# The Vocal Generosity Effect: How Bad Can Your Singing Be?

Sean Hutchins, Catherine Roquet, & Isabelle Peretz
*International Laboratory for Brain, Music and Sound research (BRAMS) & Université de Montréal, Montréal, Canada*

**PRIOR WORK INDICATES THAT LISTENERS MAY BE MORE** likely to call a note in-tune when it is sung than when it is in another timbre. The current study seeks to confirm whether this vocal generosity effect generalizes to melodies. Musicians and nonmusicians listened to pairs of single tones and scale-based melodies performed with the voice or the violin. The final note was varied in how well it was tuned to the prior context, and for each example, listeners judged whether the final note was in-tune or not. A strong vocal generosity effect was found for musicians and nonmusicians in both melodic and single tone conditions – a higher degree of mistuning was necessary for listeners to decide that sung tones were out-of-tune compared with violin notes. These results confirm the role of timbre in tuning judgments, and help explain why singers are typically less well-tuned than instrumentalists in performance.

*Received September 23, 2011, accepted March 9, 2012.*

**Key words:** tuning, pitch, voice, violin, perception

In any musical performance, it is imperative that each note be in-tune. But this raises the important question, "What counts as in-tune?" This question is especially critical for instruments such as the violin, or voice, for which the performer must control the tuning of each note each time it is played. While performers will almost always try to produce their target pitch as closely as possible, some error will generally be present. Thus we ask, how far away from the target pitch is the performer allowed to be before listeners will judge the note to be out-of-tune?

Studies of vocal pitch matching have needed to address exactly this question. Two different criteria have been suggested and used, both based on the fact that musical notes have a minimum of one semitone difference. Pfordresher and Brown (2007) and Dalla Bella, Giguère, and Peretz (2007), for example, use an accuracy criterion of ±100 cents (100 cents = 1 semitone = 1/12 octave) from a target pitch to determine pitch matching errors. Demorest and Clements (2007) and Hutchins and Peretz (2012), on the other hand, use an accuracy criterion of ±50 cents. Amir, Amir, and Kishon-Rabin (2003) point out the merits of both while advocating neither and Bradshaw and McHenry (2005) analyze their data in both ways independently. The ±100 cents criterion requires that a note be a full semitone off to be counted as wrong, but yields a 200 cent acceptable range. The ±50 cents criterion assumes that notes more than halfway to another musically valid note will be heard as more like the other note than the target note, and does not give overlapping ranges for target notes a semitone apart.

Use of each of these different criteria for determining what counts as out-of-tune, however, has generally been driven by assumptions, rather than data. A better way to determine what should be called in- or out-of-tune would be to present notes with different mistuning to listeners and determine where they draw the boundary between acceptable and unacceptable tunings. Unfortunately, this has not been a well-studied topic to date. The majority of studies that tackle this issue have looked at perception of musical intervals to find the particular frequency ratio perceived as best representative of a particular musical interval (Moran & Pratt 1926; Rakowski 1990; Rakowski & Miskiewicz 1985; Terhardt, 1969; Vurma & Ross, 2006; Ward, 1954). These studies of intonation accuracy are generally geared towards answering questions about preferred tuning, often to test whether listeners tend to prefer particular tuning systems, such as equal-tempered, just, or Pythagorean tuning, which prescribe slightly different intervallic ratios between scale degrees. As a by-product, these types of studies also provide information about the range of acceptable tuning of an interval, which can depend on the specific interval tested, the direction of the interval, and the timbre of the tones. Typically, this varies from

*Music Perception* volume 30, issue 2, pp. 147–159. issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/mp.2012.30.2.147

±10 to ±45 cents (Moran & Pratt 1926; Rakowski 1990; Rakowski & Miskiewicz 1985; Terhardt, 1969; Vurma & Ross, 2006; Ward, 1954).

Listeners are also not reliably able to make identification judgments of intervals (as a perfect fourth, major third, etc.) more precisely than about a semitone (Burns & Ward, 1978; Halpern & Zatorre, 1979; Rakowski, 1976; Schellenberg, 2001; Siegel & Siegel, 1977), which also points to a ±50 cent limit for tuning acceptability. However, these studies do not give us explicit judgments about whether particular mistuned intervals are perceived as acceptable or not. In addition, due to their use of identification or adjustment measures, almost all of these studies use only the judgments of musically trained individuals, which may be quite different from those of nonmusicians (Burns & Ward, 1982).

A few studies have explicitly looked at the range of acceptable tuning. Lindgren and Sundberg (as cited in Sundberg, 1979, 1982) used a tape-splicing technique to place notes with slightly different tunings within recordings of a vocal performance, and asked musically experienced listeners to identify any tuning errors within the performance. They found that musically experienced listeners would accept errors of up to 50–70 cents as in-tune. These were especially unlikely to be identified as out-of-tune when found in metrically unstressed positions, when the errors tended sharp rather than flat, and when they occurred in emotionally prominent points in the song. Sundberg and colleagues (Sundberg, Prame, & Iwarsson, 1996) later ran a similar study in which trained singers were asked to identify any tuning errors in ten professional recordings of Schubert's *Ave Maria*. The listeners here showed a great deal of disagreement with each other. Moreover, some tones generally perceived as out-of-tune were in fact quite close to equal temperament tuning, whereas some perceived as in-tune were among the more distant from equal temperament tuning. The authors concluded that contextual factors were highly important to tuning. However, pitch was not explicitly manipulated while controlling the context, as it was in Lindgren and Sundberg (1972), making it more difficult to disentangle these two factors.

The range of acceptable tuning in a melodic context was also explicitly studied by Fyk (as cited by van Besouw, Brereton, & Howard, 2008) and van Besouw, Brereton, and Howard (2008), who again studied trained musicians judging synthesized tones. The studies found total ranges of 20–25 cents, with tones being labeled as acceptably well-tuned when they were between +10 and –15 from perfect tuning. van Besouw et al. (2008) also showed that adding vibrato to a tone increased the lower limit (but not the upper limit) of the range of acceptable tuning by a further 10 cents.

Hutchins and Peretz (2012) also explicitly studied this factor. They asked participants to make same/different judgments about pairs of single notes differing only in their pitch. Musicians and nonmusicians were tested using actual recordings of a different set of nonmusicians' singing tones as well as synthesized vocal tones with a similar (but non-identical) steady-state timbre. Pitch differences ranging from 10 to 100 cents were tested. Overall, the results showed that listeners began to reliably describe a note as out of tune once it was off by somewhere between 20–50 cents. Interestingly, for both musicians and nonmusicians, there were large differences in their tuning judgments to actual voices compared to synthesized voice timbres. People were significantly less likely to hear the tuning differences between two tones if they were actual vocal tones, rather than synthesized voices. In particular, nonmusicians needed about 50 cents of separation between two sung notes to hear the tuning difference, compared with only 30 cents of separation between two synthesized vocal tones.

These results lend more support to the ±50 cents criterion than the ±100 cents criterion for making tuning judgments, but raise two important questions. First, is the greater acceptance of tuning errors for the voice seen in that study really due to a difference in our perception of tuning in the human voice compared with other sounds, or is it due to the fact that, in this case, it was compared to a synthesized timbre? Second, does this finding hold for more musically realistic types of stimuli, in which a note is judged as in-tune or out-of-tune due to its pitch relative to a prior musical context? The present study expands upon Hutchins and Peretz (2012) to address these questions and discover whether the greater acceptance of tuning errors for the voice is a generalizable phenomenon.

In terms of production, it has been noted that even well-trained singers do not always produce notes that are perfectly tuned. Seashore (1938) noted that opera singers often start about 9 cents flat, for first 200 ms, before correcting their tuning; this correction from a flat onset has been noted to an even greater extend in untrained singers (Hutchins & Campbell, 2009). Prame (1997) showed that professional singers, in their recorded work, could be up to +42 to –44 cents off pitch. Presumably, these recorded versions were approved by the singer and various others, and these mistunings were either unnoticed or deemed acceptable. In contrast, violinists and wind players conform rather closely to the equal-tempered scale, with average deviations of about only 11 to 17 cents, significantly less than singers (Geringer, 1978).

These same trends held for perception of scales performed by the voice versus wind and string instruments (Geringer, 1978). Based on this evidence, Seashore (1938) opined, and Sundberg (1979) agreed, that the musical ear is extremely generous and operates in the "interpretative mood" when it listens to singing. The work of Hutchins and Peretz (2012) supports this general theory, which we will call vocal generosity, but more testing is necessary to see if this is truly a voice-specific phenomenon.

In addition, this vocal generosity effect may be affected by musical context. In theory, the more musical information that the listener has to draw upon, the stronger judgment they will be able to make about the tuning of a note in the context. Studies of pitch perception across timbres have shown that listeners will make more accurate pitch judgments about a different timbre note with a melodic context than without one (Krumhansl & Iverson, 1992; Semal & Demany, 1991, 1993; Warrier & Zatorre, 2002). This may also be true for pitch tuning perception within a timbre. Hutchins and Peretz (2012) used only a single note context for listeners to make their tuning judgments; it is possible that the vocal generosity effect may diminish as the contextual information allows listeners to make more refined tuning judgments. On the other hand, listeners may be treating the direct comparison between two tones as a same/different judgment rather than an in-tune / out-of-tune judgment, and be able to make more efficient decisions without the musical context. However, Warrier and Zatorre (2002) showed that these two types of judgment methods yield no differences, when directly compared with each other. Participants do not use different criteria for same/different judgments and in-tune/out-of-tune judgments.

The current experiment builds upon Hutchins and Peretz (2012) to examine the scope of the vocal generosity effect. To ensure that the prior results were not due to using nonmusicians compared with synthesized vocal tones, in the current experiment, we use a highly trained professional opera singer (a mezzo-soprano) to create the vocal examples, and we compare these to those produced by a trained violinist. Natural instruments such as the violin are more likely than synthesized tones to elicit errors in tuning judgments, due to imperfections in playing as well as small changes in the timbre across attack, sustain, and release portions of the note. Thus, this represents a more conservative choice than synthesized voice tones used in Hutchins and Peretz (2012). Other acoustic factors, such as the vibrato or pitch stability of the notes may play a factor in the vocal generosity effect, but were not explicitly examined here, due to the already large design. We first wanted to determine whether this effect was replicable in circumstances more akin to real-world music, with trained singers performing in a melodic context.

Both the singer and the violinist performed single tones as well as melodies, which in this case were either scales or broken scales (the same eight notes performed in a different, but musically coherent, order; see Figure 1). Musicians and nonmusicians made tuning judgments about tuning variations on the single tone as well as on the final note of the melody (always the tonic). If the vocal generosity effect holds, we expect to see a larger range of acceptable tuning for the voice example than the violin examples, across both single tones and melodies, for both musicians and nonmusicians. We also expect to see musicians make more accurate tuning judgments than nonmusicians across all conditions.

## Method

### PARTICIPANTS

Twenty-two nonmusicians (9 female) and 22 musicians (15 female) between 18 and 30 years old were recruited from the Montreal university community. Nonmusicians all reported less than two years of formal music training, as most schools offer some music training as a part of the basic curriculum (mean = 0.49 years), and had a mean age of 22.73 years. Some nonmusicians had choral singing experience (mean = 0.40 years), and only one had formal voice training. Musicians all had at least six years of formal music training in a primary instrument (mean = 9.64 years), and had a mean age of 22.00 years. Almost all musicians had experience singing in a group (mean = 3.64 years), and some had formal voice training (mean = 1.50 years). No participants reported having any

FIGURE 1. *Melodies*: Example of ordered scales (left side) and broken scales (right side), both descending (top row) and ascending (bottom row).

auditory deficits or neurological disorders, except for one nonmusician who reported a minor sports-related concussion a few years prior to the study.

### STIMULI AND EQUIPMENT

Two types of stimuli were created: single tones and melodies. All stimuli were recorded in a soundproof studio with a Neumann TLM 103 microphone (Georg Neumann GmbH, Berlin, Germany). Eight single tones were recorded by a violinist with thirteen years of training, ranging from C4 (261.63 Hz) to G4 (392 Hz), in steps of one semitone. The same eight single tones were recorded by a professional opera singer with eleven years of training for a total of sixteen original stimuli. These were then tuned to be precisely on the intended pitch (to correct for minor performance errors; tuning was made with respect to an equal-temperament scale) and adjusted to be each the same duration using Melodyne (Celemony Software GmbH, Munich, Germany). All frequency adjustments made with Melodyne were applied consistently to the entirety of each note, from onset to offset; this preserves the microstructure of the pitch within each tone, including vibrato, and retains the naturalness of the performed stimuli. We also normalized each stimulus for amplitude. Following this, twenty extra tuning deviations were created from each tone by transposing them in steps of ten cents (100 cents = one semitone), in both sharp and flat directions, until the tone reached +/– one semitone. This yielded a total of twenty-one stimuli derived from each single tone.[1]

Melodies were either ordered scales or "broken scales," and either ascending or descending. Broken scales used the same first and last notes as ordered scales (both tonics, separated by an octave), but the other notes in the series were presented out of order (although still musically coherent), so that the last note was approached by a jump in the melody. Each note was held for one beat (tempo = 120 bpm, each beat = 500 ms), with the exception of the first note, which was held for two beats, and the final note, which was held for four beats. The four melody types are shown in Figure 1. Twenty-eight melodies were recorded with the voice and the violin, the fourteen ascending melodies ending on pitches from G#4 (415.30 Hz) to D5 (587.33 Hz), and fourteen descending melodies ending one octave lower. This produced a total of fifty-six original melodies. We used Melodyne to tune the frequency of each note in the melodies and correct any imprecisions in their timing. We also normalized each melody for amplitude. Each melody's last note was then transposed by steps of ten cents, in both directions, until the note reached +/– one semitone for a total of twenty-one stimuli, in the same way as the single tone stimuli. The vibrato characteristics are analyzed in the results section, below. Both single tone and melodic stimuli were presented to participants using ePrime (Psychology Software Tools, Inc., Sharpsburg, Pennsylvania, USA), through closed BeherDynamic DT 990 Pro headphones (Beyerdynamic GmbH & Co. KG, Heilbronn, Germany). Participants' responses were recorded on the computer keyboard, by ePrime.

### PROCEDURE AND DESIGN

The experimental session was divided into two tasks, the Single Tones task and the Melodies task. The order of the tasks was counterbalanced across participants within each music training group. We created seven pseudorandomized stimulus lists for the Melodies task, which included equal numbers of each melody type, direction, instrument, and final tone tuning. There were 168 trials in each list. We created two pseudorandomized stimulus lists for the Single Tones task with the same considerations, comprising 160 trials each. Before the experiment, participants were asked to fill out a consent form and a biographical questionnaire, which took five min to complete. The experiment had an average total duration of one hr and fifteen min.

The Single Tones task had 160 trials presented in pseudorandom order such that trials derived from the same note were separated by at least one intervening trial. In each trial, two tones were presented, with 500 ms of silence between tones. In half the trials, the same tone (either a perfectly tuned note or a version with a tuning deviation) was presented twice; for the other half, the perfectly tuned tone was followed by a version of the same tone with a tuning deviation (or vice versa). Participants were then asked if the second tone was exactly the same, the same note but out-of-tune, or a different note altogether. This wording was designed so that nonmusicians would not make a response of "same" to a note that they considered to be essentially the same note,

[1] This overlap among tones is a necessary feature of the experimental design. In order to exactly match the target and comparison tones, they needed to be created from the same original recording, so that the microstructures are preserved. While this could have been achieved by recording one tone and shifting this to all required pitch levels, such a high degree of pitch shifting produces artifacts that are readily noticeable. To preserve the validity of the stimuli, we thus recorded pitches at eight different pitch degrees, and shifted each only by +/– 100 cents. This way, tones are compared only to versions that are identical in all respects except global pitch differences, including their envelope, timbre, and pitch microstructure. A version of C globally shifted upwards by 60 cents and of C# globally shifted downwards by 40 cents, for example, are noticeably different, despite being the same pitch. Thus, there is more than one version of the same pitch, so that only pitch can be used as a clue for differences within each comparison.

even if they could hear the tuning discrepancy. Musicians and nonmusicians alike understood the task. For the purposes of most of our analyses, responses of "out-of-tune" and "different note" were combined into one category; a set of analyses looking for category distinctions divides these two response types and is presented later.

The rate of hits minus false alarms was calculated for each tuning deviation for each participant. Any response to non-zero tuning deviations that was either judged out-of-tune or different was counted as a hit. The false alarm rate (judgments of "out-of-tune" or "different note altogether" in response to identical stimuli) was relatively small, and did not change much across different tuning deviations. Note here that terms such as hits, false alarms, accuracy, and error are used for analysis purposes only, and are not intended to convey normative judgments.

The Melodies task comprised 168 trials presented in pseudorandom order, such that melodies in the same key were separated by at least one trial. Each trial presented one melody on the violin or the voice, with the last note either perfectly tuned or with a tuning deviation. Participants were then asked if the last note of the melody was the right note and in tune, the right note but out-of-tune, or a different note altogether. This subtle variation in wording between the two tasks has been shown not to evoke any differences in response patterns (Warrier & Zatorre, 2002). Due to the large number of conditions and stimuli, the melodies condition did not present all combinations of all five manipulated variable to each participant. For melodies, hits were calculated in the same way as in the single tone section, with answers of "out-of-tune" or "different note" to changed final tones both counted as hits. However, the false alarm rate was measured using responses to the unchanged tone (no tuning deviation) for each participant. This differed from the false alarm measurement for the single tones, as the stimuli were not presented in pairs.

For both tasks, participants controlled the pace of trials with the space bar, and participants indicated their responses on the keyboard. Both tasks were preceded by a block of five practice trials that were different from those used in the main experiment. During the practice trials, the participant was presented feedback on the accuracy of their responses ("Correct" or "Incorrect") to familiarize them with the task. No feedback was given to the participants during the main experiment.

## Results

### STIMULUS ANALYSIS

We calculated the characteristics of the vibrato for each of the original stimulus tones that were judged. Vibrato reflects a sinusoidal variation in the fundamental frequency, and can be characterized by rate and amplitude. Only final melodic tones and single tones were analyzed. The onsets and offsets were cut from each tone, to remove the initial consonant and any section too faint for frequency analysis. Vibrato data were calculated using an in-house Matlab function. Vibrato rate (measured in Hz) was significantly slower in the voice stimuli ($M$ = 4.47 Hz, $SD$ = 0.24 Hz) than in the violin stimuli ($M$ = 6.43 Hz, $SD$ = 1.82 Hz), $t(40) = 37.42$, $p < .001$ (corrected for unequal variances). Vibrato amplitude (measured in cents) was significantly greater in the voice stimuli ($M$ = ± 98.91 cents, $SD$ = 14.65 cents) than in violin stimuli ($M$ = ± 5.07 cents, $SD$ = 2.80 cents), $t(33) = 6.12$, $p < .001$ (corrected for unequal variances).

### SINGLE TONES

The hits minus false alarms rate was entered in a three-way ($2 \times 2 \times 10$) mixed design ANOVA performed on the factors of Music Experience (between-subjects; musicians versus nonmusicians), Instrument (within-subjects; violin versus voice), and Tuning Deviation (within-subjects; absolute deviation in cents from the initial tone, between 10 cents and 100 cents), see Figure 2a. All $F$ values reported in this experiment are based on Greenhouse-Geisser corrections (which adjusts the degrees of freedom).

The ANOVA showed a significant main effect of Music Experience, $F(1, 42) = 5.94$, $p = .02$, where musicians were more accurate in their responses than nonmusicians, as well as a main effect of Tuning Deviation, $F(5.83, 244.94) = 71.63$, $p < .001$, where participants were more accurate as the deviation in cents increased from the target tone. We also found a main effect of Instrument, $F(1, 42) = 64.88$, $p < .001$, where participants were more accurate overall in the violin trials than in the voice trials. The ANOVA yielded a significant Instrument by Tuning Deviation interaction, $F(6.62, 278.01) = 5.66$, $p < .001$, where participants reached higher accuracy levels at lower tuning deviations for the violin trials compared to the voice trials, indicative of a vocal generosity effect. To test the tuning deviations most affected by the vocal generosity effect, we ran $t$-tests for every tuning deviation level comparing voice and violin trials within musicians and nonmusicians, using a Bonferroni correction (adjusted $\alpha = .0025$). We found that, for nonmusicians, the tuning deviations that showed a significant difference between voice and violin judgments were 30, 40, and 50 cents, all $t(21) >= 3.47$, $p <= .0025$. For musicians, the significant differences were at 30 and 40 cents, all $t(21) >= 3.47$, $p <= .0025$.

As a follow-up analysis, we performed a four-way ($2 \times 2 \times 2 \times 10$) mixed design ANOVA, adding the factor of Direction to the previous design. This factor measures whether the second note of the pair was below or above

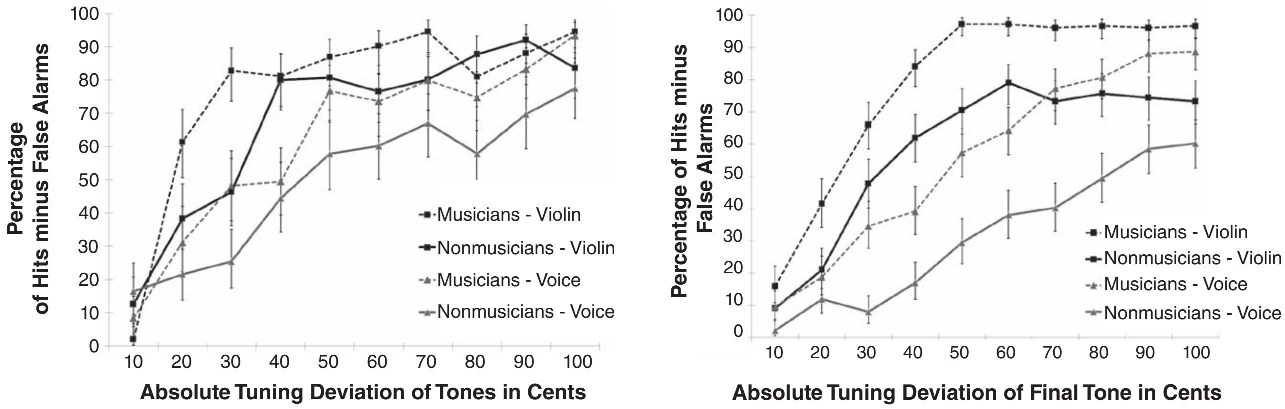


FIGURE 2. *Single Tones (left) and Melodies (right)*: The percentage of hits minus false alarms for musicians and nonmusicians in the violin and voice trials at each absolute tuning deviation, with binomial confidence intervals.

the first note. The ANOVA revealed a significant main effect of Direction, $F(1, 42) = 4.78$, $p = .035$, where participants were less accurate when the first single tone was followed by a lower note than when followed by a higher note. This effect was stronger in the voice than in the violin, but the interaction was not significant, $F(1, 42) = 2.029$, *n.s.* All previous significant main effects and interactions reported above were found, as well, but no other significant effects were found.

MELODIES

Two mixed design ANOVAs were then performed, since the factors tested in the melodic condition did not cross evenly with one another. These factors were Music Experience (between-subjects; musicians versus nonmusicians), Instrument (within-subjects; violin versus voice), Final Tone Direction (within-subjects; flat versus sharp deviation of the final tone relative to perfect tuning), Melody Direction (within-subjects; ascending versus descending melodies), Melody Type (within-subjects; ordered scales versus broken scales), and Tuning Deviation (within-subjects; absolute tuning deviation in cents from the target frequency, 10 levels).

First, a four-way ($2 \times 2 \times 2 \times 10$) mixed design ANOVA was performed to compare the factors of Music Experience, Instrument, Final Tone Direction, and Tuning Deviation. The significant main effect of Music Experience, $F(1, 42) = 22.36$, $p < .001$, showed that musicians were consistently more accurate in their responses compared to nonmusicians across all conditions. The ANOVA revealed a significant Music Experience by Tuning Deviation interaction, $F(4.44, 186.28) = 3.63$, $p = .005$, showing that musicians' response accuracy was considerably better than nonmusicians' for mid-range tuning deviations. Moreover, musicians' judgments reached ceiling at smaller tuning deviations than nonmusicians' judgments, which never exceeded eighty percent accuracy even at deviations of +/− 100 cents. This is partially due to the higher false-alarm rate found in nonmusicians (mean false alarms = 18%) than nonmusicians (mean false alarms = 5%), but the difference is also present in the hits rate (mean hits at 100 cents deviation = 84% for nonmusicians, 98% for musicians). We found a main effect of Instrument, $F(1, 42) = 63.91$, $p < .001$, where participants were more accurate in the violin trials compared to the voice trials, showing that the vocal generosity effect is present in the melodies as well. There was also a main effect of Final Tone Direction, $F(1, 42) = 12.11$, $p = .001$, as participants showed better accuracy at detecting out-of-tune or different notes when the final tone was sharp.

The ANOVA also yielded a main effect of Tuning Deviation, $F(4.44, 186.28) = 189.21$, $p < .001$, where participants did significantly better at judging mistuned final tones as different as the deviation in cents increased. A significant Instrument by Tuning Deviation interaction, $F(5.30, 222.70) = 19.09$, $p < .001$, showed that this tendency was greatly increased in the violin trials, where participants reached greater accuracy at lower tuning deviations and their responses reached a ceiling effect with high accuracy values, whereas tuning deviations in voice trials were more difficult to detect, especially for mid-range tuning deviations. Figure 2b shows this effect for each group. To test the tuning deviations most affected by the vocal generosity effect, we ran *t*-tests for every tuning deviation level comparing voice and violin trials results within each group using a Bonferroni correction (adjusted $\alpha = .0025$). We found that, for nonmusicians, the tuning deviations that showed a significant difference between voice and violin judgments were 30 to 80 cents tuning deviation, all $t(21) >= 3.47$, $p <= .0025$. Musicians showed differences on exactly the same tuning deviations as nonmusicians, all $t(21) >= 3.47$, $p <= .0025$.

We also found an Instrument by Final Tone Direction by Tuning Deviation interaction, $F(7.37, 309.42) = 3.60$, $p = .001$, such that participants' responses in the voice trials (but not the violin trials) were more accurate when the final tone was sharp compared to flat, and this effect was especially pronounced for mid-range tuning deviations. This effect is shown in Figure 3. Furthermore, the significant interaction between Instrument, Music Experience, and Final Tone direction, $F(1, 42) = 11.87$, $p = .001$, showed that musicians' responses were more accurate when the voice trials (but not the violin trials) had a sharp final tone than a flat final tone. No other interactions were found to reach significance.

A five-way ($2 \times 2 \times 2 \times 2 \times 2$) ANOVA was also performed on the factors of Music Experience, Instrument, Melody Type, Melody Direction, and Final Tone Direction. This ANOVA included all factors, with the exception of Tuning Deviation, to evaluate how the vocal generosity effect interacts with the type of melodic information. In addition to the main effects and interactions reported above, we found a significant main effect of Melody Type, $F(1, 42) = 6.65$, $p = .01$, where participants showed more accuracy in judging the tuning of the final note when the melody was an ordered scale than when it was a broken scale. Furthermore, the significant Music Experience by Melody Type interaction, $F(1, 42) = 8.45$, $p = .006$, showed that musicians were more accurate than nonmusicians in both scales and broken scales and did not show a particular sensitivity to one or the other, whereas nonmusicians did significantly better for the ordered scales than broken scales.

We also found a significant Melody Direction by Final Tone Direction interaction, $F(1, 42) = 53.58$, $p < .001$, where participants were more accurate at detecting tuning inaccuracies of sharp final tones in the descending melodies, but were more accurate with flat final tones in the ascending melodies. The Melody Direction by Final Tone Direction by Instrument interaction, $F(1, 42) = 94.30$, $p < .001$, shows that this effect was only consistent in the voice trials. These results are displayed in Figure 4. The significant Melody Type by Melody Direction by Final Tone Direction interaction, $F(1, 42) = 30.67$, $p < .001$, demonstrated that this effect was consistent in broken scale trials, but only partially so in ordered scales. Participants were significantly more accurate in judging sharp final tones in both descending broken and ordered scales, but were only more accurate in judging flat final tones in broken scales. There was no difference between judging sharp and flat final tones in ascending ordered scales.

We also found a significant Music Experience by Melody Type by Melody Direction by Instrument interaction, $F(1, 42) = 8.36$, $p = .006$, where nonmusicians' tendency to be significantly worse for broken scales was especially marked in the voice trials. They showed an even greater reduced accuracy for ascending broken scales, whereas musicians showed a significantly decreased accuracy for descending broken scales only in the voice trials. A significant Music Experience by Melody Type by Melody Direction by Final Tone Direction interaction, $F(1, 42) = 34.33$, $p < .001$, shows that this last effect held only when the final tone was flat. Finally, the ANOVA also revealed an Instrument by Melody Type by Melody Direction by Final Tone Direction interaction, $F(1, 42) = 23.01$, $p < .001$, showing that in the voice trials, participants

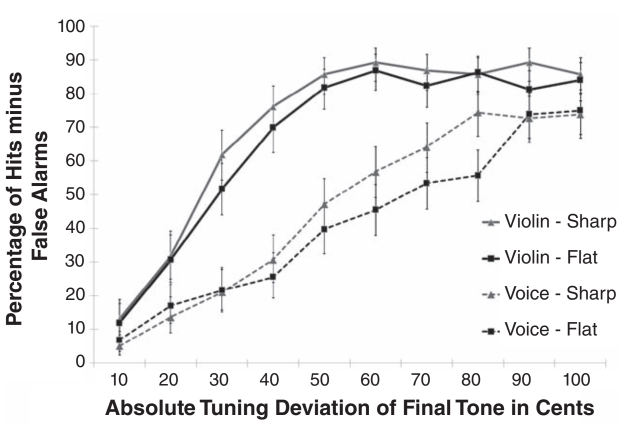


FIGURE 3. *Melodies*: The percentage of hits minus false alarms for the violin and voice trials when the final tone was either flat or sharp at each absolute tuning deviation, with binomial confidence intervals.

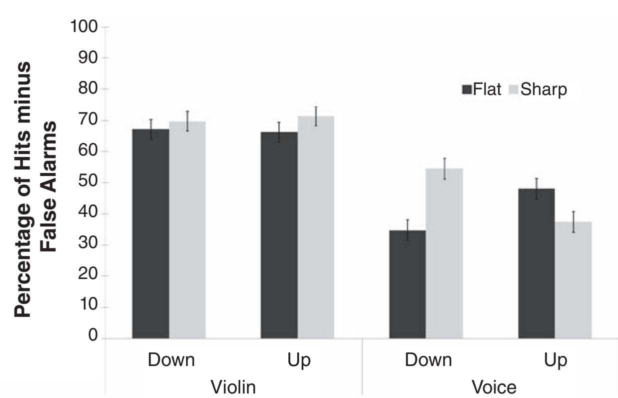


FIGURE 4. *Melodies*: The percentage of hits minus false alarms for the violin and voice trials when the final tone was either flat or sharp for each melody direction, with binomial confidence intervals.

did significantly worse for the broken scales only if the final tone was flat and the melodies were descending or if the final tone was sharp and they were ascending, otherwise the results were similar to the ordered scales, although all results were less accurate than in the violin trials.

#### COMPARISON ANALYSIS

We also tested the differences between single-tone and melodic contexts. We ran a four-way ($2 \times 2 \times 2 \times 10$) mixed design ANOVA using the hits minus false alarm rate comparing Music Experience, Context (within-subjects; single tones versus melody), Instrument, and Tuning Deviation. The ANOVA yielded a significant main effect of Context, $F(1, 42) = 7.76$, $p = .008$, where participants were less accurate in the melodic context than in the single tone context. We also found a significant interaction between Context and Music Experience, $F(1, 42) = 8.07$, $p = .007$, showing that the lower accuracy for melodic contexts only holds for nonmusicians; musicians show no differences due to context. All previous significant main effects and interactions reported in the melodic and single tone conditions were found here as well, but no other new significant effects were found.

A final analysis tested for order effects. This was implemented as a five-way ($2 \times 2 \times 2 \times 2 \times 10$) mixed design ANOVA, replicating the design from above but including a between subjects factor coding for whether the single tone trials were run in the first or second block. There was no main effect of Block Order, but there was a significant interaction between Block Order and Instrument, $F(1, 40) = 5.98$, $p = .02$. The vocal generosity effect was diminished in the second block of the experiment, regardless of which block was run first, but was still significantly different. Across all conditions, participants were 27% more likely to detect a mistuned note in the violin than in the voice in the first block, and 17% more likely to detect a mistuned note in the violin than in the voice in the second block (Tukey HSD = 6.75%). No additional main effects or interactions were found, other than those already mentioned above.

#### CATEGORIZATION ANALYSES

To assess the dividing line between categorically different notes, rather than the perception of any noticeable tuning errors, we ran a modified version of the four-way comparison analysis presented above. In this new analysis, responses of "Out-of-Tune" were combined with responses of "Same Note," rather than with responses of "Different Note," as before. Hits and false alarms were calculated across the new grouping criterion, and the results are shown in Figure 5. This was analyzed as a four-way ($2 \times 2 \times 2 \times 10$) mixed design ANOVA, comparing Music Experience, Context (within-subjects; single tones versus melody), Instrument, and Tuning Deviation.

We found a similar pattern of significant effects as in the previous analyses. There was still a large and significant effect of Instrument, $F(1, 42) = 192.72$, $p < .001$, reflecting a vocal generosity effect. Listeners were more likely to decide that a note crossed a category boundary when it was played on a violin than when it was sung. There was also a significant interaction between Instrument and Tuning Deviation, $F(5.87, 246.55) = 30.59$, $p < .001$, as before. The main effects of Music Experience, $F(1, 42) = 11.50$, $p = .002$, Context, $F(1, 42) = 25.65$, $p < .001$, and Tuning Deviation, $F(3.65, 153.44) = 276.73$, $p < .001$ all showed significant effects, and there were numerous significant higher-level interactions, including the four-way interaction, $F(5.88, 246.97) = 2.62$,

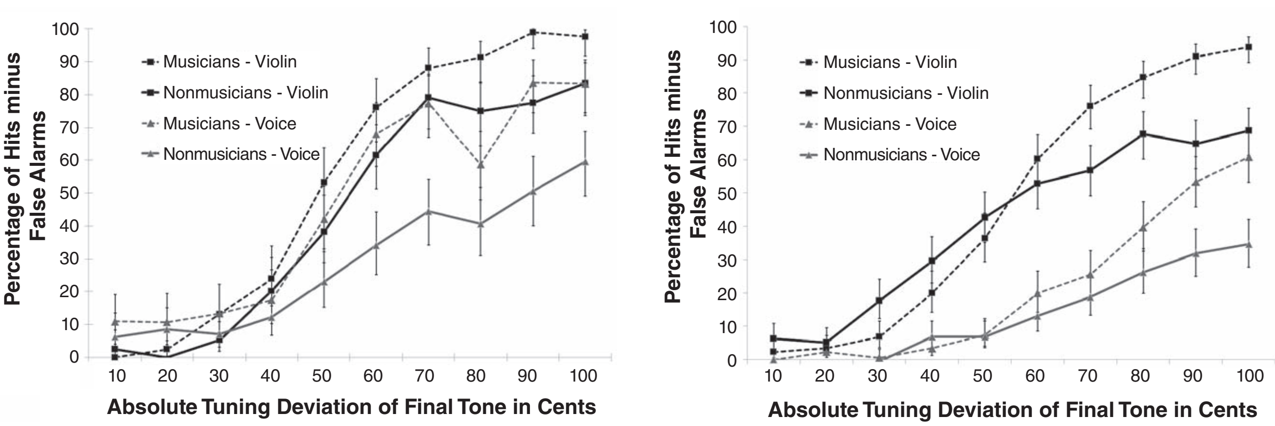


FIGURE 5. *Single Tones (left) and Melodies (right)*: A categorization analysis, combining out-of-tune responses with same note responses, rather than with different note, to evaluate perception of category boundaries. Each panel shows the percentage of hits minus false alarms for musicians and nonmusicians for the violin and voice trials at each absolute tuning deviation, with binomial confidence intervals.

$p = .018$. Just as in previous analyses, there was no interaction between Music Experience and Instrument. Within-context ANOVAs, run separately on single-tone and melody data, showed that the pattern of significant effects did not vary between the two contexts: The vocal generosity effect held for category judgments in both contexts.

## Discussion

There was a very clear vocal generosity effect present for both musicians and nonmusicians, across both single tones and melodies. For the same degree of mistuning, a note was more likely to be judged as in-tune when it was sung compared with when it was played on a violin.[2] This vocal generosity effect was especially prominent during mid-range tuning deviations, which was expected from the shape of the data (due to floor and ceiling effects). In this mid-range, participants were approximately 40% more likely to judge a sung note as in-tune than its violin equivalent. Across conditions, sung tones needed to be mistuned by at least 60 cents to be reliably labeled as out-of-tune, whereas violin tones only needed to be mistuned by just over 30 cents. Many other factors affected the tuning judgments. However, the vocal generosity effect was found across all other conditions, and created a larger effect on tuning judgments than any other experimental variable, including melodic context and musical experience.

The vocal generosity effect tended to diminish in size across the course of the session, with larger differences between judgments for violin and voice conditions in the first block than in the second block, regardless of the order in which the blocks were presented. This may be due to participants becoming more familiar with the experiment, or to them beginning to make more explicit comparisons between violin and voice trials. However, even in the second block, the vocal generosity effect was still quite large; listeners were still 30% more likely to judge a sung note as out-of-tune for mid-range mistunings, which is about 75% of the original effect size.

One particularly interesting component of the vocal generosity effect was found in the melodic contexts. Listeners were more likely to describe a final note as in-tune when the direction of the mistuning was in the same direction as the melody, but only for the sung melodies. In a rising melody, listeners were forgiving of a sharp final note, which increased the size of the final interval, but were more forgiving of flat final tones in the context of a falling melody. In other words, people were more accurate at detecting mistunings in singing when they contracted the final interval than when they expanded it.

The vocal generosity effect applies to category judgments as well. When the results were analyzed to examine the detection of category boundaries between different notes, rather than tuning judgments, there was still a significant difference between violin and voice; listeners were up to 40% more likely to judge a vocal tone as being a member of the same note category. Across contexts, violin pitches needed to be mistuned by about 50 cents to be judged as a fundamentally different note, whereas vocal tones needed at least 90 cents of mistuning to be reliably judged to be in a different category, and even a note one full semitone different would not be judged as a different note in some cases. The perceived category boundaries are distinct from the acceptable tuning ranges, showing that there is a range in which listeners perceived a note as out-of-tune, but fundamentally part of the same note category. Interestingly, in general, a mistuning that is perceived as the same note but out-of-tune on the violin is perceived as in tune for the voice, and a mistuning that is perceived as the same note but out-of-tune for the voice is perceived as a completely different note category for the violin. We should point out, however, that while these analyses can inform us of the perception of pitch categories, the experiments are not designed to speak to the psychological phenomenon of categorical perception. Although the presence of a region perceived as the same note but mistuned is suggestive of the absence of categorical perception (as listeners tend to notice a within-category distinction), further and specially designed experiments would be necessary to ascertain the presence or absence of this effect.

### THE CAUSES OF THE VOCAL GENEROSITY EFFECT

The present results replicate and extend the vocal generosity effect shown in a previous study (Hutchins & Peretz, 2012) that used nonmusicians to produce the sung tones, and compared these with a synthesized voice. Because of the difference in stimuli and the addition of melodic contexts in the present experiment, we believe that the vocal generosity effect generalized across singers, comparisons, and contexts. It seems to be a characteristic of all listeners to be particularly forgiving of vocal tuning errors.

There are two possible causes of the vocal generosity effect: It may be driven by acoustic factors that impede our ability to determine the precise pitch of a vocal tone, or there may be top-down influences that make us less

[2] There are many different ways to refer to these types of results. For clarity's sake, we remind the reader that greater accuracy is synonymous with fewer "in-tune" responses, indicating less forgiveness of mistuning. Lower accuracy is synonymous with more "in-tune" responses, indicating more forgiveness of mistuning in that condition. These different terminologies are used interchangeably throughout the discussion.

sensitive to a vocal tone's pitch precision (or more tolerant to imprecision) *after* it is identified as a voice. Acoustic features such as vibrato, steady-state timbre, jitter, shimmer, pitch variability, and its long-term spectral envelope, as well as features of the onset and offset of the tone all serve to mark a tone as a voice, and are different from those features for notes played by other instruments. Any or all of these features, as well as combinations thereof, may serve to drive the greater range of acceptable tuning found in the voice than other instruments.

Vibrato in particular is a difference worthy of consideration, as the sung tones in our experiment contained considerably more vibrato than the violin tones. Vibrato is manifested as a periodic variation in pitch, but it is generally not perceived as a change in pitch. Rather, it is heard as a part of the note's characteristic quality. Even for tones with very large vibrato amplitudes, the pitch is perceived to be at the mean of the fluctuating pitch (Shonle & Horan, 1980; Sundberg, 1978); this is true for perception of the pitch of string instruments as well (Brown & Vaughn, 1996). Both violin and voice typically have vibrato as part of skilled performance (Mellody & Wakefield, 2000; Sundberg, 1978), and singers use vibrato with a considerably greater amplitude than violinists (Mellody & Wakefield, 2000; Prame, 1997), often up to ± 100 cents or more (Prame, 1997), although certain styles of vocal music exist that employ considerably less vibrato than this (e.g., early music, barbershop).

Although Sundberg (1982, 1987) reported that vibrato does not affect the certainty with which a note is perceived, other work has shown that it can affect the range of acceptable tuning (van Besouw et al., 2008) by increasing the total range by about 10 cents among trained musicians when judging a synthesized tone. Many musicians believe that vibrato is an effective way of hiding tuning errors (Yoo, Sullivan, Moore, & Fujinaga; 1998). The pitches of tones with vibrato are also less quickly identified than those without (Yoo et al., 1998). However, this same study found that vibrato did not interfere with the listeners' final determination of pitch; listeners could make very fine tuning judgments comparing tones with and without vibrato. Shonle & Horan (1980), too, found quite precise tuning judgments, even in the case of very large vibrato amplitudes, but also found greater variability is those judgments associated with larger vibrato amplitudes.

Because of the greater variability it seems to evoke in tuning judgments, vibrato may be an underlying factor in the vocal generosity effect. Indeed, the voice typically uses more vibrato than just about any other instrument, and the singer in this study tended to employ vibrato with a wider extent and slower rate than most singers (Prame, 1997; Sundberg, 1994). However, comparisons between this experiment and Hutchins and Peretz (2012) seem to contradict that interpretation. In the latter, the vocal generosity effect was almost identical in size to the one found here, despite the much lower extent of the vibrato in the tones produced by the untrained singers (mean = ± 11.56 cents, S.D. = ± 4.16 cents). The fact that the vocal generosity effect is consistent between high and low vibrato stimuli indicates that this is probably not simply an effect of vibrato, although a controlled experiment would be necessary to verify this.

Another type of explanation for this effect posits that it is driven by cognitive influences specific to voices. That is, once we recognize a tone as being a voice, we may be more forgiving of mistunings because it is a voice. For example, it may be an effect of implicit learning. We are generally exposed to far more mistuned vocal notes than instrumental notes in our experiences with music (although the magnitude of this trend may be changing, due to more widespread use of auto-tuning in the past decade). The greater diversity of tuning in voice may have implicitly led listeners to accept a wider range of tuning as acceptable in the voice than in other instruments. A similar type of explanation concerns the more common use of the voice for speech.[3] In speech, there are no precise requirements for tuning. Thus, the listener may generalize from speech to music and consciously or unconsciously lower their requirements for acceptable tuning, operating partly in a "speech mode"; some recent evidence shows that normal listeners are more sensitive to pitch changes in instrumental tones than in spoken syllables (Tillmann et al., 2011), paralleling the instrument/singing differences found here and in Hutchins and Peretz (2012).

Further testing can shed light on the ultimate cause of this effect. For example, if we were to ask listeners to make tuning judgments while somehow rendering them unaware that they were hearing a voice, we would expect to see a similar range of acceptable tuning if the vocal generosity effect were driven by acoustical factors. On the other hand, the range of acceptable tuning would be more similar to that of instrumental tones if the effect were driven by awareness of the tone as a voice. Conversely, an instrumental tone that evokes the perception of a vocal sound (à la Peter Frampton, for example), should show a larger range of acceptable tuning in this case as well. In addition, we would expect to see a good correlation between measureable acoustic features and acceptable tuning ranges if the former are ultimately

[3] Thanks to an anonymous reviewer for suggesting this as a possible cause.

driving the effect, but not if it is being caused by higher-level processing.

IMPLICATIONS FOR SINGING RESEARCH

One of the major implications of our findings concerns research on the singing abilities of the general population. In general, researchers in this field have tended to categorize singers as good or poor based on whether their average tuning errors fall above or below a particular criterion, generally either 50 or 100 cents of error (Dalla Bella et al., 2007; Demorest & Clements, 2007; Hutchins & Peretz, 2012; Pfordresher & Brown, 2007). In most cases, however, this criterion was chosen based on assumptions derived from the theory of Western music, rather than on experimental data about the range of pitch, which is generally considered to be acceptably in tune. The current study attempted to rectify this oversight, and, in addition to Hutchins and Peretz (2012), represents the only study that explicitly looks at the range of acceptable tuning for vocal tones using a controlled experimental method. This study includes judgments of both musically trained and untrained listeners, making it a solid empirical basis for future studies of tuning in singing.

The vocal generosity effect found here represents a strong challenge to any tuning criteria based on the structure of Western music, rather than empirical methods. A structurally based criterion would posit that tuning should be evaluated the same way no matter which instrument is being performed. However, we have demonstrated that there are tunings (around 40 cents off) that are widely perceived as in-tune for the voice, but as out-of-tune for other timbres. Therefore, a structurally based criterion, which makes no allowances for different timbres, would disagree with most listeners in one case or the other, no matter how it is set.

The maximum allowable deviation considered to be in-tune varied depending on the condition and the experience of the listener. In general, though, the results of the current study fell in line with those from Hutchins and Peretz (2012), and tend to support a criterion of 50 cents to be considered in tune, although the result showing that nonmusicians will generally hear 80 cent deviations in melodic sung tones as in tune could lend some support for a wider acceptability criterion. Tuning judgments can be highly variable, and singers in the real world may also be using deliberate strategies for enhancing expressivity within the context of songs, leading to variations in preferred tuning not captured by our experimental manipulations. Despite these caveats, regular patterns in tuning judgments of singing do emerge, and these patterns can be used as a basis for criteria in acoustical studies of singing production.

These results are meant to illustrate whether a particular note is in-tune or out-of-tune; the question of whether a particular singer is in-tune or out-of-tune is a somewhat different and more complicated question. A singer's abilities may vary across different pitch ranges, melodies, or even musical styles, and certain patterns of mistunings within a melody may be perceived as better than others (e.g., correcting prior errors). However, as a first approximation, it seems reasonable to assume that out-of-tune singers will produce more out-of-tune notes. Thus, the current data represent a good starting point for an empirically based evaluation of singing ability.

The vocal generosity effect demonstrates that singing should not be held to the same standards as instrumental music, and provides an explanation for the tendency of singers to be less in-tune than instrumentalists (Geringer, 1978; Prame, 1997). Furthermore, this effect implies that other work on music cognition that uses instrumental stimuli may also not generalize to the voice, including studies that use instrumental stimuli to examine topics such as memory for music, attention, absolute pitch, and the emotional response to music. If the vocal generosity effect is driven by specific acoustical variables (such as vibrato or others), then this means that these variables need to be taken into account when constructing experimental stimuli, even when the experiment does not specifically involve the voice. On the other hand, if this effect is driven by top-down cognitive influences, this implies that the entire processing of vocal and non-vocal tones may be different.

Two other specific fields within music cognition should also take heed of the vocal generosity effect; namely research in amusia and research in the relationship between language and music. In the former, the general lack of sensitivity to pitch errors (Hyde & Peretz, 2004; Peretz et al., 2002) is similar to the specific lack of sensitivity to vocal pitch errors in the general population. Because of this, interactions between pitch accuracy and timbre should be taken into account in this research. The vocal generosity effect should also be taken into account when comparing language to music, as instrumental music is not entirely comparable to vocal music. The conclusions drawn from this research could vary significantly depending on which is used as the representative of music in the comparison, and it may be worthwhile in some cases to compare all three cases (speech, singing, and instrumental music).

Finally, the results presented here also have implications for real-world musicians and those who work with music. For example, our results of acceptable tuning across different contexts can provide guidelines for singers and audio engineers who want to manipulate the tuning of a song – this works demonstrates that vocal and instrumental parts need not be auto-tuned to the same degree to achieve the same effect.

## Conclusion

This work confirms the vocal generosity effect initially found in Hutchins and Peretz (2012), and generalizes it to different types of voices and contexts. As noted in Hutchins and Peretz (2012), this seems to be related to the small tuning errors (10–30 cents) consistently found among what are generally considered to be good singers. Listeners generally do not hear such small tuning deviations as errors. Our results show that vocal tones need around 50 cents of mistuning on average to be perceived as errors, but this is highly dependent on whether they are judging the tuning by comparing to the same note, or comparing to a musical context. Musicians have a smaller range of acceptable tuning than nonmusicians, but both groups show a strong vocal generosity effect. These results provide a basis for criteria for accuracy measurements in vocal-pitch matching experiments, demonstrate the important role of stimulus timbre in tuning judgments, and help to answer the important question raised in the introduction, "What counts as in-tune?"

## Author Note

This work was supported by grants from the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada and a Canada Research Chair in neurocognition of music to I. P. We would like to thank the action editor, Peter Pfordresher, and three anonymous reviewers for their insightful comments and suggestions for this article.

*Correspondence concerning this article should be addressed to* Dr. Sean Hutchins, Rotman Research Institute, Baycrest Hospital, 3560 Bathurst Street, Toronto, Ontario, Canada, M6A 2E1. E-MAIL: Sean.Hutchins@gmail.com

## References

AMIR, O., AMIR, N., & KISHON-RABIN, L. (2003). The effect of superior auditory skills on vocal accuracy. *Journal of the Acoustical Society of America, 113,* 1102–1108.

BRADSHAW, E., & MCHENRY, M. A. (2005). Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *Journal of Voice, 19,* 431–439.

BROWN, J. C., & VAUGHN, K. V. (1996). Pitch center of stringed instrument vibrato tones. *Journal of the Acoustical Society of America, 100,* 1728–1735.

BURNS, E. M., & WARD, W. D. (1978). Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of musical intervals. *Journal of the Acoustical Society of America, 63,* 456–468.

BURNS, E. M., & WARD, W. D. (1982). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The psychology of music* (pp. 241–269). New York: Academic Press.

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.

DEMOREST, S. M., & CLEMENTS, A. (2007). Factors influencing the pitch-matching of junior high boys. *Journal of Research in Music Education, 55,* 190–203.

GERINGER, J. M. (1978). Intonational performance and perception of ascending scales. *Journal of Research in Music Education, 26,* 32–40.

HALPERN, A. R., & ZATORRE, R. J. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception and Psychophysics, 26,* 384–395.

HUTCHINS, S., & CAMPBELL, D. (2009). Estimating the time to reach a target frequency in singing. *Annals of the New York Academy of Sciences, 1169,* 116–120.

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

HYDE, K., & PERETZ, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15,* 356–360.

KRUMHANSL, C. L., & IVERSON, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 739–751.

MELLODY, M., & WAKEFIELD, G. H. (2000). The time-frequency characteristics of violin vibrato: Modal distribution analysis and synthesis. *Journal of the Acoustical Society of America, 107,* 598–611.

MORAN, H., & PRATT, C. C. (1926). Variability of judgments of musical intervals. *Journal of Experimental Psychology, 9,* 492–500.

PERETZ, I., AYOTTE, J., ZATORRE, R., MEHLER, J., AHAD, P., PENHUNE, V., & JUTRAS, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33,* 185–191.

PFORDRESHER, P. Q., & BROWN, S. (2007). Poor-pitch singing in the absence of "tone deafness." *Music Perception, 25,* 95–115.

PRAME, E. (1997). Vibrato extent and intonation in professional Western lyric singing. *Journal of the Acoustical Society of America, 102,* 616–621.

RAKOWSKI, A. (1976). Tuning of isolated musical intervals. *Journal of the Acoustical Society of America, 59,* S50.

RAKOWSKI, A. (1990). Intonation variants of musical intervals in isolation and in musical contexts. *Psychology of Music, 18,* 60–72.

RAKOWSKI, A., & MISKIEWICZ, A. (1985). Deviations from equal temperament in tuning isolated musical intervals. *Archives of Acoustics, 10,* 95–104.

SCHELLENBERG, E. G. (2001). Asymmetries in the discrimination of musical intervals: Going out-of-tune is more noticeable than going in-tune. *Music Perception, 19,* 223–248.

SEASHORE, C. E. (1938). *Psychology of music.* New York: Dover.

SEMAL, C., & DEMANY, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America, 89,* 2404–2410.

SEMAL, C., & DEMANY, L. (1993). Further evidence for an autonomous processing of pitch in auditory short-term memory. *Journal of the Acoustical Society of America, 94,* 1315–1322.

SHONLE, J. I., & HORAN, K. E. (1980). The pitch of vibrato tones. *Journal of the Acoustical Society of America, 67,* 246–252.

SIEGEL, J. A., & SIEGEL, W. (1977). Categorical perception of tonal intervals: Musicians can't tell sharp from flat. *Perception and Psychophysics, 21,* 399–407.

SUNDBERG, J. (1978). Effects of the vibrato and the 'singing formant' on pitch. *Journal of Research in Singing, 5,* 5–17.

SUNDBERG, J. (1979). Perception of singing. *Speech Transmission Laboratory – Quarterly Progress and Status Report, 20,* 1–48.

SUNDBERG, J. (1982). In tune or not? A study of fundamental frequency in music practice. *Speech Transmission Laboratory – Quarterly Progress and Status Report, 23,* 49–78.

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

SUNDBERG, J. (1994). Acoustic and psychoacoustic aspects of vocal vibrato. *Speech Transmission Laboratory – Quarterly Progress and Status Report, 35,* 45–67.

SUNDBERG, J., PRAME, E., & IWARSSON J. (1996). Replicability and accuracy of pitch patterns in professional singers. In P. J. Davis & N. H. Fletcher (Eds.), *Vocal fold physiology*, San Diego, CA: Singular.

TERHARDT, E. (1969). Oktavspreizung und Tonhöhenverschiebung bei Sinustonen. [Octave enlargement and pitch shift in the case of sine tones]. *Acustica, 22,* 348–351.

TILLMANN, B., RUSCONI, E., TRAUBE, C., BUTTERWORTH, B., UMILTÀ, C., & PERETZ, I. (2011). Fine-grained pitch processing of music and speech in congenital amusia. *Journal of The Acoustical Society of America, 130*, 4089–4096.

VAN BESOUW, R. M., BRERETON, J. S., & HOWARD, D. M. (2008). Range of tuning for tones with and without vibrato. *Music Perception, 26,* 145–156.

VURMA, A., & ROSS, J. (2006). Production and perception of musical intervals. *Music Perception, 23,* 331–344.

WARD, W. D. (1954). Subjective musical pitch. *Journal of the Acoustical Society of America, 26,* 369–380.

WARRIER, C. M., & ZATORRE, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception and Psychophysics, 64,* 198–207.

YOO, L., SULLIVAN, D. S., JR., MOORE, S., & FUJINAGA, I. (1998). The effect of vibrato on the response time in determining the pitch relationship of violin tones. In S.W. Yi (Ed.), *Proceedings of the 5th International Conference on Music Perception and Cognition* (pp. 209–211). Seoul, Korea: Seoul National University.