.

I. INTRODUCTION

Inaccuracies in singing constitute an interesting paradox for the music psychologist. On the one hand, inaccurate singing appears to be a pervasive problem. Most people feel they have heard a poor singer and by a recent estimate possibly more than half of young adults feel they cannot imitate a melody by singing (Pfordresher & Brown, 2007). On the other hand, the enjoyment of music appears to be universal across persons and cultures, and expression with the voice is a natural and intrinsic part of emotional expression in both music and language (cf. Juslin & Laukka, 2003).

In the current paper we review research designed to disentangle the paradox of inaccurate singing. We are explicitly concerned with the way in which people imitate pitch while singing, and secondarily interested in the reproduction of timing and the articulation of words during singing. In general, these data support a model of singing like that shown in Figure 1. This model constitutes an attempt to represent the basic components involved in singing melodies (cf. Harvey, 1985). In principle, singing deficits could involve one or more of these components.

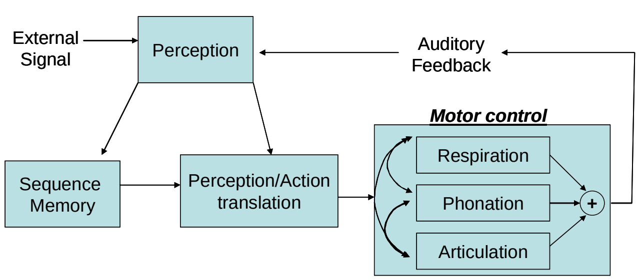

Figure 1. Functional components involved in singing.

A. Background for our model

The model shown above was motivated both by our intuitions and by past research. For instance, the contribution of "pitch perception" is of interest given the results of recent research. Specifically, Peretz and colleagues have identified a musical disorder termed *Congenital Amusia*. Congenital amusia is a music perception deficit that, unlike acquired amusia, is not attributable to neural insult. Though congenital amusia is a perceptual deficit, possibly related to basic pitch discrimination (Hyde & Peretz, 2004), those with congenital amusia are – not surprisingly – also typically deficient in production (Ayotte, Peretz, & Hyde, 2002). Congenital amusia is rare (estimated at approximately 5% of the population, cf. Kalmus & Fry, 1980) — perhaps rarer than the frequency of inaccurate singing. Nevertheless, its presence poses an important question: To what degree does inaccurate singing in general reflect deficiencies in the perceptual system? In other words, is inaccurate singing truly 'tone deafness', as described by the general population (Cuddy, Balkwill, Peretz, & Holden, 2005)?

It is important to note that pitch perception during singing is not identical to pitch perception in typical perceptual experiments. First, perception during sensorimotor tasks may differ from perception that occurs with no motor involvement; in singing, this most obviously relates to the role of bone conducted feedback (e.g., Howell, 1985). Second, in singing one often has to coordinate feedback from oneself with auditory information from other singers (cf. Fine, Berry, & Rosner, 2006).

On the other hand, it is intuitive to consider singing inaccuracies as related to motor control. Singing, after all, is a multifaceted complex motor activity (Sundberg, 1987). In Figure 1 we have represented motor control of singing with respect to three interacting components, respiration (breathing), phonation (control of laryngeal muscles) and articulation. In terms of pitch control during singing, phonation is obviously of central importance. Of course, phonation would be impossible without respiration and articulation allows a singer to associate pitches with lexical meanings and clearly segment notes.

Mediating the relationship between pitch perception and motor control are two components that may play a central role in singing-related deficits. One component, sequence memory, involves the degree to which a singer has retained the structure of a to-be-imitated sequence in memory. Recent research suggests that novice and expert musicians can both form sophisticated representations of musical structure at an implicit level (Tillmann, Bharucha, & Bigand, 2000). However, it may be the case that musically deficient populations form less defined representations. Though sequence memory can be updated via pitch perception, it does not necessarily require pitch perception. This is illustrated by the fact that singers can continue to produce melodies (albeit with somewhat less precision) when auditory feedback is masked (Ward & Burns, 1978). Finally, our model includes a component involved in sensorimotor translation, which is the locus of imitation. This component receives information from sequence memory and can be separately updated by pitch perception (e.g., when matching pitch with a song one hears for the first time).

B. Measuring inaccuracy

In the broader population, people often speak as though the distinction between accurate and inaccurate singing were transparent. However, formally distinguishing groups of singers as good or bad is not a trivial issue. The research described here, in line with current trends (e.g., Dalla Bella, Giguère, & Peretz, 2007) focuses on acoustic measurements of F0 in vocal productions, rather than ratings. Although subjective ratings of vocal performance can provide informative and reliable results (e.g., Ayotte et al., 2002; Welch, 2006), that may converge with results from acoustic measures (Wise & Sloboda, 2008), acoustic measures are often preferable because they allow the evaluation of vocal performance on a qualitative ("accurate" versus "inaccurate") and quantitative basis.

One complication in defining inaccuracy is the issue of criterion. Based on standard statistical practice, a plausible criterion would be to define 'inaccurate' singers as those who fall more than 2 standard deviations away from mean performance, in a direction consistent with the idea of a 'deficit' (e.g., their pitch errors are higher than the mean by 2 standard deviations). Unfortunately, such a standard is problematic in that it leads to a dependency between the criterion for inaccurate singing and performance characteristics of the group used as a basis for comparison. For instance, if a sample of 10 singers contains 9 professionals, the single non-professional would likely be considered an inaccurate singer when compared to the professionals, though that singer may demonstrate "normal" performance among a sample of 10 non-singers.

As such, most researchers use fixed cutoffs to distinguish accurate from inaccurate singers. For instance, in the aforementioned research on congenital amusia, musically deficient individuals were first identified by self-report. This personal assessment was later verified in a battery of tasks that eventually formed the Montreal Battery of the Evaluation of Amusia (MBEA, Peretz, Champod, & Hyde, 2003). Tests of the MBEA have been used to determine specific cutoff scores (based on norming studies) that define congenital amusia.

Our research focuses on vocal production more so than perception; thus, we analyze the extent of mistuning as an index of singing inaccuracy. Specifically, we examine the difference (in cents) between produced notes in the imitation and target notes in the to-be-imitated stimulus. Vocalists whose difference scores are greater than 100 cents (in absolute terms) on average are deemed inaccurate singers. We originally based this criterion on the common characterization of inaccurate singers as being 'out of tune' (cf. Hyde & Peretz, 2004). There are, however, several other ways that inaccuracy in pitch can be determined.

It is important to note that the criterion described in the previous paragraph (which is the criterion used in the studies we will describe) is based on accuracy in *absolute pitch*. Thus, it is possible that a singer who is defined as 'inaccurate' based on our criterion could sing a tune perfectly with respect to *relative pitch*, but the song would be in the wrong key. Accordingly, an alternate pitch accuracy metric would be to only label as inaccurate those singers who distort the size of pitch intervals when they sing. In this case, inaccurate singing would be based on deficits in the imitation of relative pitch.

Finally, we bring up an important distinction between two ways of measuring pitch production that could be applied to the imitation of relative or absolute pitch. So far, we have referred to 'accuracy' as the primary characteristic of measurable singing performance. There is, however, an important formal distinction in motor control between *accuracy* and *precision* (Schmidt & Lee, 1999). Accuracy refers to whether performance deviates from some target in a specific direction. With respect to singing, accuracy simply refers to whether produced tones are 'sharp' or 'flat' relative to target tones. Because accuracy is typically determined based on average performance, it is possible for a participant to be 'accurate on average', although he or she may actually generate a sloppy performance on a note-by-note basis. Consider a singer who sings every odd note 50 cents sharp and every even note 50 cents flat: though the performance will be accurate on average, it may not sound 'good' to a listener. Such a singer would be exhibiting high accuracy but low *precision*. Thus, precision refers to the degree to which performance is variable, irrespective of whether performance is accurate or not.

C. Experimental procedures

Prior to our research summary we offer a brief overview of the method common to each study described below.

1) Participants. In these studies, we attempted to limit the number of participants with extensive musical training and in most cases eliminated any participants with vocal training. Persons with vocal training were only included if their data did not differ significantly from others (suggesting non-expert performance). Furthermore, our participants were healthy young adults. Thus, we have attempted to describe the vocal performances of normal, untrained persons rather than the performances of highly trained individuals or individuals with neuro-cognitive deficits.

2) Materials. In the experiments described here, participants vocally imitated simple 4-note melodic sequences. We used a voice synthesis package (Vocaloid Leon, Zero-G Ltd.) to generate sequences for male singers, and a female counterpart (Vocaloid Lola) to generate sequences for female singers. Certain studies also include perception tasks, which comprised sine wave tones.

3) Procedure. Vocalization tasks reported here follow an overt imitation procedure in which participants first listen silently to a sequence and then reproduce it immediately afterwards. Perceptual tasks are likewise simple, involving listening to a pair of tones or tone intervals and then making 'same/different' judgments.

II. DEFINING INNACURATE SINGING

As mentioned before, serious music perception deficits may exist in about 5% of the population. However a different rate may obtain for singing deficits, particularly if singing deficits are not solely due to perception (see Figure 1). And, as we will demonstrate, the rate for singing deficits does differ.

A. Pitch-matching abilities

As part of an attempt to pre-screen a large group of participants for vocal imitation abilities, we had 106 randomly selected non-musicians vocally imitate 4-note monotone sequences. These sequences represented a range of pitch values surrounding each individual's comfort pitch (based on warm-up trials). We measured the mean signed difference between produced and target notes across these sequence to determine each participant's accuracy. Figure 2 illustrates the cumulative distribution of these scores after conversion of the signed differences to absolute values. As can be seen, the majority of participants (76%) generate average error scores of less than 100 cents. According to this metric, most participants can accurately sing within one semitone of the target.

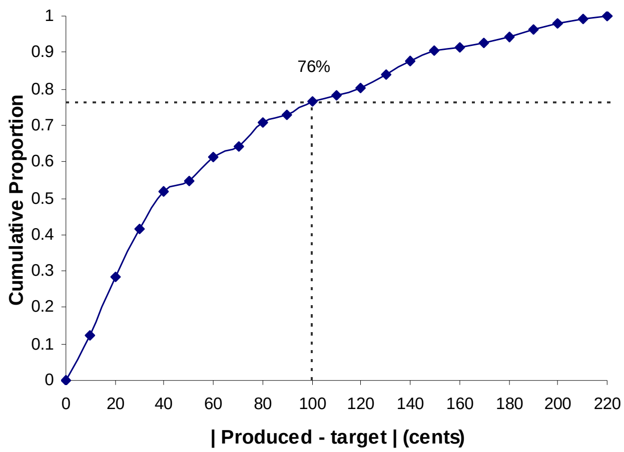

Figure 2. Cumulative proportion of absolute differences between produced and target notes for the imitation of monotone sequences. The dashed lines highlight the cumulative frequency for a 100-cents error score.

One might expect the rarity of inaccuracies shown in Figure 2 merely to reflect the simplicity of the task. In fact, data from tasks that involve the imitation of more complex sequences have yielded strikingly similar figures. Pfordresher and Brown (2007) had participants imitate 4-note sequences with changing pitches (more melodic) and Dalla Bella and colleagues (2007) had participants perform a popular melody from long-term memory (*Gens du pays*). Both studies yielded approximately the same rate of accurate versus inaccurate singing, 10-15% across studies.

B. Relative pitch

As stated before, the ability to match pitch is not the only (and may not be the best) measure of singing accuracy. Our perspective is that an optimal measure of singing ability is one that clearly discriminates between accurate and inaccurate singers.

Accuracy in relative pitch, as mentioned earlier, may be considered more important to music production than accuracy in absolute pitch, at least when people are singing alone (obviously, absolute pitch takes on great importance in group performances). In fact, the ability to imitate absolute and relative pitch is positively correlated (Dalla Bella et al., 2007; Pfordresher & Brown, 2007). However, when vocal performances are assessed strictly on the basis of relative pitch accuracy, individual differences are not as great as when they are scaled based on absolute pitch. Pfordresher & Brown (2007) standardized performance data on an absolute and a relative pitch accuracy measure and found that individual differences were greater by an order of magnitude for absolute as compared to relative pitch errors. Thus, accuracy based on absolute pitch may be a better way to distinguish accurate from inaccurate singers than is accuracy in relative pitch.

That being said, inaccurate singers – when defined with respect to pitch matching – typically do exhibit inaccurate imitation of relative pitch. One crucial finding from Pfordresher and Brown (2007) is that inaccurate singers have a tendency to *compress* pitch intervals. As target interval size increases, inaccurate singers tend to shrink the size of the produced interval. Moreover this tendency to compress intervals was proportional to the degree of mistuning a singer exhibited.

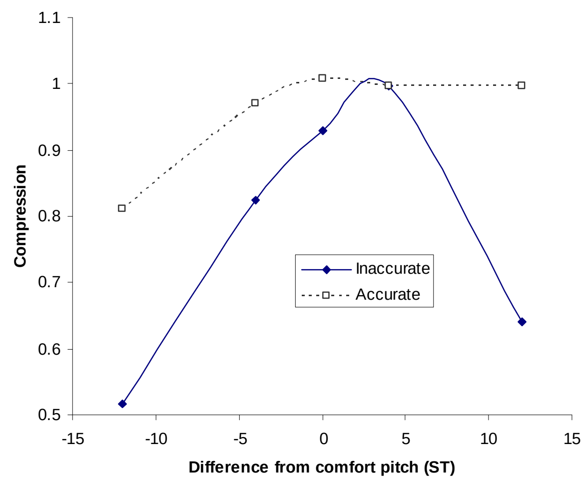

Figure 3. Compression of pitch intervals as a function of the difference between the starting pitch of a sequence and individual comfort pitches.

Recently, we ran an experiment in which participants were required to imitate melodies that could be near to or far from their comfort pitch. (Proximity to one's comfort pitch was established by the starting note of target sequences; pitch intervals in sequences equally often were higher or lower than the starting pitch.) In this way we attempted to experimentally manipulate "mistuning" of the target melody relative to an individuals vocal range. We measured compression of pitch intervals (as in Pfordresher & Brown, 2007) by regressing produced interval size on target interval size. This analysis technique provides an intuitive representation of compression; compression is suggested when the slope of the regression line is less than 1, whereas a slope of 1 indicates no compression. Figure 3 plots these slope values as a function of singing category (accurate/inaccurate) and how distant a melody's starting pitch is from one's comfort pitch (in semitones). As can be seen, inaccurate singers compress intervals more than accurate singers, and these differences increase as component pitches differ from one's comfort pitch. Moreover, a hallmark of inaccurate singing is an inability to reproduce relative pitch across a wide range of keys.

C. Precision

As mentioned previously, motor control research often distinguishes between *accuracy* and *precision*. Thus, one may argue that singing precision matters as much as singing accuracy. In any case, it is valuable to address individual differences with respect to both accuracy and precision. Table 1 shows an analysis of accuracy and precision from the data of Pfordresher and Brown (2007), who initially did not report analyses of precision. Imprecision was defined as a standard deviation in note production of 100 cents or more for repeated production of the same pitch (note that Table 1 represents different participants and target stimuli from the data shown in Figure 2). An imprecise singer may reproduce the note C4 as C4 on average but for any single instance may instead produce a tone that is closer to C#4, or B3, etc. Accuracy and precision measures were correlated across participants, $r(76) = .51$, $p < .01$.

Table 1. Accuracy and precision rates from Pfordresher and Brown (2007).

	Accuracy		
Precision	***Accurate***	***Inaccurate***	**Row Sums**
Precise	53%	1%	54%
Imprecise	35%	12%	46%
Column Sums	87%	13%	

As can be seen in Table 1, imprecision is more common than inaccuracy. More important to us, however, is the degree to which imprecision predicts inaccuracy and vice-versa. According to these data, the conditional probability of being inaccurate when one is imprecise, p(inaccurate | imprecise) is 0.26. By contrast the conditional probability of imprecision given inaccuracy, p(imprecise | inaccurate) is 0.92. Thus it appears that the state of being inaccurate predicts imprecision more so than the reverse. These relationships support the idea that accuracy, more so than precision, is the most diagnostic performance characteristic to measure when investigating deficits of singing.

III. ETIOLOGY

A major issue of importance in poor-pitch singing concerns the locus of the deficit. Pfordresher and Brown (2007) identified four 'canonical' models of poor-pitch singing based on a modular framework like that shown in Figure 1. That study, and subsequent research, has attempted to determine which of the several possible components may be deficient in poor-pitch singing. The following sections address each hypothesis and suggest tentative conclusions.

A. Absence of perceptual deficits

Most poor-pitch singers do not exhibit a companion deficit in basic pitch perception tasks. Thus, most inaccurate singers do not suffer from a syndrome comparable to congenital amusia (cf. Hyde & Peretz, 2004). As such, the colloquial term "tone deaf" is, for most persons, inappropriate. That being said, truly tone deaf individuals are likely to be inaccurate singers; as such, a minority of inaccurate singers will also be deficient in pitch perception. This sentiment accords with Pfordresher and Brown (2007) who found that 1 of 10 inaccurate singers was deficient in basic pitch perception, based on performance in a pitch discrimination task.

B. Absence of disordered phonatory control

Another possibility is that inaccurate singers have difficulty controlling laryngeal muscles, which in turn guide phonation (see Figure 1). This hypothesis seems reasonable, considering that singing is a nontrivial motor task (cf. Sundberg, 1987). If faulty muscle control is responsible for poor singing, inaccurate singers might present a lack of control and/or a restricted range of movement for laryngeal muscles. However, our data argue against such a view—inaccurate singers do not seem to have a restricted vocal range for non-imitative warm-up tasks that explicitly probe phonation limits (Pfordresher & Brown, 2007). Furthermore, recent re-analyses of the data from Pfordresher and Brown suggests that inaccurate singers can sustain level tones with similar precision to accurate singers (SD of F0 = 14 Hz for each group).

C. Mixed evidence for sequence memory

The basis for inaccurate singing may result in part from degraded representations of sequence memory. Thus far, however, evidence for this hypothesis is mixed. Wise and Sloboda (2008) reported that differences in pitch accuracy between participants who self-define as "tone deaf" and those who do not was greater for longer than for shorter sequences. An implication of this result is that the less accurate (and apparently less confident) group had relatively more difficulty maintaining complex pitch sequences than did the more accurate group. Similarly, Wise and Sloboda found that differences across groups were reduced when singers were accompanied by the correct melody. The accompanying information would reduce memory demands and allow singers to rely on the immediate perceptual feedback resultant from the difference between their pitch and the model's pitch.

In contrast, Pfordresher and Brown (2007) found exactly the opposite results with a similar set of conditions. In their study, poor-pitch singers performed more accurately on more complex melodies (here defined by the number of unique pitches in a 4-note sequence rather than the sequence length) and performed worse with corrective feedback. The main difference between these studies was that Wise and Sloboda (2008) categorized singers based on self-report, whereas Pfordresher and Brown defined singer categories based on produced pitch matching errors. However, it is unclear why these different classifications would lead to such different results; future research should explore the fine-grained effects of accurate auditory feedback on intoned vocal productions.

D. Imitation: Mapping perception to action

Thus far, our data downplay the importance of perceptual, motor, and memory components in singing deficits. As Figure 1 suggests, another potential candidate for inaccurate singing resides at the confluence of perception and action; that is, mapping pitch to motor gestures. This component must be contrasted to pure muscle/motor control in that it is not a description of motor potential, but of the neural translation that precedes vocal production. We suggest that the singer (accurate or not) intends to imitate the target (the to-be-imitated stimulus) and produce the correct pitch. Inaccuracy may result if the translation from intention to action is not ideally transmitted.

Because we believe that the mapping hypothesis has greater explanatory potential than purely perceptual, memory, or motor hypotheses, our ongoing research has focused on tasks that test individual differences in vocal imitation, broadly construed. We have adopted an intentional imitation paradigm in which participants are instructed to imitate the sequence they have just heard. Imitations, unlike spontaneously produced speech or song, allow measurement of vocal performance based on the ideal target, including fine-grained fluctuations in pitch (e.g., vibrato), timing, and articulation. Importantly, we can vary numerous stimulus parameters and investigate their influence on vocal imitation accuracy. These manipulations can provide critical insight into the foundations of singing deficits.

IV. SINGING AS IMITATION

If deficits in singing reflect general-purpose vocal imitation mechanisms, then those who are inaccurate singers should likewise be inaccurate when imitating a non-musical target. We have tested this possibility by having participants imitate both sung melodies and spoken sentences. Both sequence types contain time-varying pitch information that conveys meaning and is culturally relevant. However, the structure of pitch contours in spoken language (American English, in our research) differs strikingly from pitch structures in tonal music. Moreover, neuroscientific evidence suggests that different brain regions are used to process pitch in music and language (e.g., Peretz, 2006).

One important way that the use of pitch differs across music and language has to do with the degree to which phonetic information constrains pitch structures. Whereas mapping between pitch and text is relatively flexible in singing (e.g., Palmer & Kelly, 1992), this link is not as flexible in speech, where pitch conveys important suprasegmental information that is linked to the text (e.g., Cruttenden, 1997). With respect to the model shown in Figure 1, one might expect greater coupling of phonation and articulation in the imitation of sentences than in the imitation of melodies. Thus, we have addressed the degree to which one's ability to imitate melodies or sentences depends on the presence of accompanying phonetic information.

In this research, participants imitated pre-recorded tokens of sentences and melodies that were produced naturally by two individuals who both sang and spoke the tokens. Starting with recorded sentences of three to five syllables in length, we then created melodies that matched the sentences in terms of pitch range, contour, and word content. These tokens constituted our *worded* conditions in that they paired pitch contours with matching phonetic information. We then created *wordless* melodic and sentence stimuli by synthesizing the pitch-time traces of the worded utterances. These wordless exemplars retained all of the dynamic pitch information of their worded counterparts, but lost all semantic and phonetic information.

Figure 4 shows the accuracy with which a sample of 71 individuals matched the pitch of sentences or melodies during imitations; these data are averaged across worded and wordless sequences. As can be seen, there is a strong positive correlation between imitation performance on both tasks, $r(69) = .78$, $p < .01$. Dashed lines are used to highlight the 100-cent boundaries used by Pfordresher and Brown (2007) to distinguish accurate from inaccurate singers. Although these boundaries are not similarly valid within the domain of speech, it is nonetheless interesting to note that individuals who are inaccurate singers (by this criterion) tend also to be inaccurate speakers given the very same (music-based) criterion. Similar relationships were observed regardless of whether participants imitated worded or wordless sequences and the relationship is still significant (though attenuated) when statistical outliers (inaccurate singers) are removed, $r(61) = .50$, $p < .01$.

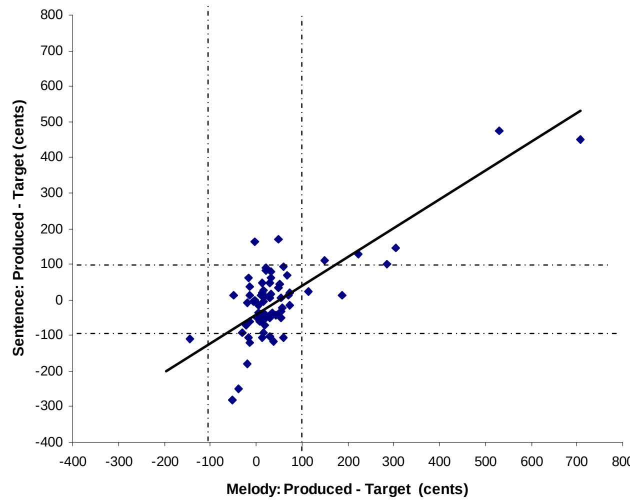

Figure 4. Scatterplot relating signed pitch error scores (positive = sharp) for the imitation of melodies and the imitation of sentences. Each data point represents mean performance for an individual. Dashed lines highlight boundaries of +/- 100 cents and the solid black line represents the best-fitting linear regression.

Although individual differences suggest common mechanisms for imitation of music and speech, mean performance across tasks suggests a more nuanced interpretation. For both melodies and sentences, the presence of phonetic information in worded sequences improved performance relative to wordless sequences. However, this advantage may be greater for sentences than for melodies. This implication arose when we addressed the accuracy of imitating relative pitch (not shown in Figure 4) and in the accuracy with which people imitated the timing (rate) of sequences. These results will be discussed in more detail in a forthcoming paper.

Taken together, these results suggest that individual differences in singing harness general-purpose vocal imitation mechanisms that are not limited to music but may also be used when imitating non-musical sequences like spoken sentences. Moreover, these domain-general individual differences may co-exist with domain specific constraints on imitation, such as the degree to which phonation is coupled to articulation.

V. CONCLUSION

In this paper we have summarized several recent studies concerning individual differences in singing. Inaccurate singing, though rare, has important implications for the understanding of singing and potentially for vocal imitation in general. In general, the results of these studies suggest that inaccurate singing constitutes domain-general failure of vocal imitation. Inaccurate singing, moreover, is borne out in two consistent error patterns, mistuning of pitch and compression of pitch intervals. In future research we hope to explore further the cognitive and neural bases of this disorder, and to test the degree to which training strategies can lead to the improvement of inaccurate singers.

ACKNOWLEDGMENT

This research was supported by San Antonio Life Sciences Institute Grant #121075 and by NSF grants BCS-0344892, 0704516 and 0642592. We thank Erik Gallemore, Danielle Maddock, Julliann Mejia, Brian Benitez, Lilly-Ann Flores, Zachary Clay, Jennifer Walsh, Ece Yildirim, David Ricotta, Amy Mazur and Meredith Zintz for help with data collection and analysis.

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.

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge, UK: Cambridge University Press.

Cuddy, L. L., Balkwill, L., Peretz, I., & Holden, R. R. (2005). 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.

Fine, P., Berry, A., & Rosner, B. (2006). The effect of pattern recognition and tonal predictability on sight-singing ability. *Psychology of Music, 34*, 431-447.

Harvey, N. (1985). Vocal control in singing: A cognitive approach. In P. Howell, I. Cross, & R. West (Eds.) *Musical structure and cognition* (pp. 287-332). London: Academic Press.

Howell, P. (1985). Auditory feedback of the voice in singing. In P. Howell, I. Cross, & R. West (Eds.) *Musical structure and cognition* (pp. 259-286). London: Academic Press.

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

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770-814.

Kalmus, H., & Fry, D. B. (1980). On tune deafness (dysmelodia): Frequency, development, genetics and musical backgrounds. *Annals of Human Genetics, 43*, 369-382.

Palmer, C., & Kelly, M. H. (1992). Linguistic prosody and musical meter in song. *Journal of Memory and Language, 31*, 525-542.

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition, 100*, 1-32.

Peretz, I., Champod, A., & Hyde, K. (2003). Varieties of musical disorders: The Montreal battery of evaluation of amusia. *Annals of the New York Academy of Sciences, 999*, 58-75.

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

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

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

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review, 107*, 885-913.

Ward, W. D., & Burns, E. M. (1978). Singing without auditory feedback. *Journal of Research in Singing, 1*, 24-44.

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

Wise, K. J., & Sloboda, J. A. (2008). Establishing an empirical profile of self-defined “tone deafness”: Perception, singing performance and self-assessment. *Musicae Scientiae, 12*, 3-26.