Journal of Experimental Psychology: Human Perception and Performance

Pitch Perception in Music: Do Scoops Matter?

Pauline Larrouy-Maestri and Peter Q. Pfordresher

Online First Publication, July 5, 2018. http://dx.doi.org/10.1037/xhp0000550

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Human Perception and Performance

0096-1523/18/$12.00

http://dx.doi.org/10.1037/xhp0000550

Pitch Perception in Music: Do Scoops Matter?

Pauline Larrouy-Maestri
Max Planck Institute for Empirical Aesthetics
and University at Buffalo

Peter Q. Pfordresher
University at Buffalo

Studies of musical pitch perception typically treat pitches as if they are stable within a tone. Although pitches are represented this way in notation, performed tones are rarely stable, particularly in singing, which is arguably the most common form of melody production. This paper examines how brief dynamic changes at the beginnings and endings of sung pitches, a.k.a. “scoops,” influence intonation perception. Across three experiments, 110 participants evaluated the intonation of four-tone melodies in which the third tone’s tuning could vary within the central steady-state (the asymptote), or by virtue of scoops at the beginning and/or end of the tone. As expected, listeners were sensitive to mistuning. Importantly, our results also point to unique contributions of scoops. As in the language domain, dynamic changes in a small time window are perceptually significant in music. More specifically, this study revealed the coexistence of two distinct mechanisms: sensitivity to the average pitch across the duration of the tone (assimilating the scoop), and processing the relationship of the scoop to the surrounding context. In addition to clarifying intonation perception in music, the identification of these mechanisms paves the way to cross-domain comparisons and, more generally, to the better understanding of auditory sequences processing.

Public Significance Statement
This study highlights the perceptual relevance of small pitch dynamic changes, such as the scoops performed by singers at the start and end of tones, in music perception. Listeners combine two different strategies when processing scoops: averaging of the pitch information within the larger unit and using the small unit (i.e., scoop) in relation to the inferred goals of the producer. By using music as a window to examine auditory sequence processing, this study demonstrates parallels with pitch information processing in the language domain and thus opens the door to direct comparisons.

Keywords: auditory sequence processing, music perception, pitch accuracy, scoops, singing

One of the most important functions of the auditory system is to process pitch. Without the ability to perceive pitch, the intentions or emotions of speakers could be difficult to discern (Banse & Scherer, 1996; Hellbernd & Sammler, 2016), word meanings associated with lexical tones could be misconstrued (Yip, 2002), and it would not be possible to recognize one’s favorite melody or perceive a music performance as sounding “right.”

We focus here on the use of pitch in music perception. Through enculturation, listeners become sensitive to categorical distinctions among pitch classes (Bigand & Poulin-Charronnat, 2006; Burns & Ward, 1978) and develop an internal representation of what is “correct” in terms of pitch accuracy (Larrouy-Maestri, Lévêque, Schön, Giovanni, & Morsomme, 2013; Larrouy-Maestri, Magis, Grabenhorst, & Morsomme, 2015). The foundation of our understandings for pitch perception comes from psychoacoustical studies in which pitches are typically level (unchanging within a tone). However, in practice, pitches are rarely stable throughout a tone (Larrouy-Maestri, Magis, & Morsomme, 2014a).

Consider singing, which is probably the most frequent way melodies are relayed. Even the most highly trained singers often do not start a sung pitch on the precise target fundamental frequency (Hutchins & Campbell, 2009; Mori, Odagiri, Kasuya, & Honda, 2004; Saitou, Unoki, & Akagi, 2005; Stevens & Miles, 1928). Singers typically exhibit a *scoop*: A relatively brief pitch transition toward or away from the target pitch. These scoops may be used expressively, but also represent difficulties in reaching a target

Pauline Larrouy-Maestri, Department of Neuroscience, Max Planck Institute for Empirical Aesthetics and Department of Psychology, University at Buffalo; Peter Q. Pfordresher, Department of Psychology, University at Buffalo.

This research was supported in part by a travel grant awarded to Pauline Larrouy-Maestri from the Patrimoine de l’Université de Liège and FNRS (Fond National de la Recherche Scientifique) and NSF Grant BCS-1256964 awarded to Peter Q. Pfordresher For the resources to complete this work, we sincerely thank David Poeppel. We thank the Laboratory LIST (department of signals, images, and acoustics, at the Université Libre de Bruxelles) for technical support. We are grateful to Zahra Malakotipour and Malak Sharif for assistance with data collection and to Natalie Holz for comments on an earlier version of this paper.

Correspondence concerning this article should be addressed to Pauline Larrouy-Maestri, Department of Neuroscience, Max-Planck-Institute for Empirical Aesthetics, Grüneburgweg, 14, 60322 Frankfurt-Am-Main, Germany. E-mail: plm@aesthetics.mpg.de

pitch, adjusting vocal folds, or maintaining subglottal pressure. It is important to note that scoops are deterministic (i.e., oriented toward a target pitch) and do not reflect mere noise in the motor system.

Figure 1 shows an excerpt from a trained singer performing the third phrase of "Happy Birthday to You" (which features an octave jump), in French. In this performance, the singer was instructed to sing without a specific singing technique and therefore the fundamental frequency (F0) pattern does not include pronounced vibrato. Even so, there are pronounced scoops at certain pitches. The octave jump (tone 3, the syllable "an") features a scoop up at the beginning and down at the end, reflecting its relationship to surrounding pitches. The last two syllables of "*Anniversaire*" include starting scoops that move up ("*ver*") and down ("*saire*"). In other words, this figure represents the succession of discrete tones in a musical phrase as specified in the Western tonal music system (e.g., Krumhansl, 1979; Lerdahl & Jackendoff, 1983) as well as the typical pitch fluctuations when the phrase is performed by a singer (see Appendix A).

Our concern in this paper is not with the singer, but with the listener. Namely, we consider the perceptual impact that such fluctuations have on one's perception of how well a pitch fits into the context of a tone or melody. The multitime resolution hypothesis, proposed by Poeppel (2001, 2003) for speech and supported by Teng, Tian, and Poeppel (2016) for pure tones, suggests that dynamic changes in small time windows are relevant to listeners. In language, different linguistic elements (e.g., word, phrase, sentence) are tracked and temporally integrated (Ding, Melloni, Zhang, Tian, & Poeppel, 2016; Ding et al., 2017). Using different methods, they examined the cortical activity tracking the rhythms of the syllable/words, phrases, and sentences, and observed a link between strength of the entrainment and intelligibility of the linguistic material, suggesting that neural tracking of hierarchical linguistic structures is a plausible functional mechanism for temporal integration of small units into larger ones. In other words, speech comprehension relies on the parsing and integration of information processing at different timescales.

Although it is not known whether this same process applies to music perception, a first step in this direction consists in identifying how units in music are segmented and integrated (Fritz, Poeppel, Trainor, et al., 2013). In other words, it is necessary to examine the perception/relevance of dynamic changes at small timescales as well as their integration into larger timescales or units (i.e., tones, melodies). As discussed in Fritz et al. (2013), such knowledge (i.e., nature and processes of the units) would definitely allow cross-domain comparisons of neural mechanisms. Indeed, units themselves are difficult to compare (e.g., is there a musical correspondence to a syllable?) but the cognitive processes (segmentation, merging, hierarchical structure, categorization) of specific units might be comparable across domains.

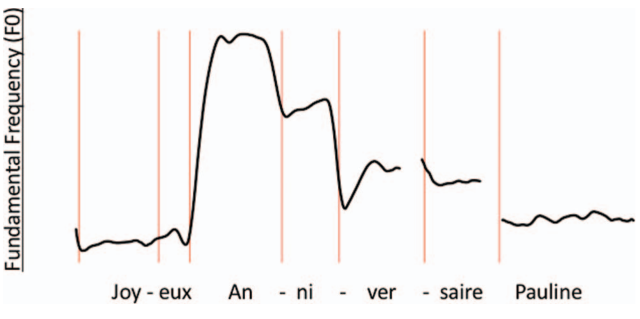

Figure 1. Illustration of the fundamental frequency (F0) of a trained singer producing a phrase from *Happy Birthday to You* in French. See the online article for the color version of this figure.

It has been shown that dynamic changes in pitch are relevant to listeners (e.g., glides between pure tones: Lyzenga, Carlyon, & Moore, 2004; glides at the end of pure tones: Wang, Tan, & Martin, 2013; frequency modulations: Gockel, Moore, & Carlyon, 2001; vibrato: van Besouw & Howard, 2009). In the psychophysical literature, dynamic changes involving some form of frequency modulation are often referred to as "glides" or "sweeps." We use the music-related term "scoop" to refer to a specific and highly frequent type of dynamic change in singing. The perceptual threshold to identify direction in frequency-modulated signals has been identified at about 20 ms (Gordon & Poeppel, 2002; Luo, Boemio, Gordon, & Poeppel, 2007). Such dynamic changes influence our perception of tone sequences (Kerivan & Carey, 1976). However, the question of the relevance of dynamic changes in music listening/appreciation remains open. Indeed, listeners are able to discriminate small pitch differences (Micheyl, Delhommeau, Perrot, & Oxenham, 2006; Moore, 1973), but small pitch deviations do not necessarily make a melody sound out-of-tune (Hutchins, Roquet, & Peretz, 2012; Warrier & Zatorre, 2002). Along the same lines, it might be that dynamic changes to pitch (i.e., scoops at the start or end of tones) are perceived but are not treated as informative, and are therefore discarded in the memory trace used for further processing (cf. Raffman, 1993). In such case, listeners would not be influenced by the presence of scoops when evaluating performances. In contrast, if listeners assimilate scoops to the closest neighbor tone or use scoops in relation to surrounding tones, the tone might not be the smallest unit on which listeners rely when evaluating performance quality.

The importance of understanding the influence of scoops on perception/appreciation goes beyond the context of music. In understanding this phenomenon, we aim to address deeper questions concerning how variability within an auditory event influences how pitch events are categorized. We consider two hypotheses as a starting point. One is that auditory processing is *statistical* in nature. That is, the listener's perception is based on the average F0 across a sung tone. The other hypothesis is that auditory processing is *teleological*, meaning that the listener uses assumptions about the singer's goals when evaluating the effect of scoops. Under this assumption, the listener focuses on a particular point in the sung tone (most likely the center) which is taken to represent the goal of the producer. The significance of vocal scoops are based on the relationship of the scoop to this central value as well as this surrounding context. In this way, the teleological hypothesis treats scoops as distinct from the way in which a scoop influences average tuning across the entire tone.

How may we distinguish these predictions? Two illustrative examples are shown in Figure 2. In Figure 2A there is an upward scoop at the start of the tone, followed by another upward scoop at the end of the tone, whereas in Figure 2B the tone is initiated with an upward scoop and ends with a downward scoop. The statistical hypothesis outlined above predicts different perception in each case, with perceived pitch being lower (and perhaps less accurate)

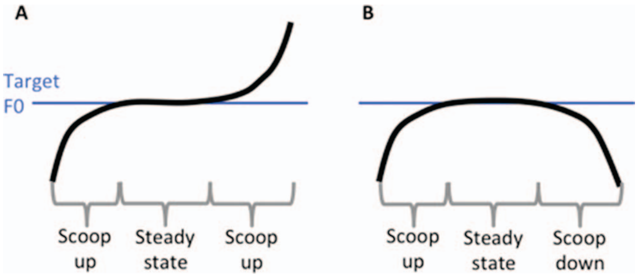

Figure 2. Schematic illustration of scoops (i.e., up vs. down) around the asymptotic steady state of sung tones. See the online article for the color version of this figure.

in Figure 2B than in Figure 2A. Note that if the asymptote is sharp, the presence of scoops such as in Figure 2B would *compensate* the deviation of the asymptote and thus enhance pitch accuracy of the tone.

By contrast, a teleological view might interpret these examples with respect to how they relate to the surrounding context. A specific hypothesis we test in this respect comes from whether scoops enhance *continuity* across successive discrete tones. For instance, both scoop patterns in Figure 2 would be considered continuous if left-side pitch was found in the middle of an ascending pattern, and if the right side pitch formed a local peak in the melodic contour. An *anticontinuous* pattern would be obtained for the obverse situation. For instance, if the scoops in Figure 2A were found in a descending overall contour, the scoops would serve to enhance the distinctiveness of the selected tone by detracting from continuation. Regularities in the preference for either situation—continuity or anticontinuity—would be in line with a multiple time-scale representation in the auditory system (e.g., Teng et al., 2016). By this logic, scoops would be processed using the small-scale system while the large-scale system processes global properties of the tone and their organization in melodies.

In this paper, we report the results of three experiments that were designed to examine the perception of pitch fluctuations within tones (i.e., scoops at the start or end of tones) when listening to melodies and to investigate the relevance of such pitch fluctuations in comparison to pitch deviations of the steady state, which we refer to as *asymptotic* tuning. In each experiment, participants listened to synthesized four-tone melodies in a vocal timbre. In certain melodies, the pitch of the third tone was manipulated so that the asymptotic F0 was mistuned and/or scoops were positioned at the start of the tone, the end of the tone, or both. Because we were concerned about the demands of explicitly matching perceived pitches to an internalized schema (e.g., by asking participants: "Was this tone/melody sung in tune?"), we adopted a pairwise comparison procedure in which participants rated which of two alternative performances was more accurate with regard to pitch.

To summarize, we addressed the following research questions:

1. Do scoops influence melodic perception, or are they treated as irrelevant?

2. If scoops influence listeners' perception, is it driven by:

   (a) The statistical average across the entire tone, or

   (b) The relationship between scoops and inferred goals of the producer (i.e., the melodic context)?

Concretely, if listeners perform statistical averaging of the tone unit, we expect to observe a preference for *compensatory* scoops over scoops enlarging the mistuning of the tone. An effect of *continuity* (i.e., relation between scoops and the surrounding tones) would support the teleological hypothesis.

General Method

Stimuli

Four melodies (shown in Figure 3) were created from synthesized vocal tones using a male timbre (Vocaloid, Zero-G Limited, Okehampton, England). Each tone was 900 milliseconds in length and generated using a synthesized articulation of the syllable /da/, including a fade in and out of 50 ms. Each melody contained 4 tones: C3 (131 Hz), D3, E3, and G3, arranged in different melodic contours using equal temperament (100 cents per semitone).

We manipulated the characteristics of the third tone of each melody (D3 for Melody 1 and E3 for Melodies 2, 3, and 4). We singled out this tone because it is far enough into the melody that listeners would interpret pitch alterations with respect to surrounding context. At the same time, we wanted to avoid any biases associated with the perception of closure or particularly high expectation that might be associated with the final tone (Pearce & Wiggins, 2006). Manipulations were implemented at two different levels, which reflect different temporal scales within a sung tone. At a broad or coarse-grained level, we manipulated the central portion or *asymptote* of the tone to be in-tune, flat, or sharp. At a finer-grained level we manipulated the presence, location, and direction of *scoops* at the start and/or end of the tone. The audio material of the three experiments is available at https://edmond.mpdl.mpg.de/imeji/collection/tKj5Oo5FLMglWzWu?q=.

Manipulations of the tones' asymptote were done using a dynamic transposition with preservation of the envelope (AudioScupt, Ircam, Paris). Sharp and flat asymptotes were ±50 cents (a half semitone or quarter tone) away from ideal equal tempered tuning, on the third tone of the melodies. This magnitude of deviation has been shown as big enough to be discriminated (Micheyl et al., 2006; Moore, 1973), yet small enough to not disrupt pitch category perception (i.e., Burns & Ward, 1978; Zarate, Ritson, & Poeppel, 2012).

Manipulations of scoops (on the fine-grained timescale) incorporated the same dynamic transposition algorithm as used for asymptotic manipulations. It was a major concern to us that the manipulated scoops were representative of the kind of dynamic changes to pitch that a human singer might make. Otherwise, obtained effects related to scoops could simply reflect a listener's

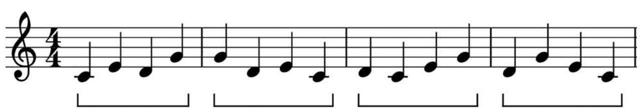

Figure 3. Musical notation of the four melodies used in the present experiments. Each melody is one measure long; thus, each bracketed set of tones constitutes an independent stimulus melody.

ability to detect unnatural or artificial perturbations to the pitch sequence. We based the timing and extent of scoops on a large-scale analysis of scoops in a previously published dataset (Pfordresher & Mantell, 2014). In that study, singers representing a wide range of singing skills imitated four-tone melodies similar to the ones used here. Based on the analyses of pitch fluctuations (see Appendix A), synthesized scoops in the current data set were inserted into the initial or final 220 ms of tones—just under 25% of the total duration. Transitions were based on an exponential curve from the start of the scoop to the asymptote, as illustrated in Figure 2. The direction of scoops was defined relative to the asymptote. Thus, for scoops at the beginning of a tone, an "upwards" scoop starts lower than the asymptote, and a "downwards" scoop starts higher than the asymptote. Figure 2 shows upward starting scoops in both panels. By contrast, an upward scoop at the end of the tone proceeds up in pitch from the asymptote and thus ends at a higher pitch. The ending scoop in Figure 2A would be termed upward, whereas the ending scoop in Figure 2B would be termed downward.

Our analysis of scoops in human singing suggested that singers vary primarily in the extent of their scoops—the difference between the most extreme pitch in the scoop and the asymptote—and not in the duration between the extreme pitch and the asymptote (see Appendix A). Based on this result, we speculated that scoop magnitude may be an important variable for listeners and we were thus careful to base all end points on data from human singing. For this reason, scoop magnitude was not held constant across all scoop types; we address how scoop magnitude may influence responses in regression analyses. Table 1 specifies the average magnitude of scoops examined in the dataset of Pfordresher and Mantell (2014) and used in all experiments of the present study. As can be seen, a general rule is that scoop magnitudes are larger when scoops preserve continuity between tones. For example, when there is upward pitch motion (tone 2 is a lower pitch than tone 3), an upward scoop at the start of tone 3—consistent with this pitch motion—is 74 cents, whereas a downward scoop is only 58 cents. These differences are magnified for scoops at the end of tones.

Procedure

Participants were asked to evaluate the pitch accuracy of the manipulated melodies with a pairwise comparison paradigm. Specifically, listeners compared all possible pairs of performances

Table 1
Magnitude of Scoops Used in All Experiments

		Relation of adjacent tone	
Scoop position	Scoop direction	Higher pitch	Lower pitch
Start of tone	Upwards	68	74
	Downwards	71	58
End of tone	Upwards	115	57
	Downwards	86	159

Note. All units are in cents, representing the absolute values of the difference between the pitch at the start or end and the asymptote. For starting scoops, the adjacent tone refers to the pitch of tone 2 in the melody, whereas for ending scoops, the adjacent tone refers to the pitch of tone 4.

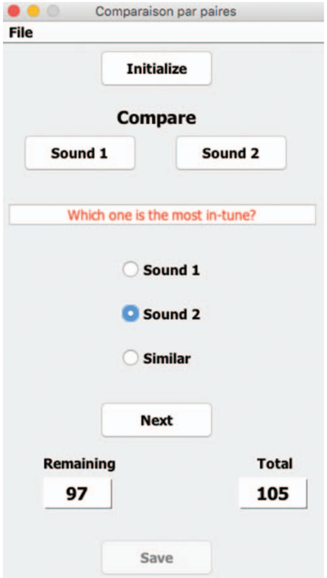

Figure 4. Screen shot of the experimental interface. See the online article for the color version of this figure.

across the entire stimulus set and reported whether the first or the second sequence was the most in-tune or if the sequences were equally in-tune (3 alternatives). Thus, participants were exposed to every possible pair of sequences within a single melody in every order (number of trials for a single melody = $N(N–1)/2$, where N refers to the total number of sequences resulting from pitch manipulations). This paradigm has been used in previous studies for evaluations of healthy versus disordered voices (Kacha, Grenez, & Schoentgen, 2005) and the perception of intonation in operatic voices (Larrouy-Maestri, Magis, & Morsomme, 2014b; Larrouy-Maestri, Morsomme, Magis, & Poeppel, 2017). Note that such a procedure is not a systematic comparison to an "ideal" sequence in same/different tasks (e.g., Hyde & Peretz, 2004; Marmel, Tillmann, & Dowling, 2008; Stalinski, Schellenberg, & Trehub, 2008) but allows a direct comparison of the effect of contrasted pitch manipulations (i.e., *scoops* and *asymptote* tuning in our case) on listeners' evaluation of melodic performances.

Each trial was self-paced. Participants initiated each melody in a pair by pressing one of two buttons on a customized graphical user interface programed in Java (see Figure 4). Stimuli were presented over Sennheiser HD 280 Pro Headphones at comfortable intensity. After listening, participants reported their preference by answering the question (presented at the bottom of the interface): "Which one is the most in-tune?" by pressing one of the three buttons (i.e., Sound 1, Sound 2, Similar). After saving their choice (i.e., Next button), a new trial was presented, that is, another pair of sung performances to compare.

Of course, an experiment that combines all levels associated with our manipulations of pitch (3 asymptotic × 5 scoop manipulations, which includes no scoops), across all four melodies would lead to a prohibitively long experiment. Thus, we kept the length of sessions manageable by dividing manipulations into three related experiments, summarized in Table 2. Each experiment was designed to address a specific aspect related to the effect

Table 2
Illustration of the Manipulations Proposed in the Three Experiments

Asymptotic deviation	Scoop at the start	Scoop at the end: None	Up	Down
None	None	Exp 1 & 2	Exp 1 & 2	Exp 1 & 2
	Up	Exp 1 & 2	Exp 2	Exp 2
	Down	Exp 1 & 2	Exp 2	Exp 2
+50 cents	None	Exp 1	Exp 1	Exp 1
	Up	Exp 1	Exp 3	Exp 3
	Down	Exp 1	Exp 3	Exp 3
−50 cents	None	Exp 1	Exp 1	Exp 1
	Up	Exp 1	Exp 3	Exp 3
	Down	Exp 1	Exp 3	Exp 3

Note. Exp. = experiment.

of scoops on pitch perception, so that the three experiments taken together provide the fullest account.

Data Analyses

Ratings. The task was scored in three steps. For each participant and each block (i.e., melody), all stimuli were initialized to a score of zero. Every time a participant indicated preference for one stimulus over another (i.e., heard it as being more in-tune), the score for the preferred stimulus was increased by one, and the nonpreferred stimulus remained at its current score. If both stimuli of the pair were judged to be "equal," the total score of both stimuli was increased by 0.5 points. The total score for each stimulus was computed by accumulating points over trials, ranging from 0 (i.e., stimulus never selected as the most in-tune) to $N - 1$ (i.e., stimulus always selected as the most in-tune). As a consequence, this proposed rating procedure allows ranking the manipulated sequences from the most out-of-tune to the most in-tune. The melodic sequences considered to be in-tune received higher scores, lower scores were given to melodic sequences considered as out-of-tune. Ratings of the three experiments are available at https://edmond.mpdl.mpg.de/imeji/collection/tKj5Oo5FLMglWzWu?q=.

Statistical analyses. In each experiment, preference scores were analyzed using mixed-model analyses of variance (ANOVA), with the primary focus being the main effects of factors related to the manipulation of scoop (including presence, location, and direction) and interactions with other factors. For example, main effects related to stimulus melody are not of interest, but interations of this variable with the direction of starting scoops are. Moreover, we follow each omnibus ANOVA with a series of complex planned comparisons based on whether scoops preserve or disrupt continuity between adjacent melody pitches, as well as whether the direction of the scoop compensates for mistuning of the asymptote. Complex contrasts were based on the sum of cross-products between mean preference scores, and contrast coefficients represent the degree to which a particular participant responds to continuity or compensation. Interactions were analyzed based on how these contrasts interacted with other experimental factors (see Keppel & Wickens, 2004, for further discussion). Finally, we assessed whether predictors based on these contrasts (i.e., continuity and compensation) account for unique portions of variance in preference scores with multiple regression analyses. All statistical decisions were made with $\alpha = .05$, applying type-I error correction as necessary. Unless stated otherwise, all results in which $df_{\text{effect}} > 1$ that are reported as significant were significant after applying the Greenhouse-Geisser correction.

Experiment 1

As outlined in Table 2, Experiment 1 investigated pitch deviations at both timescales (asymptotic deviations, as well as scoops), but limited scoops to be present only at the beginning or ending of tones (i.e., not at both points). The manipulations were presented in two different melodic contexts (Melodies 1 and 2 in Figure 3), to investigate the effect of alterations at a relatively high (Melody 1) or a relatively low (Melody 2) point in the overall melodic contour.

Method

Participants. Fifty-two University at Buffalo students (24 females), from 18 to 22 years old (M = 19.40, SD = 1.24), participated in exchange for course credit.[1]

Participants reported normal hearing abilities and few of them reported a limited amount of formal music training (up to 8 years, M = 0.94 years, SD = 1.85). Each participant was randomly assigned to one of two melody conditions and responded based on these stimuli for the entire session.

Design. For each participant, we manipulated the asymptotic level of the third tone in each melody and the presence of fine-timescale (i.e., scoops), at the beginning or at the end of this tone, across trials. The asymptotic tuning of tones was varied across three levels: in-tune, 50-cents flat, and 50-cents sharp, as described in the General Method section. As shown in Table 2, scoops in Experiment 1 were restricted to be only at the beginning or at the end, but were never present at both locations. In all, there were five levels of the within-subjects factor scoop. This came from the factorial combination of scoop location (beginning or end), and scoop direction (downward or upward), plus a single control condition with no scoops. In addition to these within-subjects variables (asymptotic tuning and scoop), we manipulated melody between subjects, with half the participants listening to variations of Melody 1 across trials, and the other half listening to Melody 2. As shown in Figure 3, tone 3 is a local peak in the contour of Melody 1 but a local valley in Melody 2. Thus, each participant was exposed to 15 variants of the melody, with all possible pairings yielding 105 pairs for a single melody.

[1] The justification for this sample size stems from a conservative assessment of statistical power. Although no research existed on the perceptual effects of vocal scoops before this study, manipulations of intonation in vocal patterns typically yield very strong effects. For instance, we estimated the omega-squared effect size (according to Olejnik & Algina, 2003) for a manipulation of vocal intonation in Hutchins and colleagues (2012) to be .79, whereas a "large" effect size based on Field (2013) would be .14. A significant main effect of scoop, our primary factor of interest, based on a "large" effect size would yield a significant effect with a power of .8 with a sample size of 9 participants. We chose a much larger sample than this, given the assumption that interactions with the factor scoop may be borne out in smaller effects.

Procedure. After reading and signing the information sheet and consent form, participants evaluated 105 pairs of melodic sequences with the pairwise comparison paradigm (see General Method). Participants were randomly assigned to one of two orders of trials. The entire session lasted about 40 min.

Results

As discussed earlier, we report streamlined analyses that focus on effects associated with the manipulation of scoops, including contrast analyses designed to focus on the effect of continuity.[2] The omnibus ANOVA yielded a significant main effect of scoop, $F(4, 200) = 15.99$, $p < .001$, $\eta_p^2 = .24$, a significant Melody × Scoop interaction, $F(4, 200) = 10.03$, $p < .001$, $\eta_p^2 = .17$, and a significant Asymptote × Scoop interaction, $F(8, 400) = 5.49$, $p < .001$, $\eta_p^2 = .10$.

The main effect of scoop reflected the fact that participants preferred melodies with no scoops (*M* preference score for no scoops = 7.95, *SD* = 1.26) over all conditions with scoops. Lowest preferences were for downward ending scoops (*M* = 5.96, *SD* = 1.48), and other conditions yielded intermediate preferences (upward starting scoops, *M* = 7.01, *SD* = 1.12, downward starting scoops, *M* = 7.08, *SD* = 1.18, upward ending scoops *M* = 7.00, *SD* = 1.08). These observations were verified with post hoc pairwise comparisons that adopted a Bonferroni correction for familywise $\alpha = .05$.

The interaction of scoop with asymptote is plotted in Figure 5 (tables of means across all conditions for each experiment are provided in Appendix B). We addressed the generality of scoop effects by analyzing simple effects of scoop at each asymptote level. The effect of scoop was significant ($p < .05$) and of a similar effect size when the asymptote was in-tune, $\eta_p^2 = .24$, or flat, $\eta_p^2 = .21$. The effect was considerably smaller when the asymptote was sharp, $\eta_p^2 = .05$, and was not significant when adopting the Greenhouse-Geisser correction (*p* without correction = .02, with correction = .05).

We then analyzed the Asymptote × Scoop interaction using a complex planned comparison analysis that addressed whether scoops *compensated* for asymptotic mistuning. As discussed in the introduction, a statistical listening approach would cause listeners to respond positively to downward scoops at the start (or upward scoops at the end) if the asymptote were flat, but not if the asymptote were sharp. For this analysis, we discarded trials in which the asymptote was in tune, and coded scoops as compensatory (+1) if the extreme point of the scoop counteracted the mistuning (e.g., a downward starting or upward ending scoop if the asymptote is flat), coded scoops as anticompensatory (−1) if the scoop exaggerates the asymptotic mistuning (e.g., an upward starting scoop or downward ending scoop for a flat asymptote), and coded scoops as neutral (0) for conditions with no scoop. Figure 5 highlights compensatory scoops with downward arrows.

Across all conditions, the linear contrast coefficient based on compensation was significant with a large effect size, $t(51) = 4.11$, $p < .001$, $r^2 = .25$. As can be seen in Figure 5, there is a tendency for listeners to favor compensatory scoops, although this is only obvious when the asymptotic tuning is flat. In line with this observation, a further analysis showed that the magnitude of the contrast varied significantly with asymptotic tuning, $F(1, 50) = 3.48$, $p = .001$, $\eta_p^2 = .10$, with a significant difference from zero for flat asymptotes, $t(51) = 5.09$, $p < .001$, but not for sharp asymptotes ($p = .393$).

We next addressed the role of musical context by probing deeper into the interaction of scoop with melody. We were specifically interested in whether the effect of scoops on intonation ratings was driven by *continuity* across successive pitches, as a teleological approach would suggest. Figure 6 shows this interaction, highlighting those scoop conditions that enhance continuity from tone to tone with downward arrows. As can be seen, there is a tendency for listeners to prefer anticontinuity here. Thus, we explored the effect of continuity more deeply using the contrast analysis described in the General Method section. Specifically, we coded each condition that enhances continuity (a downward starting scoop or upward ending scoop for Melody 1; an upward starting scoop or downward ending scoop for Melody 2) as +1, all anticontinuous conditions (the other scoop conditions) as −1, and neutral conditions (no scoop) as 0.

First, we evaluated the role of continuity across all conditions after averaging scores within each continuity category. The magnitude of the contrast was statistically significant, $t(51) = -5.89$, $p < .001$, $r^2 = .40$. Preferences were highest for conditions with no scoop (*M* = 7.95, *SD* = 1.26) or anticontinuity (*M* = 7.30, *SD* = 0.70), and lowest for scoops that preserved continuity (*M* = 6.23, *SD* = 0.75). We went on to analyze whether contrasts associated with continuity further varied with melody, by defining the contrast separately for each participant and melody. A one-way between-subjects ANOVA revealed that the contrast did vary significantly across conditions, $F(1, 50) = 11.59$, $p < .01$, $\eta_p^2 = .19$. The effect of continuity was stronger for Melody 2 than for Melody 1, as is apparent in the highlighted conditions in Figure 6. Importantly, however, one-sample *t* tests within each melody condition showed that the contrast effect was significant for each melody ($p < .01$ in each case).

We used multiple regression to assess whether both predictors—continuity and compensation—account for unique portions of variance in intonation judgments. We removed all conditions with an in-tune asymptote for this analysis, and averaged scores across all trials for a subject that represented a unique contribution of both predictors. The multiple regression was significant, $F(2, 257) = 7.71$, $p < .001$, $r^2 = .06$, with each predictor accounting for a significant unique portion of the variance. A follow-up comparison between the magnitude of each simple correlation suggested that continuity may be a stronger predictor than compensation (*r* for continuity = .21, *r* for compensation = .12).

Finally, we addressed an important straightforward question: To what extent do listeners simply respond to the average tuning across the entire tone? We addressed this issue by regressing preference ratings on the absolute difference between the mean F0 of evaluated tones and ideal tuning. This regression, which operationalizes the prediction from a statistical learning perspective, was statistically significant, $r(28) = -.78$, $p < .001$. Thus, listeners are sensitive to the overall average, and thus prefer scoops whose direction counteracts asymptotic mistuning over scoops that exaggerate asymptotic mistuning. At the same time, when we entered this measure as an additional predictor into the aforementioned

[2] As highly expected, the main effect of asymptote on listeners' ratings was significant: $F(2, 100) = 104.02$, $p < .001$, $\eta_p^2 = .68$.

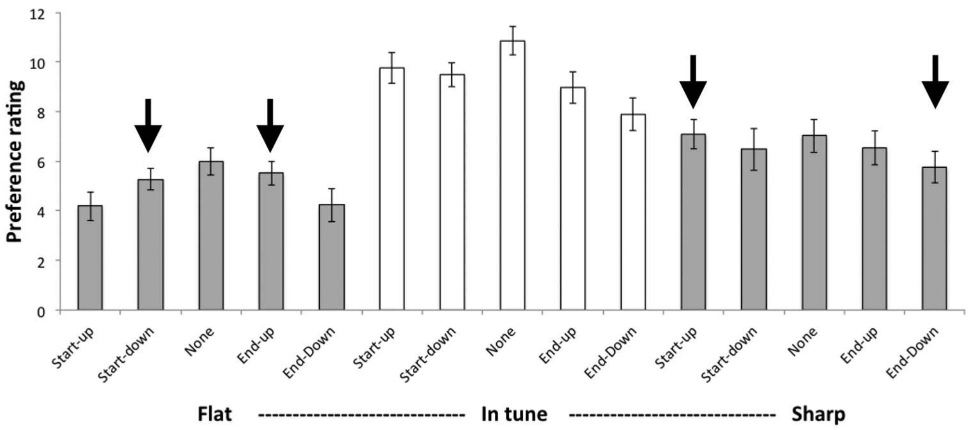

Figure 5. Mean preference rating by scoop and asymptotic tuning conditions in Experiment 1. Error bars display 95% confidence intervals. Downwards arrows indicate scoops that compensate for asymptotic mistuning.

regression model, it did not account for a significant unique portion of the total variance. In other words, although the magnitude of scoops was linked to ratings, the continuity and compensatory tuning effects within each scoop accounted for participant responses beyond this influence.

Discussion

The results of Experiment 1 make several important contributions. First, in a basic but important sense, this experiment confirmed that temporary instabilities in F0 surrounding the center of a tone contribute significantly to judgments of intonation. Listeners do not simply disregard these scoops, even though the asymptotic tuning is an important criterion in pitch accuracy evaluation (Larrouy-Maestri et al., 2013, 2015). Thus, the common practice of omitting these instabilities from tones during the analysis of singing accuracy may not properly reflect all features of the acoustic signal that relate to listeners' perception. More generally, this finding supports that smaller time windows (i.e., smaller than the usual musical tone) are actually processed (Gordon & Poeppel, 2002; Luo et al., 2007; Teng et al., 2016) and dynamic changes at the start or end of tones are relevant to listeners when listening to singing performances.

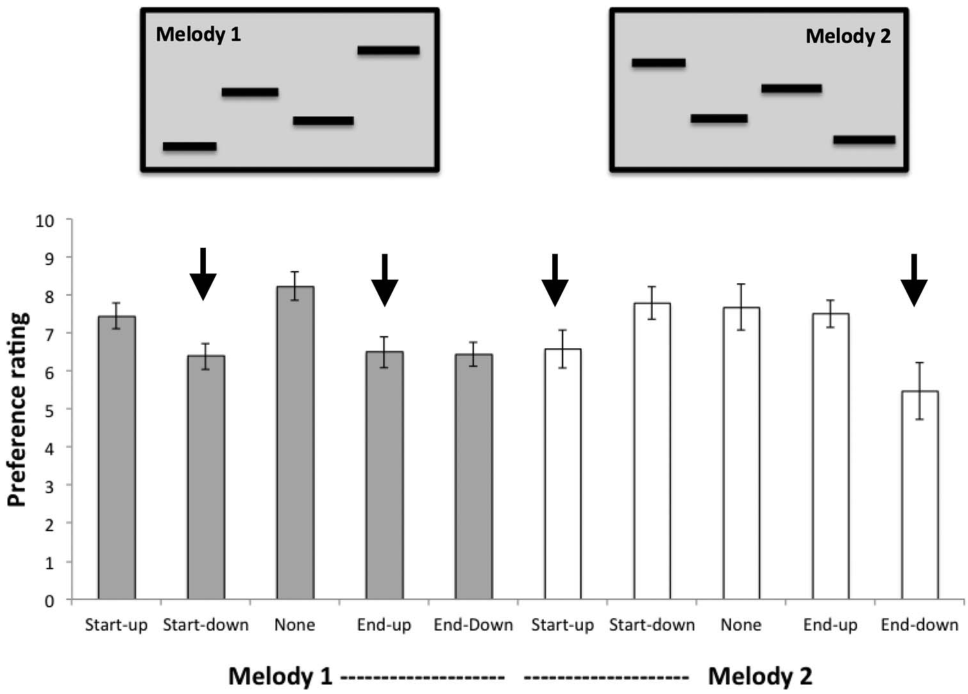

Figure 6. Mean preference ratings by scoop condition and melody in Experiment 1. Error bars represent 95% confidence intervals. Downwards arrows highlight conditions in which the scoop enhances the continuity of pitch change between successive tones. Melodies are displayed in piano-roll format above corresponding data.

Second, listeners seem to use a combination of both teleological (i.e., goal-oriented) and statistical listening strategies, with results not fully conforming to predictions of either hypothesis. Consider the interaction of asymptotic tuning with scoop, shown in Figure 5. Here there was some evidence in favor of a statistical listening strategy, but it was limited to conditions in which asymptotic tuning was flat. Another result in favor of a statistical listening strategy is the strong relation observed between proximity of mean F0 to ideal tuning and listeners' ratings. However, this measure did not reach significance level when entered in our model predicting listeners' variance in rating pitch accuracy. In addition, compensatory scoops were not powerful enough to reverse the effect of an asymptotic mistuning; in every case the presence of a scoop led to equivalent or lower preference ratings than a mistuned asymptote. Although scoops do contribute to pitch perception, listeners may still place more weight on static than dynamic portions of tones in forming statistical estimates of tone properties (cf., Gockel et al., 2001).

A stronger predictive factor in the effect of scoops seems to be the relationship between scoop direction and the surrounding melodic context, supporting a goal-oriented strategy. The direction of this relationship runs against an intuition that may arise from classic Gestalt theories of perception. According to the grouping principle, *Good Continuation* observers prefer smooth contours over those that have abrupt trajectory changes when perceiving a Gestalt. Melodies are commonly considered to be perceived as Gestalts (i.e., whole); thus, one might assume that listeners prefer scoops that enhance continuity across tones. Also, a recent study has pointed out that singing is associated with the impression of glides between tones (Merrill & Larrouy-Maestri, 2017) and thus a kind of continuity between tones. Yet, the results from Experiment 1 indicate a preference against continuity. Listeners preferred tones with scoops that enhanced contrast across tones. However, the replication of this effect, and its observation in other melodic contexts (see Experiment 3), would be necessary to generalize this counterintuitive finding.

Interestingly, the lowest ratings were attributed to scoops at the end of the tone. We suspect that this disfavor comes from an internalized sense for what accurate and inaccurate singers do. On a physiological level, scoops could represent the fine adjustment and tension of vocal folds (Sundberg, 2013; Titze, 1989, 2000). Whereas a motor adjustment (i.e., starting scoop) seems common, because even trained singers show pitch fluctuations at the start (Mori et al., 2004; Saitou et al., 2005), the lack of stability and decrease in tension of the vocal folds at the end of the tone could correspond to poor abilities in sustaining subglottic pressure (i.e., lack of breath support). The rating of the listeners might be supported by their implicit evaluation of the vocal instrument of the performer, with a preference for "normal" perturbation. Recall that the manipulations were designed to be representative of human singing (see Appendix A), so that our results would not be confounded by having some scoops that sounded "unnatural."

An important limitation of Experiment 1 is that scoops were only present at the beginning or the ending of tones, but never at both positions. In practice, scoops at the start and end are usually combined. Experiment 2 was designed to address interactions across scoops at the beginning and end of tones.

Experiment 2

In Experiment 2, we investigated the effects of factorial combination of starting and ending scoops, where a tone may have scoops at one or both locations. As in Experiment 1, we assessed the influence of scoops with respect to whether they preserve continuity of pitch transitions. In addition, Experiment 2 was designed to analyze whether increasing the number of scoops yields additive or interactive effects on perception by including scoops at one or both positions. To keep the number of trials within a reasonable limit, Experiment 2 included only manipulations of scoops and no manipulation of asymptotic tuning; thus we do not specifically address compensatory effects of scoops (i.e., compensation for a mistuned asymptote) in Experiment 2.

Method

Participants. The same participants from Experiment 1 took part in Experiment 2, following a short break.

Materials. As illustrated in Table 2, all conditions in Experiment 2 involved "ideal" asymptotic tuning. This constraint was implemented to make the total number of comparisons per sequence manageable for participants. Instead, Experiment 2 focused on manipulations of scoops that could be present at the start and/or end of the third tone of the melody, using the values for scoop magnitudes shown in Table 1.

Procedure. Each stimulus melody was crossed with 9 conditions involving a factorial combination of starting scoops (none, upward, downward) and ending scoops (none, upward, downward). As in Experiment 1, participants experienced all possible parings (36 in total). Experiment 2 lasted about 20 min.

Data analysis. The rankings were computed following the procedure described in the General Methods section. As in Experiment 1, we analyzed the way in which scoops preserved or counteracted continuity between tones. In addition, because the number of scoops per tone varied in Experiment 2, we also analyzed the contribution of number of scoops in a complex planned comparison. Because the asymptote was never mistuned in Experiment 2, we did not analyze compensatory effects of scoops.

Results

As in Experiment 1, we report an omnibus ANOVA that includes all relevant factors, followed by planned contrasts. In addition to contrasts based on the presence of continuity, used as in Experiment 1, we also ran a contrast analysis that evaluates how much the total number of scoops—irrespective of their direction or effect on overall tuning—contributes to intonation judgments. This second analytical comparison was based on the aforementioned hypothesis that scoops may function as independent perceptual units. Means for all conditions in Experiment 2 are shown in Figure 7; a table of all means and standard deviations is presented in Appendix B.[3]

[3] The astute reader will notice that five of the nine conditions in Experiment 2 are also shared with Experiment 1. We used this overlap as a test–retest reliability test. The pattern of means across these conditions and both melodies ($n = 10$) was significantly correlated across experiments, $r(8) = .88$, $p < .01$, offering support for reliability of our participants.

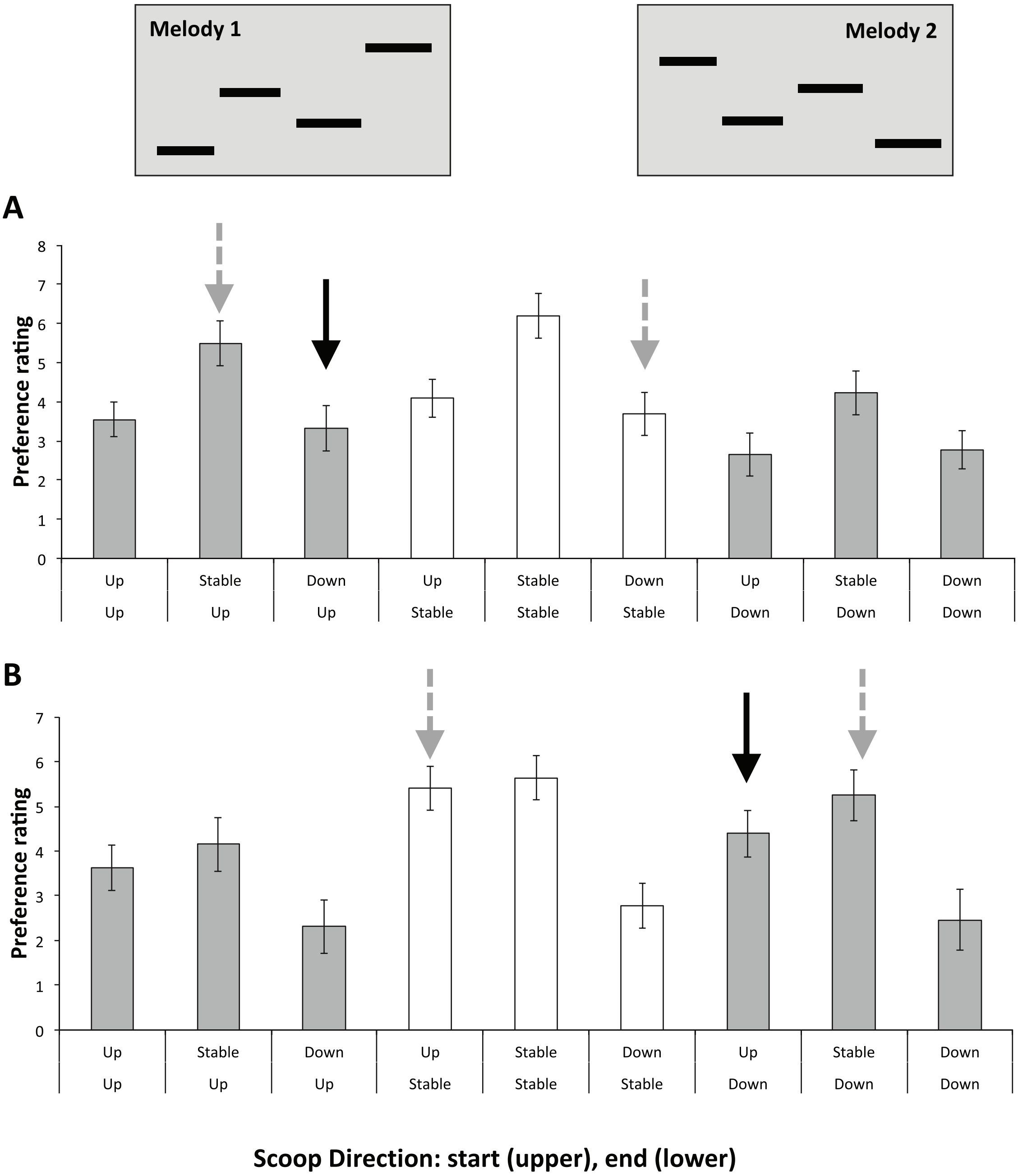

Figure 7. Mean preference rating by scoop condition for Melody 1 (A) and Melody 2 (B) in Experiment 2. Schematic illustrations of each melody appear at the top. *x* axis labels are arranged so that the upper label refers to the starting scoop and the lower label refers to the ending scoop. Error bars represent 95% confidence intervals. Dark downward arrows highlight conditions in which both scoops preserve continuity; gray dashed arrows highlight conditions with only one scoop preserving continuity.

The omnibus ANOVA included the within-subjects factors starting scoop (none, upward, downward) and ending scoop (same levels), and the between-subjects factor melody. There were significant main effects of starting scoop, $F(2, 102) = 18.47$, $p < .001$, $\eta_p^2 = .27$, and ending scoop, $F(2, 102) = 60.05$, $p < .001$, $\eta_p^2 = .54$, and significant two-way interactions of melody with starting scoop, $F(2, 102) = 9.45$, $p < .001$, $\eta_p^2 = .16$, and ending scoop, $F(2, 102) = 10.19$, $p < .001$, $\eta_p^2 = .17$. The three-way interaction among all factors was only significant when the Greenhouse-Geisser correction was not applied ($p = .047$ without correction, $p = .096$ with the correction), and the Starting × Ending scoop interaction was not significant ($p = .084$). Taken together, these results suggest that scoops at the beginning and ending of tones have an additive effect on intonation judgments, with possibly stronger effects for ending scoops than starting scoops given differences in effect size and interactions with melody type.

Similar to Experiment 1, we analyzed the influence of continuity on listener ratings. Because we varied starting and ending scoops factorially in Experiment 2, we constructed separate contrast coefficients for starting and ending scoops. First, we addressed the main effect of starting scoop via contrasts. The magnitude of this contrast was statistically significant, $t(52) = -3.73$, $p < .001$, $r^2 = .21$, with highest preferences for conditions with no scoops ($M = 4.63$, $SD = 0.69$) or anticontinuity ($M = 4.08$, $SD = 0.77$), and lowest for scoops that preserved continuity ($M = 3.29$, $SD = 0.87$). We went on to analyze whether contrasts varied by melody, as in Experiment 1, but the ANOVA yielded a nonsignificant effect ($p = .58$, $\eta_p^2 = .01$). Moreover, the magnitude of the contrast was significant within each melody ($p < .05$ for each).

When focusing on ending scoops, the overall contrast effect was significant, $t(52) = -3.62$, $p < .001$, $r^2 = .20$. However, for ending scoops, there was a large difference in judgments between trials with no ending scoops ($M = 5.15$, $SD = 0.87$) and those with anticontinuous scoops ($M = 3.88$, $SD = 0.97$), whereas these conditions were nearly equal for starting scoops. As in other cases, lowest preference scores were assigned to ending scoops that preserved continuity ($M = 2.97$, $SD = 1.07$). Further exploration suggested that the effect of continuity for ending scoops interacted significantly with melody, $F(1, 51) = 26.00$, $p < .001$, $\eta_p^2 = .34$, with a significant effect of contrast within Melody 2 ($p < .001$) but not Melody 1 ($p = .76$). Thus, Experiment 2 offered some evidence that listeners preferred anticontinuous scoops, although the results were more variable than in Experiment 1.

An important difference between Experiments 1 and 2 was that more than one scoop could be present for manipulated tones in Experiment 2. To investigate how the number of scoops influenced responding, we undertook an exploratory analysis based simply on the frequency and location of scoops independent of their direction. We constructed a mixed-model ANOVA with the "number of scoops" (none, start only, end only, both) as the within-subjects variable and melody as a between-subjects variable. As illustrated in Figure 8, there was a large significant main effect for number of scoops, $F(3, 49) = 39.694$, $p < .001$, $\eta_p^2 = .71$, but no interaction with melody ($p = .43$, $\eta_p^2 = .05$). Preference scores were highest for conditions with no scoops ($M = 5.91$, $SD = 1.35$), followed by one scoop at the beginning ($M = 4.78$, $SD = 0.78$), one at the end ($M = 4.00$, $SD = 0.79$), and two scoops ($M = 3.14$, $SD = 0.59$). Post hoc pairwise comparisons between all adjacent pairs of means, using a Bonferroni correction, verified that all adjacent means differed from each other.

As in Experiment 1, we used multiple regression to test the relative contribution of both predictors used in planned comparison analyses: Continuity of scoops and the total number of scoops. In the first regression model, we combined two predictors based on starting and ending scoops, respectively, along with a single predictor for number of scoops (0, 1, or 2). All three predictors yielded significant bivariate correlations with participant ratings ($p < .001$ in each case). The regression equation accounted for a significant proportion of total variance in ratings, $F(3, 473) = 76.77$, $p < .001$, $R^2 = .32$. More important, each predictor accounted for a significant unique portion of total variance. Also, as in Experiment 1, we tested whether these variables accounted for ratings when overall deviation of the tone from ideal tuning was included as an additional predictor, $F(4, 472) = 59.38$, $p < .001$, $R^2 = .33$, and when predictors coding the magnitude of starting and ending and scoops (2 additional predictors) was included, $F(5, 471) = 62.65$, $p < .001$, $R^2 = .40$. In each case, coefficients that coded the unique properties of scoops (continuity, or number of scoops) contributed independently to the regression model.

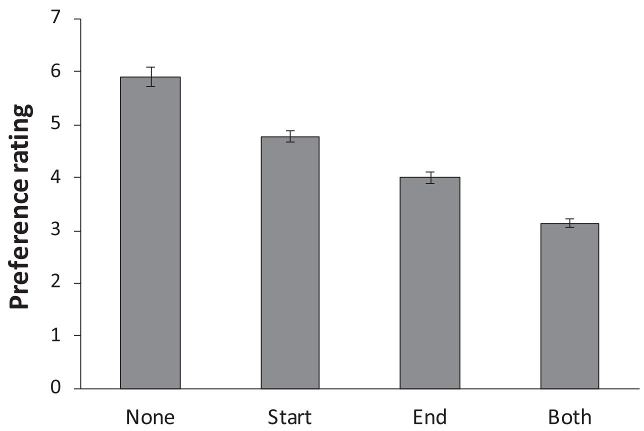

Figure 8. Mean preference rating by scoop number/position in Experiment 2 (i.e., no scoop, a scoop at the start, a scoop at the end, and scoops at both start and end). Error bars represent 95% confidence intervals.

Discussion

Experiment 2 demonstrated the effect of dynamic changes (i.e., scoops) in the absence of any deviations in asymptotic tuning. Scoops might be considered as nuances of melodic performances that are not processed consciously by listeners (Raffman, 1993), but participants do not simply disregard these brief transitions in F0 when evaluating melodic intonation.

Although we did not thoroughly analyze compensatory effects of scoops, as Experiment 2 did not include conditions with asymptotic mistuning, it can be seen that the participants' responses to scoop did not simply reflect statistical averaging of tuning across a tone. For instance, in responding to altered versions of Melody 1, there was no difference in preference across scoops that proceed in the same direction (e.g., both upward) versus those that head in different directions. In fact, the strongest predictor of how scoops influenced listener judgments may simply be based on the number of scoops that are present in a tone. Regardless of the direction of a scoop, listeners preferred tones that had only one scoop rather than two, with greatest preference for tones with no scoops. In line with Experiment 1, Experiment 2 revealed a difference between scoops at the start versus the end of a tone, with stronger effects for the latter than the former. In other words, listeners may be more 'forgiving' about scoops at the beginning versus the end of the tone. Also, they are particularly sensitive to combination of scoops, which is the most usual case in occasional singers' performances.

Finally, by examining listeners' ratings of pitch accuracy in various "natural" contexts (i.e., factorial combination of scoops), this experiment confirms the first results (i.e., preference of anticontinuity), and supports the idea that listeners' processing of scoops follows a teleological strategy, beyond any sensitivity to the tone's statistical properties. Our choice to only present scooped tones (without asymptote deviations) allowed to further examine listeners' sensitivity to scoops (i.e., dynamic changes in small time windows, depending on their position and number) but also lead to experimental limitations. The "ideal" asymptote is exceptional in singing performances (Dalla Bella, Giguère, & Peretz, 2007; Hutchins, Larrouy-Maestri, & Peretz, 2014; Pfordresher & Brown, 2007) and might not perfectly reflect ecological performances. The lack of asymptotic mistunings might have artificially enhanced the apparent importance of scoops. In addition, both Experiments 1 and 2 revealed some interactions between the melody and the factors manipulated. Such influence of the melody on these processes (even if limited) might be attributable to the particular melodic contour of the melodies (up-down or down-up) under study. Thus, the generalization of our findings to music perception requires the replication of the reported effects in representative performances (i.e., combination of asymptotic mistuning and scoops) and contrasted melodic material (i.e., different melodic contours).

Experiment 3

To better understand the processes behind scoop perception and to overcome the limitations previously discussed, Experiment 3 combined elements of the previous two experiments. Like Experiment 1, it included manipulations of asymptotic tuning as well as scoops. Like Experiment 2, scoops were present at both the beginning and ending of the third tone. To keep the number of trials within a reasonable limit, all trials in Experiment 3 included scoops at both the beginning and the end of the third tone.[4] To more fully address the role of melodic context in the effect of scoops, we added two additional melodies to the stimulus set of Experiment 3: Melodies 3 and 4 from Figure 3. Taken together, these four melodies include every combination of directional change from tone 2 to tone 3 and from tone 3 to tone 4. In addition to replicate and generalize previous finding, this experiment was designed to further clarify (and compare) the relevance of both teleological and statistical approach in music perception.

Method

Participants. Fifty-four University at Buffalo students (17 females), 18 to 23 years old ($M = 19.28$, $SD = 1.25$), participated and received University credit. Two participants reported considerable musical training (nine years of saxophone lessons and 10 years of piano lessons with actual training of four and three hours per week, respectively). Of the remaining participants, a few of them reported a limited amount of formal music training (up to seven years, $M = 1.04$ years, $SD = 1.89$). One of the participants reported having absolute pitch. Each participant was randomly assigned to two of the four melodies and responded based on these stimuli for the entire session.

Materials. To exhaustively explore the role of melodic context, all four melodies shown in Figure 3 were used in Experiment 3. Thus, we can address how scoops influence intonation judgments when positioned at a local minimum (Melody 1) or maximum (Melody 2) in the contour, as well as situations in which the tone is framed within an ascending trajectory (Melody 3) or descending trajectory (Melody 4).

There were four scoop conditions resulting from the factorial combination of starting scoops (upward, downward) and ending scoops (upward and downward). When these four conditions were further crossed with two asymptotic tunings (+50 cents, or −50 cents deviation), there were 12 experimental conditions for each of the four melodies, and thus 66 pairwise comparisons per melody.

Procedure. The melodic sequences were presented with the pairwise comparison paradigm described in the general methods. Each block contained one melody (i.e., 12 versions of each). Because of time constraints, one subset of the participants ($n = 25$) listened to and compared Melodies 1 and 2, and the other participants ($n = 29$) listened to and compared Melodies 3 and 4. Within blocks, the presentation of pairs ($n = 66$ pairs per block) was randomized. This experiment lasted about 50 min.

Design. Three within-subjects factors influenced the tuning of the third tone in a melody: starting scoop (upward, downward), ending scoop (upward, downward), and asymptote (flat or sharp) as reported in Table 2. The four melodies were arranged according to two factors to limit the number of trials for a given participant. Half of the participants heard Melodies 1 and 2, whereas the other half heard Melodies 3 and 4. This constituted the between-subjects factor "melody set." Within each melody set, the two individual melodies constituted the within-subjects factor "melody instance." It is important to note that the two-factor arrangement of melodies was done only for convenience of participants; there is no compelling theoretical distinction between these factors. Thus, when we discuss ANOVA analyses we only focus on the interaction of these factors and disregard any main effects of either factor related to melody.

Results

The omnibus ANOVA for Experiment 3 revealed a significant main effect of ending scoop, $F(1, 52) = 43.08$, $p < .001$, $\eta_p^2 = .45$, supporting a preference for upward scoop, but no main effect of starting scoop ($p = .193$, $\eta_p^2 = .03$).[5] There were also significant two-way interactions between asymptote and starting scoop, $F(1, 52) = 28.54$, $p < .001$, $\eta_p^2 = .35$, asymptote and ending scoop, $F(1, 52) = 13.70$, $p = .001$, $\eta_p^2 = .21$, as well as an interaction between starting and ending scoop, $F(1, 52) = 21.21$, $p < .001$, $\eta_p^2 = .29$. The two factors that comprised the four different melodies only interacted with ending scoop, $F(1, 52) = 4.97$, $p = .03$, $\eta_p^2 = .09$.

In comparison with the other effects, the interaction with melody type yielded a small effect, suggesting that most of the treatment variance was carried by the effect of scoops and asymptotic tuning. Mean ratings broken down by these factors are plotted in Figure 9 (full Table is presented in Appendix B). As in previous figures, conditions in which both scoops compensated for asymptotic tuning are highlighted with downward arrows. These results are similar to Experiment 1, in that there was a strong preference for compensatory scoops when the asymptote was flat, whereas the

[4] In the interest of brevity, we have omitted four conditions from Experiment 3 that were also included in Experiment 2, which featured in-tune asymptotes. The means from these conditions were positively correlated across experiments when broken down across the two melodies that were used in both experiments, $r(6) = .81$, $p < .01$.

[5] As highly expected and confirmed in Experiment 1, the main effect of asymptote on listeners' ratings was significant: $F(1, 52) = 44.71$, $p < .001$, $\eta_p^2 = .46$.

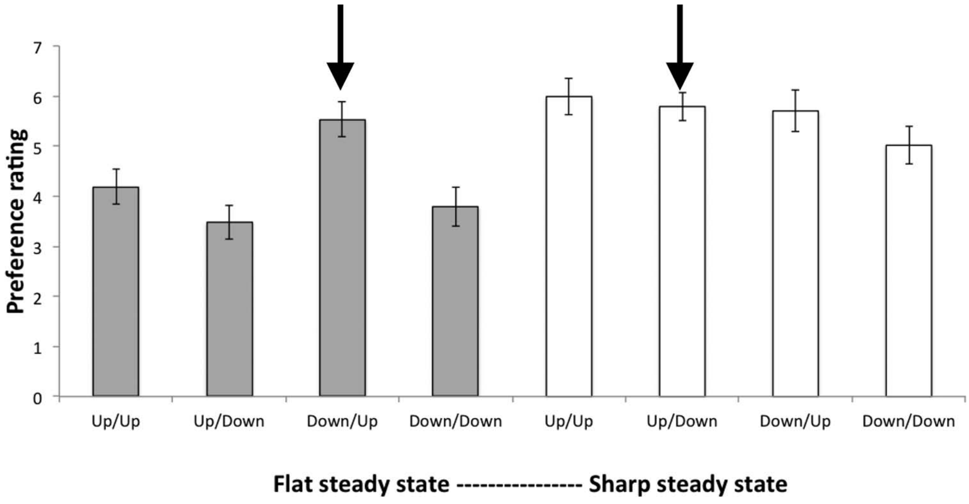

Figure 9. Mean preference rating by scoop and asymptotic tuning conditions in Experiment 3. *x* axis labels are arranged so that the first label refers to the starting scoop and the second label refers to the ending scoop. Error bars display 95% confidence intervals. Downwards arrows indicate scoops that compensate for asymptotic mistuning.

effect of compensatory scoops was greatly reduced when the asymptote was sharp.

The role of compensation in the effect of scoops was further analyzed with planned contrasts as carried out in Experiment 1, but with coefficients being computed separately for starting and ending scoops. Across all conditions the effect of compensation within starting scoops was significant with an even larger effect size than found in Experiment 1 across all scoops, $t(53) = 5.30$, $p < .001$, $r^2 = .35$. Unlike Experiment 1, the magnitude of the contrast did not vary significantly with asymptotic tuning ($p = .19$, $\eta_p^2 = .02$) and was significant for both sharp, $t(53) = 3.10$, $p = .002$, and flat, $t(53) = 4.06$, $p < .001$, asymptotes. Likewise, this contrast did not vary as a function of the four melodies. Thus, compensation within starting scoops had a strong role in how participants responded to scoops.

The compensatory status of ending scoops yielded a significant effect that was slightly weaker than was found for starting scoops, $t(53) = 3.73$, $p < .001$, $r^2 = .21$. This effect did not vary significantly as a function of melody but further analysis showed that the contrast was only significant for Melodies 1 and 4 and not for 2 and 3. The magnitude of this contrast did, however, vary significantly as a function of asymptotic tuning, $F(1, 53) = 41.73$, $p < .001$, $\eta_p^2 = .44$. Compensation for ending scoops had a stronger effect when the asymptote was flat than when it was sharp, yet the effect was significant in both cases.

Two further analyses based on compensatory effects combined the coefficients for starting and ending scoops. One model was based on an interactive (i.e., multiplicative) combination, whereas the other was additive. Both models led to significant contrasts overall [interactive model, $t(53) = 4.65$, $p < .001$, $r^2 = .29$, additive model, $t(53) = 6.47$, $p < .001$, $r^2 = .44$]. Because the additive model accounts for considerably more variance, we conclude that perception of intonation is largely driven by an additive weighting of starting and ending scoops.

We next focused on the effect of scoop continuity, which is of particular interest given that the pattern of starting and ending scoops yielding continuity varies across the four melodies. Moreover, the fact that melody type had little influence in the omnibus ANOVA may reflect the absence of focus on variance specially associated with continuity. We coded continuity in the same way as in Experiment 2, separating continuity effects for starting and ending scoops.

The overall effect of continuity for starting scoops was not significant ($p = .11$, $r^2 = .03$), in contrast to Experiment 2, nor did it interact as a function of the four melodies. However, a closer look within each stimulus melody showed that the effect of continuity was significant for all melodies except for Melody 2, which was also used in Experiments 1 and 2 and did yield a significant effect there. Of course, continuity is fundamentally more complex than in the other experiments given the role of scoops at the beginning and ending of tones that can also vary in asymptotic tuning. Ultimately, the effect of continuity for starting scoops was weaker in Experiment 3 than in Experiment 2.

By contrast, the continuity of ending scoops yielded a strong and significant overall effect, $t(53) = -4.46$, $p < .001$, $r^2 = .27$, which, like Experiments 1 and 2, suggested a preference for anticontinuity. This contrast did not interact significantly as a function of melody or asymptote.

We also explored contrasts based on additive and interactive combinations of starting and ending scoops, the latter of which was based on multiplying coefficients. The interactive contrast model did not predict listener ratings ($p = .33$, $r^2 < .01$). However, the additive contrast did yield a strong significant effect, $t(53) = -4.39$, $p < .001$, $r^2 = .27$, which did not vary as a function of melody or asymptotic tuning.

Taken together, ratings in Experiment 3 were particularly influenced by the continuity of ending scoops. At the same time, the role of continuity in Experiment 3 was not as reliable as in Experiments 1 and 2, possibly due to the aforementioned greater complexity of pitch manipulations in Experiment 3.

As in Experiment 1, we used multiple regression to assess the independent contributions of continuity and compensation to intonation ratings. Because starting and ending scoops appeared to contribute additively in ANOVA analyses, we entered four separate predictors based on continuity and compensation coding for starting and ending scoops. The multiple regression equation predicted a small but significant proportion of variance in ratings, $F(4, 859) = 12.79$, $p < .001$, $R^2 = .06$. All predictors accounted for a significant independent portion of the variance except for the continuity of starting scoops, in line with the contrast analysis reported above. We then added mean deviation (up to about 88 cents) from perfect tuning across the entire tone as a predictor. This predictor, which accounted for a significant portion of variance (independent of others), caused both predictors for starting scoops to become nonsignificant, supporting that ending scoops have a strong impact on ratings independent of overall mistuning.

Finally, to test whether the predictors accounted for ratings when scoop magnitude was included, we ran a model with two predictors that coded the magnitude of starting and ending scoops (6 parameters in total). The regression model was significant, $F(5, 858) = 16.2$, $p < .001$, $R^2 = .09$. All predictors accounted for a significant unique portion of the variance with the exception of two parameters related to starting scoops: continuity, and deviation magnitude.

Discussion

In several critical respects, Experiment 3 verified results from the other experiments. The tendency to prefer scoops compensating for asymptotic mistuning was found (as in Experiment 1) and went beyond effects related to the magnitude of scoops. Likewise, listeners preferred scoops that counteracted the continuity of pitch changes, as found in Experiments 1 and 2, which also went beyond the magnitude of scoops. Finally, there was evidence that ending scoops had a larger effect on listener ratings than starting scoops.

Experiment 3 also extended these results across a wider range of melodies and thus to all combinations of pitch motion between the critical third tone and the two tones adjacent to it. Results were variable across melodies in Experiment 3 more than in other experiments, which may relate to random variability based on the greater complexity of information processed by participants in Experiment 3. Indeed, they had to track variability in asymptotic tuning as well as scoops at both the beginning and end of the third tone. Despite this limitation, Experiment 3 confirmed the co-occurrence of two mechanisms behind the perception and process of scoops: statistical average across the entire tone and relation to the general melody.

General Discussion

By examining participants' ratings of melodies containing pitch manipulations at the start/asymptote/end of tones, this study undoubtedly highlights the relevance of scoops in music perception. As summarized in the upper part of Table 3, the three experiments confirm that scoops matter to listeners when evaluating pitch accuracy of sung performances.

The present findings corroborate other evidence that the auditory system processes dynamic changes in small time windows (i.e., smaller than the usual musical tone, Gordon & Poeppel, 2002; Luo et al., 2007; Teng et al., 2016). Dynamic changes in pitch are not only perceived (Lyzenga et al., 2004; Teng et al., 2016; Wang et al., 2013) but also taken into account when evaluating music performances, particularly when they occur at the end of tones (Experiments 1, 2, and 3) or when they are present at both the start and the end of tones (Experiment 2). Note that the sensitivity to dynamic changes in small time windows (and/or their integration) might vary depending on the question under study. For instance, the relevance of scoops might be lessened or enhanced when the judgment refers to melodic recognition, or aesthetic/beauty judgments rather than to pitch accuracy evaluation (as tested here). Also, if scoops are relevant, whatever the question, the ratings might still differ (e.g., preference for continuity for aesthetic ratings and the opposite for pitch accuracy ratings). Finally, the mechanisms themselves (statistical or teleological strategies) might differ. The specific comparison of different types of judgments, examined with similar methods (and adequate material), would be necessary to address the effect of the question on pitch processing and thus clarify the role of top-down processes involved in music perception.

Table 3

Summary of the Findings: For Each Research Question, the Presence of the Effects (i.e., Position and Direction, Number of Scoops, Compensation, Averaging, and Continuity) Examined in the Three Experiments

Research question	Exp 1	Exp 2	Exp 3
Do scoops matter (as a unit)?			
Position and direction	V	V	V
Number of scoops	—	V	—
Statistical average across the entire tone			
Compensation	V	—	V
Averaging	X	V	—
Relationship between scoop and inferred goal			
Continuity	V	V	V

Note. Check mark (V) represents significant effects, cross mark (X) represents nonsignificant effect, and dashes (—) indicate effects not tested.

In the case of pitch accuracy judgments, the low ratings for ending scoops and particularly for downward ending scoops could be attributed to the implicit evaluation of the poor vocal abilities of the performer, such as a lack of breath support. Starting scoops have been observed in trained singers (Mori et al., 2004; Saitou et al., 2005) and might thus be more accepted. To the best of our knowledge, the ending scoops of trained singers have not been specifically described, which might suggest that they are not as pronounced as the starting ones, owing to adequate breath control (or awareness that ending scoops are not appreciated by listeners). A systematic investigation of the physiological causes of ending scoops and of their presence and magnitude in trained singers would certainly shed light on the reason of listeners' preferences. Nevertheless, this finding supports that the common practice of omitting scoops when analyzing singing accuracy might thus not properly reflect those features of the acoustic signal that relate to listeners' perception. More generally, this study provides experimental ground for pedagogical guidance with regard to scoops in singing development and training (and particularly the stability of subglottal pressure and breath control at the end of tones).

The material of the current study was based on the analysis of singing performances from Pfordresher and Mantell (2014). By averaging scoop magnitude over a large number of tones, the dynamic changes inserted in our material were expected to sound representative of performances from untrained singers (see Appendix A). A tradeoff of this approach was that ending scoops were in general larger in magnitude than starting scoops. Although this difference influenced listeners' judgments to some extent, regression analyses showed that scoop magnitude on its own cannot account for listener judgments.

Besides answering to the main question of this study (i.e., scoops matter when evaluating the correctness of melodies), this study highlights the coexistence of two mechanisms behind pitch perception: statistical and teleological. As reported in Table 3, listeners perform statistical averaging of the tone unit (as shown by the preference for *compensation* in Experiments 1 and 3, and *averaging* in Experiment 2) but also interpret scoops as a unit in a larger context (as shown by the preference for *noncontinuity* in all three experiments). In other words, listener reactions to scoops are based both on the relationship of the scoop to the tuning of the associated tone, as well as the relationship of the scoop to the broader melodic context.

The relevance of scoops (as units) might seem contradictory to Western tonal music theory, which states discrete tones as the smallest units from which melodies are constituted (Krumhansl, 1979; Lerdahl & Jackendoff, 1983; Patel, 2008). However, this study does not argue against tones as units but rather clarifies the use of pitch information in music perception. Indeed, our findings support both the integration of scoops as part of the tone (confirming the relevance of tones as major units) and the preference for discrete tones (containing noncontinuous scoops avoiding glides between tones). In fact, this study suggests that scoops might facilitate the segmentation of tones and thus the implicit learning of musical structure (Bigand & Poulin-Charronnat, 2006). This hypothesis would have to be specifically tested, for instance with psychophysics studies manipulating the magnitude of scoops and the influence on the ability to segment tone streams.

The relation between scoops and melodic context, a novel finding here, also needs further examination. Although we observed preference for noncontinuity in the three experiments, some inconsistences appeared depending on the melodic context and the position of the scoop. For instance, the effect of continuity was stronger for Melody 2 than for Melody 1 in Experiment 1, and the effect of continuity was more important for ending scoops than starting scoops in Experiment 3. Such inconsistencies could be driven by different expectations depending on the melodic context (Pearce & Wiggins, 2006). A systematic examination, by varying the degree of expectations of the tones containing the scoops in the melodic context, would clarify this issue. Nevertheless, the definition of small units, which relate to elements of larger size, supports the existence of multiple time-scale representations in the auditory system (Poeppel, 2001, 2003; Teng et al., 2016). When listening to speech, Ding and collaborators (2015, 2017) highlighted the neural tracking of hierarchical linguistic structures and the integration of units of different sizes. Using similar methods with musical material would allow to examine the existence of such functional mechanism in music processing.

Finally, the identification of two mechanisms co-occurring when listening to melodies exceeds the scope of the music perception and paves the way to a better understanding of auditory sequence processing. Pitch processing is one of the most important functions of the auditory system and is crucial in various activities (beyond music perception), such as comprehension of intentional or emotional prosody (e.g., Banse & Scherer, 1996; Hellbernd & Sammler, 2016). As highlighted by Fritz et al. (2013), similarities and differences across domains are challenging to examine without identifying the respective units and processes. The present data bring us closer to an understanding of what constitutes the basic units of perception in music. Applying the methods proposed here to linguistic material and taking into account the units which are specific to this domain would allow to define the mechanisms involved in prosodic comprehension and, ultimately, to examine cross-domain difference/similarities with regard to auditory processing.

Conclusion

Through three experiments, this study demonstrates that brief dynamic changes to pitch at the beginnings and at endings of sung pitches, which are common in singing performances, influence intonation perception. In addition to highlighting the perceptual relevance of small pitch dynamic changes, as in the language domain, this study revealed the coexistence of two distinct mechanisms underlying scoops processing: average of the pitch information within larger units and use of small unit (i.e., scoop) in relation to the inferred goals of the producer. By using music as a window to examine auditory sequence processing and identifying the combination of statistical and teleological strategies, this study clarifies one of the most important functions of the auditory system.

References

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70,* 614–636. http://dx.doi.org/10.1037/0022-3514.70.3.614

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition, 100,* 100–130. http://dx.doi.org/10.1016/j.cognition.2005.11.007

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer (Version 5.3.23) [Computer software]. Retrieved from http://www.praat.org/

Burns, E. M., & Ward, W. D. (1978). Categorical perception—Phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *The Journal of the Acoustical Society of America, 63,* 456–468. http://dx.doi.org/10.1121/1.381737

Dalla Bella, S. (2015). Defining poor-pitch singing: A problem of measurement and sensitivity. *Music Perception, 32,* 272–282. http://dx.doi.org/10.1525/MP.2015.32.3.272

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *The Journal of the Acoustical Society of America, 121,* 1182–1189. http://dx.doi.org/10.1121/1.2427111

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in Human Neuroscience, 11,* 481. http://dx.doi.org/10.3389/fnhum.2017.00481

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19,* 158–164.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London, England: Sage.

Fritz, J., Poeppel, D., Trainor, L., Schlaug, G., Patel, A. D., Peretz, I., . . . Parsons, L. M. (2013). The reurobiology of language, speech, and music. In M. A. Arbib (Ed.), *Language, music, and the brain*. Cambridge, MA: MIT Press.

Gockel, H., Moore, B. C., & Carlyon, R. P. (2001). Influence of rate of change of frequency on the overall pitch of frequency-modulated tones.

Journal of the Acoustical Society of America, 109, 701–712. http://dx.doi.org/10.1121/1.1342073

Gordon, M., & Poeppel, D. (2002). Inequality in identification of direction of frequency change (up vs. down) for rapid frequency modulated sweeps. *Acoustics Research Letters Online - Acoustical Society of America, 3,* 29–34. http://dx.doi.org/10.1121/1.1429653

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language, 88,* 70–86. http://dx.doi.org/10.1016/j.jml.2016.01.001

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. http://dx.doi.org/10.1111/j.1749-6632.2009.04856.x

Hutchins, S., Larrouy-Maestri, P., & Peretz, I. (2014). Singing ability is rooted in vocal-motor control of pitch. *Attention, Perception, & Psychophysics, 76,* 2522–2530. http://dx.doi.org/10.3758/s13414-014-0732-1

Hutchins, S., Roquet, C., & Peretz, I. (2012). The vocal generosity effect: How bad can your singing be? *Music Perception, 30,* 147–159. http://dx.doi.org/10.1525/mp.2012.30.2.147

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15,* 356–360. http://dx.doi.org/10.1111/j.0956-7976.2004.00683.x

Kacha, A., Grenez, F., & Schoentgen, J. (2005, September). *Voice quality assessment by means of comparative judgments of speech tokens*. Paper presented at the InterSpeech Conference, Lisbon, Portugal.

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

Kerivan, J. E., & Carey, B. J. (1976). Pattern identification of pure tones and frequency glides by untrained listeners. *Perception & Psychophysics, 20,* 489–492. http://dx.doi.org/10.3758/BF03208287

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology, 11,* 346–374. http://dx.doi.org/10.1016/0010-0285(79)90016-1

Large, E. W., Fink, P., & Kelso, J. A. S. (2002). Tracking simple and complex sequences. *Psychological Research, 66,* 3–17.

Larrouy-Maestri, P., Lévêque, Y., Schön, D., Giovanni, A., & Morsomme, D. (2013). The evaluation of singing voice accuracy: A comparison between subjective and objective methods. *Journal of Voice, 27,* e251–259. http://dx.doi.org/10.1016/j.jvoice.2012.11.003

Larrouy-Maestri, P., Magis, D., Grabenhorst, M., & Morsomme, D. (2015). Layman versus professional musician: Who makes the better judge? *PLoS ONE, 10,* e0135394. http://dx.doi.org/10.1371/journal.pone.0135394

Larrouy-Maestri, P., Magis, D., & Morsomme, D. (2014a). Effects of melody and technique on acoustical and musical features of western operatic singing voices. *Journal of Voice, 28,* 332–340. http://dx.doi.org/10.1016/j.jvoice.2013.10.019

Larrouy-Maestri, P., Magis, D., & Morsomme, D. (2014b). The evaluation of vocal pitch accuracy: The case of operatic singing voices. *Music Perception, 32,* 1–10. http://dx.doi.org/10.1525/mp.2014.32.1.1

Larrouy-Maestri, P., Morsomme, D., Magis, D., & Poeppel, D. (2017). Lay listeners can evaluate the pitch accuracy of operatic voices. *Music Perception, 34,* 489–495. http://dx.doi.org/10.1525/mp.2017.34.4.489

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Luo, H., Boemio, A., Gordon, M., & Poeppel, D. (2007). The perception of FM sweeps by Chinese and English listeners. *Hearing Research, 224,* 75–83. http://dx.doi.org/10.1016/j.heares.2006.11.007

Lyzenga, J., Carlyon, R. P., & Moore, B. C. J. (2004). The effects of real and illusory glides on pure-tone frequency discrimination. *The Journal of the Acoustical Society of America, 116,* 491–501. http://dx.doi.org/10.1121/1.1756616

Marmel, F., Tillmann, B., & Dowling, W. J. (2008). Tonal expectations influence pitch perception. *Perception & Psychophysics, 70,* 841–852. http://dx.doi.org/10.3758/PP.70.5.841

Merrill, J., & Larrouy-Maestri, P. (2017). Vocal features of song and speech: Insights from Schoenberg's Pierrot Lunaire. *Frontiers in Psychology, 8,* 1108. http://dx.doi.org/10.3389/fpsyg.2017.01108

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219,* 36–47. http://dx.doi.org/10.1016/j.heares.2006.05.004

Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America, 54,* 610–619. http://dx.doi.org/10.1121/1.1913640

Mori, H., Odagiri, W., Kasuya, H., & Honda, K. (2004, April). *Transitional characteristics of fundamental frequency in singing*. Paper presented at the 18th International Congress on Acoustics, Kyoto, Japan.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8,* 434–447. http://dx.doi.org/10.1037/1082-989X.8.4.434

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

Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception, 23,* 377–405. http://dx.doi.org/10.1525/mp.2006.23.5.377

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

Pfordresher, P. Q., & Larrouy-Maestri, P. (2015). On drawing a line through the spectrogram: How do we understand deficits of vocal pitch imitation? *Frontiers in Human Neuroscience, 9,* 271. http://dx.doi.org/10.3389/fnhum.2015.00271

Pfordresher, P. Q., & Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive Psychology, 70,* 31–57. http://dx.doi.org/10.1016/j.cogpsych.2013.12.005

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science, 25,* 679–693. http://dx.doi.org/10.1207/s15516709cog2505_3

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41,* 245–255. http://dx.doi.org/10.1016/S0167-6393(02)00107-3

Raffman, D. (1993). *Language, music, and mind*. Cambridge, MA: MIT Press.

Saitou, T., Unoki, M., & Akagi, M. (2005). Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis. *Speech Communication, 46,* 405–417. http://dx.doi.org/10.1016/j.specom.2005.01.010

Stalinski, S. M., Schellenberg, E. G., & Trehub, S. E. (2008). Developmental changes in the perception of pitch contour: Distinguishing up from down. *Journal of the Acoustical Society of America, 124,* 1759–1763. http://dx.doi.org/10.1121/1.2956470

Stevens, F. A., & Miles, W. R. (1928). The first vocal vibrations in the attack in singing. *Psychological Monographs, 39,* 200–220. http://dx.doi.org/10.1037/h0093347

Sundberg, J. (2013). Perception of singing. In D. Deutsch (Ed.), *The psychology of music* (pp. 69–105). Cambridge, MA: Academic Press. http://dx.doi.org/10.1016/B978-0-12-381460-9.00003-1

Teng, X., Tian, X., & Poeppel, D. (2016). Testing multi-scale processing in the auditory system. *Scientific Reports, 6,* 34390. http://dx.doi.org/10.1038/srep34390

Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *The Journal of the Acoustical Society of America, 85,* 901–906. http://dx.doi.org/10.1121/1.397562

Titze, I. R. (2000). *Principles of voice production*. Denver, CO: National Center for Voice and Speech.

van Besouw, R. M., & Howard, D. M. (2009). Effects of carrier and phase on the pitch of long-duration vibrato tone. *Musicae Scientiae, 8,* 139–161.

Wang, W. J., Tan, C. T., & Martin, B. A. (2013). Auditory evoked responses to a frequency glide following a static pure tone. *The Journal of the Acoustical Society of America, 133,* 3429. http://dx.doi.org/10.1121/1.4806040

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics, 64,* 198–207. http://dx.doi.org/10.3758/BF03195786

Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139164559

Zarate, J. M., Ritson, C. R., & Poeppel, D. (2012). Pitch-interval discrimination and musical expertise: Is the semitone a perceptual boundary? *Journal of the Acoustical Society of America, 132,* 984–993. http://dx.doi.org/10.1121/1.4733535

Appendix A

Description of Scoops in the General Population

To examine the effect of "natural" scoops on melodic perception, we incorporated pitch manipulations designed to be representative of human singing, that is, typical pitch fluctuations performed by occasional singers. To the best of our knowledge, pitch fluctuations (at the start and end of sung tones, a.k.a. "scoops") in occasional singers have not been quantified so far. To do so, we examined the sung performances from the data of Pfordresher and Mantell (2014).

This dataset includes recordings in which participants imitated 4-tone sequences with different melodic contours. Participants were occasional singers who vocally imitated these tone sequences by singing after listening to prerecorded performances of accurate singers (Pfordresher & Mantell, 2014, Experiment 1). Importantly, the occasional singers varied greatly in singing accuracy. In Pfordresher and Mantell (2014), participants were categorized as accurate if the mean difference between the central pitch of sung tones differed from target pitches by less than 50 cents (a quartertone) on average, or inaccurate if their average deviation was equal to or greater than 50 cents (following Hutchins, Roquet, & Peretz, 2012; see Dalla Bella, 2015, or Pfordresher & Larrouy-Maestri, 2015, for discussion of cutoff criteria).

Our main interest consisted in describing the characteristics of scoops at the start and end of musical sung tones, which were excluded from previous analyses of this data set. Specifically, we sought to measure the amplitude of scoops and rate to reach the stable middle part of the tone (i.e., asymptote). To provide representative values of "natural" scoops that serve as a basis for the pitch manipulation of the main study, we examined the effect of the position of the scoops (i.e., start vs. end), direction (i.e., upward vs. downward), and melodic context (i.e., pitch height relative to the previous or following tone) on the magnitude and rate of scoops.

Materials

Single tones were extracted from the performances of 29 occasional singers (7 females) aged from 18 to 23 years ($M = 19.1$). These participants reflected a wide range of singing accuracy from the general population (see Pfordresher & Mantell, 2014 for details). In total, 2145 tones were available. F0 values were extracted with the autocorrelation algorithm in Praat (Boersma & Weenink, 2018). After visual inspection of the signal, 12.63% of the tones were discarded because of pitch detection errors or because of pitch class errors that disrupted melodic contour. Finally, 1874 tones of about 1 s long were selected for analysis.

Analytical Procedure

To describe the properties of sung tones performed by occasional singers, with a focus on scoops, we used a model of pitch adaptation that was derived from Large, Fink, and Kelso (2002). Although this model was originally designed to model adaptation to a temporal perturbation, we found that it exhibited properties that resembled pitch perturbations within sung tones as well, possibly reflecting a similar adaptive process. Equation A1, from Large and colleagues (2002), was used to plot the adaptation of the phasing of taps in the presence of a temporal perturbation. We here use the same function to predict how a participant's fundamental frequency (F0) adapts to the change in a target pitch at the beginning of a sung tone.

$$Y_{s_t} = A_s \times \exp(-b_s t) \times \cos(2\pi f_s t + \theta_s) \quad \text{(A1)}$$

where A_s refers to the beginning perturbation (amplitude of the preparation, the subscript *s* indicating starting scoop), b_s represents the approach to the asymptote (rate of the preparation), f_s represents the oscillation around the asymptote (which models any tendency to overshoot or undershoot the target), and theta is used to vary the initial direction of the approach (downward or upward scoop).

To examine the pitch fluctuation at the end of the tone (i.e., the parameters A_e, b_e, f_e, and theta), the time values are inverted in a second element (Y_{et} in Equation A2).

$$Pitch_t = Y_{s_t} + Y_{e_t} + asym \quad \text{(A2)}$$

where *asym* is a scalar that relates to the pitch in the center of the tone. Figure A1 (left panel) depicts the three elements of the Equation A2 separately, with a black line for the first element (i.e., scoop at the start of the tone), a gray line for the second element (i.e., scoop at the end of the tone), and a dashed gray line for the asymptote.

(*Appendices continue*)

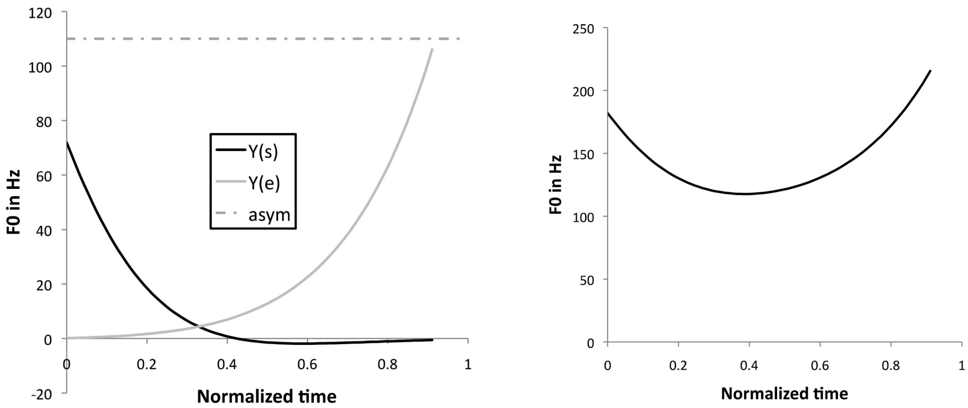

Figure A1. Illustration of the model applied on one tone from a random accurate singer. This particular tone contains upward scoops at the beginning and end. In the left panel, the dark line corresponds to the Equation A1 and the gray line corresponds to the element with inversed time values (Y_e in Equation A2). The right panel corresponds to the sum of the all the elements depicted in Equation A2. For $Y(s)$, parameters were $A = 71.91$, $b = 5.27$, $f = .59$, $\Theta = 0$. For $Y(e)$, parameters were $A = 106.07$, $b = 4.55$, $f = -.26$, $\Theta = 0$. The asymptote value was 110 Hz.

As represented in Equation A2 and illustrated in Figure A1 (right panel), the model predicts how an F0 value at time t can be influenced by perturbations at both the start and the end of the tone, surrounding a stable asymptote at the center of the tone. Note that despite the complexity of the model, several values reflect fixed properties of the data (i.e., all sampled values of each tone): The asymptote corresponds to the median of the middle portion of the tone, A_s (or A_e) is the difference between the beginning (or end) of the tone and the asymptote, and theta is the direction of the overshoot. The fitted parameters concern the rate of the approach (b_s and b_e) and the oscillation around the asymptote (f_s and f_e). For the present analyses, pitch was scaled in cents relative to the lowest starting pitch used in target melodies (C2 for men = 131 Hz, C3 for women = 262 Hz).

Each of the 1874 tones was fitted with a least-squares approximation using the optimization toolbox in Matlab. The goodness of fit

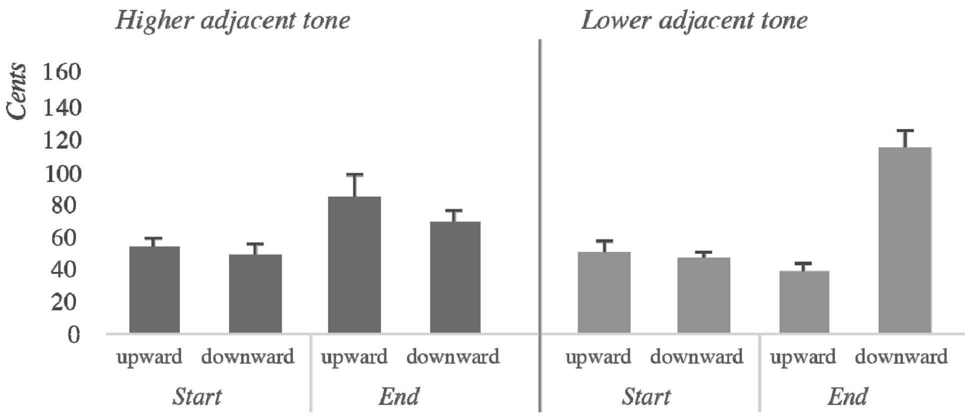

Figure A2. Illustration of the mean amplitude (absolute values, in cents) and standard error of upward and downward scoops at the start and end of the target tone, depending on the pitch height of the tones adjacent to the scoop (i.e., higher or lower than the target tone).

(*Appendices continue*)

(VAF) of the nonlinear model was then computed on the basis of the sum of squared errors between the model and the data (SSE), and the total sum of squared deviations from the mean in the data (SST): VAF = 1 − (SSE/SST). To avoid bias attributable to low fitting and the risk of unrealistic values, a conservative threshold of VAF >25% was chosen. Based on the degrees of freedom present in individual fits (*M* number of samples used in fit = 89), this is a highly conservative criterion. Given this criterion, 1461 tones (78% of the sample of analyzable tones) yielded good fits with VAFs ranging from 0.251 to 0.994 ($M = 0.62$, $SD = 0.22$). Parameters from these fits were used to describe characteristics of scoops in the general population.

Description of the Scoops

The two parameters relative to scoops (i.e., *A*: magnitude of the scoop, *b*: rate of the scoop), were individually subjected to repeated measures ANOVAs with direction (upward vs. downward), position (start vs. end), and context (adjacent tone above vs. under the target tone) as within-subjects variables.

The ANOVA on scoop magnitude (*A*) yielded a significant main effect of direction, with larger magnitude for downward scoops ($M = 89.67$, $SD = 35.36$, $SE = 6.57$) than for upward scoops ($M = 72.83$, $SD = 30.46$, $SE = 5.66$), $F(1, 27) = 8.94$, $p = .006$, $\eta_p^2 = .249$. The main effect of position was also significant, with larger magnitude for ending scoops ($M = 99.23$, $SD = 46.60$, $SE = 8.65$) than for starting scoops ($M = 63.27$, $SD = 22.76$, $SE = 4.23$), $F(1, 27) = 21.18$, $p < 001$, $\eta_p^2 = .440$. Most importantly, this analysis revealed a significant three way interaction among position, direction, and context, $F(1, 27) = 34.89$, $p < .001$, $\eta_p^2 = 564$. As illustrated in Figure A2, the magnitude of the scoop varied greatly depending on the position, direction, and surrounding context. Therefore, it seemed necessary to apply specific values in the creation of the material to be used in listening tasks.

Although the magnitude of measured scoops varied considerably, it is important to note that every mean shown in Figure A2 is significantly greater than zero ($p < .01$ for all one-sampled *t* tests with a test value of 0). Thus, scoops are a significant factor in pitch production, even though these portions of sung pitches are typically removed from analyses of singing accuracy.

With regard to the rate of the scoop to reach or following the asymptote (*b*), all main effects (i.e., direction, position, and context) as well as interactions were nonsignificant ($p > .05$). Thus, in the stimuli we developed, the rate parameter was fixed for all scoops at about 6.76.

Table A1
Average (Standard Error) of the Magnitude of Scoops (in Cents) of the 1461 Tones Analyzed (i.e., Not Aggregated per Singer), According to the Position, Direction, and Pitch Height of the Adjacent Tone

		Relation of adjacent tone	
Scoop position	Scoop direction	Higher pitch	Lower pitch
Start of tone	Upwards	68 (4)	74 (5)
	Downwards	71 (4)	58 (3)
End of tone	Upwards	115 (11)	57 (11)
	Downwards	86 (4)	159 (5)

Pitch Manipulations of the Melodic Material

The analysis described above suggests that singers vary primarily in the magnitude of scoops but not in the duration between the start or end and the stable part of the tone. To examine the perception of "natural" scoops, the synthesized scoops of the main study were always inserted into the initial or final 220 ms of tones—just under 25% of the total duration, whatever the position, the direction, and the context of the scoop. Transitions were based on an exponential curve from the start of the scoop to the asymptote, similar to the model used to describe recorded singing. In keeping with the model fits reported above, scoop magnitude depended on the position and direction of the inserted scoops, as well as on the melodic context of the stimuli to be evaluated (Figure A2). To present representative magnitude of scoops, we computed the average of the magnitude values depending on the position, direction, and surrounding melodic context, on the raw data (i.e., non aggregated, $n = 1461$) of Pfordresher and Mantell (2014). The magnitude of scoop (Table 1, also reported in the main article) ranged from 58 to 159 cents.

By taking into account the effects (i.e., position, direction, and relation to the surrounding context) observed in the singing performances of occasional singers, our pitch manipulations were expected to be natural and thus appropriate to examine the perception of scoops in melodic contexts.

(*Appendices continue*)

Appendix B

Descriptive Statistic of Experiments 1, 2, and 3

Table B1
Means and Standard Deviations for All Conditions in Experiment 1

			Starting scoop		Ending scoop	
Melody	Asymptote	None	Up	Down	Up	Down
1	Flat	*M* = 5.90	4.31	4.75	5.04	4.52
		SD = 1.79	1.92	1.44	1.46	1.80
	In-tune	*M* = 11.42	11.06	9.38	8.96	9.04
		SD = 1.82	1.85	1.73	2.47	1.89
	Sharp	*M* = 7.35	6.96	5.02	5.50	5.79
		SD = 2.11	1.64	2.59	1.93	2.12
2	Flat	*M* = 6.04	4.06	5.81	5.98	3.94
		SD = 2.19	2.22	1.56	1.80	2.92
	In-tune	*M* = 10.27	8.46	9.60	8.98	6.73
		SD = 2.09	1.85	1.79	2.13	2.16
	Sharp	*M* = 6.69	7.21	7.94	7.56	5.73
		SD = 2.75	2.51	2.77	2.44	2.45

Table B2
Means and Standard Deviations for All Conditions in Experiment 2

		Starting scoop		
Melody	Ending scoop	None	Up	Down
1	None	*M* = 6.19	5.50	4.23
		SD = 1.41	1.44	1.39
	Up	*M* = 4.10	3.54	2.65
		SD = 1.22	1.09	1.38
	Down	*M* = 3.69	3.33	2.77
		SD = 1.37	1.46	1.23
2	None	*M* = 5.63	4.15	5.24
		SD = 1.25	1.52	1.42
	Up	*M* = 5.41	3.63	4.39
		SD = 1.25	1.30	1.34
	Down	*M* = 2.78	2.31	2.46
		SD = 1.27	1.52	1.74

Table B3
Means and Standard Deviations for All Conditions in Experiment 3

		Starting = up		Starting = down	
Melody	Asymptote	End = up	End = down	End = up	End = down
1	Flat	*M* = 5.00	3.50	5.22	3.40
		SD = 1.42	1.80	1.68	1.77
	Sharp	*M* = 5.62	5.82	3.78	3.78
		SD = 2.01	1.27	2.13	1.93
2	Flat	*M* = 4.48	3.54	4.80	3.08
		SD = 1.88	1.99	1.75	1.75
	Sharp	*M* = 6.18	5.64	6.94	4.80
		SD = 2.24	1.69	2.76	2.23
3	Flat	*M* = 3.33	3.78	5.29	4.93
		SD = 1.48	2.14	1.84	1.50
	Sharp	*M* = 5.95	6.40	5.69	5.95
		SD = 1.48	1.23	1.84	1.64
4	Flat	*M* = 4.10	3.12	6.66	3.60
		SD = 1.67	1.60	1.65	1.69
	Sharp	*M* = 6.19	5.29	6.31	5.34
		SD = 1.50	1.31	1.73	1.32

Received December 16, 2017
Revision received February 12, 2018
Accepted March 27, 2018 ■