A note by any other name: Intonation context rapidly changes absolute note judgments

Stephen C. Van Hedger* [1] [2]
Shannon L.M. Heald [1] [2]
Sophia Uddin [1] [2]
Howard C. Nusbaum [1] [2]

[1] *Department of Psychology: The University of Chicago:* Chicago, IL, USA
[2] *Center for Practical Wisdom: The University of Chicago:* Chicago, IL, USA

* *Corresponding author*
5848 S. University Avenue G302
Chicago, IL USA
svanhedger@uchicago.edu

Abstract

Absolute pitch (AP) judgments by definition do not require a reference note, and thus might be viewed as context-independent. Here, we specifically test whether short-term exposure to particular intonation contexts influences AP categorization on a rapid timescale, and whether such context effects can change from moment to moment. In Experiment 1, participants heard duets in which a “lead” instrument always began before a “secondary” instrument. Both instruments independently varied on intonation (flat, in-tune, or sharp). Despite participants being instructed to judge only the intonation of the secondary instrument, we found that participants treated the lead instrument’s intonation as “in-tune” and intonation judgments of the secondary instrument were relativized against this standard. In Experiment 2, participants heard a short antecedent context melody (flat, in-tune, or sharp) followed by an isolated target note (flat, in-tune, or sharp). Target note intonation judgments were once again relativized against the context melody’s intonation, though only for notes that were experienced in the context or implied by the context key signature. Moreover, maximally contrastive intonation combinations of context and target engendered systematic note misclassifications. For example, a flat melody resulted in a greater likelihood of misclassifying a “sharp F-sharp” as a “G”. These results highlight that both intonation and note category judgments among AP possessors are rapidly modified by the listening environment on the order of seconds, arguing against an invariant mental representation of the absolute pitches of notes. Implications for general auditory theories of perception are discussed.

Keywords: absolute pitch, perception, plasticity, categorization, adaptation

Public Significance Statement

Our experiments demonstrate that the rare ability of absolute pitch is highly influenced by the immediate listening environment. Our results highlight the dynamic nature of absolute pitch perception and refute the notion that absolute pitch is insensitive to auditory context.

Absolute pitch (AP) is conventionally defined as the rare ability to name or produce any musical note without a reference note (see Deutsch, 2013 for a review). While the mechanisms underlying the acquisition of AP are still debated, an influential body of work posits that learning pitch-label associations within a critical period of development is necessary, though perhaps not sufficient, for successful AP acquisition (e.g., Crozier, 1997; Levitin & Zatorre, 2003; Zatorre, 2003, though see Ross, Olson, & Gore, 2003). Congruent with the rhetoric of critical periods, AP categories within an adult population are often discussed as being fixed in structure and content (e.g., Ward & Burns, 1982). While some exceptions to this immutability have been discovered, including systematic AP distortions caused by aging (Athos et al., 2007; Vernon, 1977), side effects from particular drugs (Gur-Ozmen, Nirmalananthan, & Von Oertzen, 2013; Kobayashi, Nisijima, Ehara, Otsuka, & Kato, 2001; Konno, Yamazaki, Kudoh, Abe, & Tohgi, 2003), and even reproductive cycles (Wynn, 1973), these distortions are commonly explained in terms of biochemical or physiological changes to the auditory system (e.g., changes in basilar membrane stiffness to explain age-related AP changes) rather than psychological changes stemming from listening experience or perceptual expectations.

Yet, previous research has clearly established that AP is not, by any means, absolute or “perfect.” AP possessors are faster and more accurate to judge “white key” notes than “black key” notes, with the notes “C” and “G” being easiest to identify (Miyazaki, 1990). Certain timbres, such as sine tones (Lockhead & Byrd, 1981) and sung tones (Vanzella & Schellenberg, 2010) can reduce accurate note identification for AP possessors compared to instrument timbres such as piano or violin. Moreover, AP possessors are often more accurate at identifying notes that come from their primary (most played) instrument (Bahr, Christensen, & Bahr, 2005). While these observations suggest that both systematic and idiosyncratic factors shape any particular individual’s AP ability, they do not inherently refute a crystallized, immutable view of AP, as it could be the case that experience with particular notes and timbres within a malleable, critical period of learning are what ultimately influence an adult’s AP “fingerprint” (cf. Miyazaki, 1990).

An alternative explanation, however, is that AP categories are constantly influenced by the environment, strengthening and weakening based on distributional experiences playing and hearing particular pitch classes and instrumental timbres. From this perspective, then, pitch-label associations may appear to be stable due to consistencies in one’s listening environment rather than solidified mental representations. Indeed, AP possessors appear perform best on octave ranges and instrumental timbres that reflect their *recent* musical activity (Bahr et al., 2005). Treating AP categories as constantly shaped through statistical learning mechanisms (cf. Saffran, Aslin, & Newport, 1996) allows for a more parsimonious explanation of recent research, demonstrating that particular experiences (or a lack thereof) can actually be detrimental to accurate AP categorization.

For example, Dohn, Garza-Villarreal, Riisgaard Ribe, Wallentin, and Vuust (2014) found that recent musical activity “tunes up” absolute pitch ability, in that AP musicians who were actively playing an instrument were able to more accurately generate pitches (via a computerized pitch oscillator) compared to AP musicians who did not have extensive recent musical activity. These results suggest that AP ability perhaps should be considered a skill, strengthened by particular sensorimotor experiences even in adults (Heald, Van Hedger, & Nusbaum, 2017b. In a similar vein, Wilson, Lusher, Martin, Rayner, and McLachlan (2012) found evidence that AP is best maintained in a “fixed *do*” environment (e.g., when playing an instrument in which the culturally accepted pitch-label associations are maintained). These results, taken together, suggest that the reinforcement of pitch-label associations is important in maintaining AP ability.

While this evidence indicates that AP categories are continuously stabilized through the relative constancy of the typical listening environment, it does not address if AP categories can be influenced or altered by short-term changes in listening experience once they are established. One means of assessing short-term modifications to AP categories is through the perception of intonation. Indeed, an implied subcomponent of AP is the ability to provide a goodness rating of an isolated note reflecting its intonation (e.g., Levitin & Rogers, 2005). Given that the vast majority of Western tonal music adheres to an absolute tuning standard (in which the A above “middle C” is tuned to 440 Hz), it is possible that this absolute intonation ability is not fixed, but rather continuously reinforced by a consistently tuned listening environment. If the manipulation of intonation in the immediate listening environment affects the intonation ratings of adult AP listeners, this would indicate that AP categories remain mutable throughout life, and are shaped by to short-term changes in one’s listening environment.

Indeed, we have previously found that after listening to a symphony that had been flattened by one-third of a semitone (an approximately 2% change in auditory frequency), AP participants’ intonation judgments of subsequent isolated notes were recalibrated to the tuning of the symphony, such that “flat” notes were judged to be significantly more “in-tune” after listening to the symphony (Hedger, Heald, & Nusbaum, 2013). Moreover, these results demonstrated that the effect of hearing flattened music is not restricted to the experienced notes, but generalizes to notes that were not heard, albeit not to unheard timbres. These results demonstrate that the immediate listening environment is important in holding AP categories in place, and therefore support the idea that AP categories are maintained by the listening environment. However, this research used exposure on the order of tens of minutes, suggesting that the tuning environment causes alterations to AP categories via a relatively slow adaptation process.

The present experiments go beyond the demonstration of context effects on AP, and explore whether intonation context can influence AP judgments on an even more rapid timescale (i.e., on a trial-to-trial basis), as adopting this kind of timescale provides several advantages for understanding how AP categories are influenced by the listening environment. First, by varying intonation context several times within a single experiment, we are able to better assess whether there are any asymmetries in how preceding intonation influences note perception (e.g., whether establishing a relatively flat intonation context changes AP judgments more or less than establishing a relatively sharp intonation context). Second, our prior work does not make it clear how much prior intonation context is sufficient for influencing note category judgments. While Hedger et al. (2013) had participants listen to flattened music for approximately 40 minutes, it is possible that less exposure can reliably influence note intonation judgments, particularly if intonation context effects in AP categorization are similar to context effects in speech categorization, which can be observed on the order of seconds (e.g., Watkins, 2005; Watkins & Makin, 1994) and even fractions of a second (e.g., see Holt & Kluender, 2000; Sawusch & Nusbaum, 1979; Sawusch, Nusbaum, & Schwab, 1980). Third, rapid manipulations of intonation context will allow for a more systematic investigation of generalization to unheard timbres and unheard note categories, which are separately considered in Experiments 1 and 2, respectively. Taken together, the present experiments will not only further our understanding of category malleability within an AP population, but will also potentially help situate AP within a larger framework of context effects on auditory perception and perceptual plasticity (cf. Heald, Van Hedger, & Nusbaum, 2017a; Heald, Van Hedger, & Nusbaum, 2017b; Lotto & Holt, 2006).

Experiment 1

In Experiment 1, AP possessors judged the intonation of several short duets. These duets consisted of a target “secondary instrument,” which was always delayed in onset relative to a context “lead instrument.” Participants were explicitly instructed to judge the intonation of the secondary instrument. Crucially, the lead and secondary instruments were independently manipulated to be flat, in-tune, or sharp relative to a canonical tuning standard (A = 440 Hz), which allowed us to test whether the intonation of the context lead instrument systematically influenced ratings of the target secondary instrument, despite task instructions and the relatively short duration of each trial. Additionally, the use of duets provides a more ecologically valid means of testing whether establishing a particular intonation context in one instrument generalizes to a secondary instrument (cf. Hedger et al., 2013).

Depending on the degree to which participants’ intonation judgments are context-independent versus context-dependent, we would hypothesize fundamentally different effects of the lead instrument’s intonation on secondary instrument intonation judgments (see Figure 1). If participants’ judgments are context-independent, intonation judgments of the secondary instrument should be consistent and accurate (e.g., if the secondary instrument is flat, it should be rated as flat regardless of the lead instrument’s intonation). If, however, participants’ judgments are context-dependent, then we would expect identical intonations of the secondary instrument to be rated differently based on the intonation of the lead instrument. More specifically, a flat lead instrument should systematically make the secondary instrument sound sharper, and a sharp lead instrument should systematically make the secondary instrument sound flatter.

While prior work suggests that AP intonation judgments could be context-dependent (Hedger et al., 2013), these results were based on a single manipulation of intonation that lasted over half an hour. Moreover, in that study the context was detuned gradually below a just noticeable difference for frequency discrimination so that listeners might not have been aware of the flattening of the music. Shifting intonation from trial to trial of relatively short contexts does not permit this kind of “trick.” AP listeners should detect the “out-of-tune” aspect of the lead (when that is the case) and therefore may disregard it perceptually. Further, manipulating the intonation of the lead instrument on a trial-by-trial basis may lead to a more context-independent strategy, as the lead instrument, from trial to trial, is “unstable” in its intonation and thus may be disregarded by listeners.

Methods

Participants

We recruited 17 AP participants for Experiment 1 (M_{age} ± *SD*: 25.76 ± 5.63 years, range: 19 to 36 years, 7 male, 10 female). Participants were recruited from an existing database of AP possessors (based on AP self-identification) and were required to perform sufficiently well on an AP test to be included in the experiment (see Procedure for details). Participants had an average of 19.59 years (*SD:* 6.59, range: 7 to 32 years) of training on their primary musical instruments. The mean age of beginning musical instruction was 4.88 years old (*SD:* 1.76, range: 2 to 9 years old), and 12 of the 17 participants completed (or were working towards completing) a college degree in music. All participants were compensated for their participation in the experiment and were treated in accordance with the NIH guidelines for interacting with human participants. The experimental protocol was approved by the University of Chicago Social and Behavioral Sciences Institutional Review Board.

Sample Size Determination

Across both experiments in the present paper, we determined our sample size from a power analysis of our previous study (Hedger et al., 2013). Given that our previous study only presented participants with flattened intonation and had participants respond on an “in-tune / out-of-tune” scale (i.e., not responding whether an out-of-tune note was “flat” or “sharp”), we calculated post-hoc power from this study by analyzing the difference in intonation judgments of just the flat notes before and after listening to flattened music. In both Experiment 1 and 2, participants rated flat notes as significantly more “in-tune” post-listening, and the statistical power was 0.724 and 0.937 with 13 and 14 participants, respectively. Given the estimated effect size (averaged across both experiments), we would require a minimum sample size of 12 participants to achieve power of 0.8 and a minimum sample size of 16 to achieve power of 0.9. As such, we set our sample size at 17 participants. We took this relatively conservative approach for two reasons. First, given the differences in experimental design, we wanted to provide sufficient power our current experiments even if the effect size was smaller than what was found in Hedger et al. (2013). Second, the relatively limited sample sizes of absolute pitch studies (given the rarity of the ability), increases the possibility that observed effects sizes may be less precise (e.g., Asendorpf et al., 2013).

Materials

The experiment was coded in Inquisit 4 (Millisecond Software: Seattle, WA) and was administered online. The 54 duet stimuli were composed by the first author and two undergraduate research assistants. Each duet was specifically composed so that the entrance of the two instruments was staggered, resulting in both a “lead” and “secondary” instrument. The duets had an average length of 18.16 seconds (*SD*: 5.16, range: 8.28 to 30.64 seconds), and the secondary instrument began an average of 7.01 seconds after the onset of the lead instrument (*SD*: 2.74, range: 2.39 to 14.94 seconds). The duets were composed so that the lead and secondary instruments were not identical in timbre, though 8 duets consisted of instruments that came from the same general instrument family (e.g., both were brass instruments). The duets were digitally notated and converted to MIDI files, where we then used the sound library of Reason 4 (Propellerhead: Stockholm, Sweden) to assign instrument sounds to each instrument. Each instrument was separately exported as an audio file as this allowed us to independently manipulate the intonation of both the lead and secondary instrument for each duet.

We pitch shifted each instrument of the duets using the “Change Pitch” function in Audacity, an open-source program for audio recording and editing. This pitch-shifting function uses the SoundTouch Audio Processing Library (http://www.surina.net/soundtouch/), which in turn manipulates pitch through the Pitch Synchronous Overlap and Add (PSOLA) algorithm, and does not significantly affect the overall tempo of the recording. Moreover, PSOLA is well suited for small pitch shifts such as those used in the present experiment. Duets were either shifted up (sharp) or down (flat) by one-third of one semitone (33 cents, approximately a 2% change in auditory frequency) in two steps of 16.5 cents. In-tune stimuli also underwent pitch shifting (either up by 16.5 cents and then down by 16.5 cents, or vice versa) to minimize the chances of the pitch-shifting process introducing artifacts that could be used to differentiate in-tune from out-of-tune stimuli. Thus, all duet instruments had undergone the pitch shifting procedure exactly twice under the same pitch shifting magnitude. This general approach of shifting all stimuli has been previously used (with the same pitch-shifting algorithm) to demonstrate that “original” (correctly-tuned) stimuli cannot be differentiated from “altered” (incorrectly-tuned) stimuli solely based on acoustic features (Van Hedger, Heald, & Nusbaum, 2017). After creating flat, sharp, and in-tune versions of each of the lead and secondary instruments, we combined the lead and secondary instruments for each duet in Adobe Audition. Each duet thus had nine versions (as both the lead and secondary instruments could be flat, in-tune or sharp in a fully-crossed manner).

We used 24 notes for the AP prescreening test. All AP prescreening notes were 400ms in duration. Half of the notes were non-instrumental triangle tones, generated in Reason 4, in which a fundamental frequency was accompanied by odd harmonics (giving the waveform a triangular shape). . These notes ranged from C4 (261.63 Hz) to B4 (493.88 Hz) and had a zero-length attack, 360ms sustain, and 40ms decay. The remaining 12 notes, which were also created in Reason 4, consisted of a harpsichord timbre and ranged from C5 (523.25 Hz) to B5 (987.77 Hz). As these notes were sampled from a real harpsichord, they had a sharp attack (approximately 6ms), immediately followed by a long decay (approximately 394ms) with no true sustain. Between trials, listeners were presented with a selection of 50 microtonal pitches (200ms in duration) and a 5-second sample of white noise, which were both generated in Adobe Audition. The microtonal pitches divided the octave in approximately 20-cent increments (300 Hz to 592 Hz) and consisted of the fundamental frequency and the first four harmonics, while the white noise was created using the “generate noise” function. All audio files had a 44.1 kHz sampling rate, 16-bit depth and were RMS normalized to the same digital level (-15 dBFS). However, given the online nature of the experiment, participants were ultimately responsible for adjusting their computer volume to a comfortable listening level to complete the experiment..

Procedure

Participants first completed the AP prescreening measure. The instructions stated that participants would be labeling 24 musical notes with their respective note names. Additionally, participants were told that they had to achieve an accuracy of 70% or higher and respond, on average, in under 5 seconds to qualify for the study. Qualifications of both speed and accuracy have been used previously in online tests of AP (e.g., Athos et al., 2007). Participants responded to each note by entering a number on their computer keyboard corresponding to a particular pitch class (in which “C” was assigned 0, “C#” was assigned 1 etc.). Given that this method of response was likely not familiar to participants, we displayed a pitch wheel at all times in the center of the screen, in which the pitch class / number associations were prominently displayed. On each trial, participants would hear an isolated musical note (randomly selected from our set of 24) and then have to respond as quickly and accurately as possible. Trial-by-trial feedback was not given, and we played 1500ms of white noise between trials. At the end of the prescreening measure, all participants received feedback (both overall accuracy and mean response time).

The instructions of the duet task told participants to imagine that they were judging a middle school band competition that consisted of a series of instrumental duets. Participants were specifically told that they were assigned to judge the intonation of the second instrument *only*, and that they should make their judgments based on their initial feeling. The intonation of the lead instrument was not mentioned (i.e., we did not notify participants that the lead instrument could be out of tune, nor did we state that the lead instrument would always be in-tune). We encouraged participants to rely on an initial feeling because we did not want them to rely on a post-perceptual decision-making strategy to infer the intonation of the secondary instrument on the basis of the lead instrument given that its intonation was varied across secondary instrument intonation. Participants heard each of the 54 duets in a random order. For each duet, we pseudo-randomly selected one of the nine possible intonation combinations (flat, in-tune, or sharp lead instrument * flat, in-tune, or sharp secondary instrument), such that participants were presented with exactly six trials of each of the nine intonation combinations. After each duet, participants judged the secondary instrument’s intonation on a three-point scale, in which -1 corresponded to “too flat,” 0 corresponded to “perfectly in-tune and 1 corresponded to “too sharp.” Between duets, participants heard 25 consecutive microtonal pitches, which were randomly selected from our 50 microtonal pitches. Each pitch was 200ms in duration, meaning participants heard a total of 5s of these microtonal pitches between each duet. After the microtonal pitches, participants also heard 5s of white noise. These inter-trial sounds were meant to disrupt short-term auditory memory, which could potentially cause prior trials’ intonation to influence current responses. Figure 2a provides a sketch of the experimental design for the duet portion of the experiment.

After participants judged all 54 duet stimuli, we collected basic demographic and musical experience information, such as age, sex, age of first beginning musical instruction (age of onset), name and number of years on one’s primary musical instrument, self-reported AP and musical abilities (on scales of 0-100), tonal language proficiency, handedness, and whether participants had (or were working toward) a college degree in music. These measures were not specifically collected because of any a priori hypotheses related to intonation context – rather, we included this component of the experiment to provide a more detailed description of our AP participants. At the conclusion of the questionnaire, participants were debriefed and given a unique code to email to the experimenter, and after doing so they were compensated for participating.

Statistical Analyses

To test for the influences of intonation context on ratings across Experiments 1 and 2, we used a cumulative link mixed model (CLMM) from the {ordinal} package in R (Christensen, 2011). We chose this type of model given the nature of participants’ responses (-1 = “flat”, 0 = “in-tune,” 1 = “sharp), which were discrete but clearly ordinal. Additionally, the mixed nature of the model allowed us to model random intercepts and slopes, which is advantageous as it allows both participant and item variation to be accounted for in the same model (i.e. we did not have to collapse across participants or duets to obtain mean scores). The fixed-effect betas from a CLMM reflect the log odds of responding “sharp” versus “in-tune” or “flat”, as well as the log odds of responding “sharp” or “in-tune” versus “flat.” Given the model and the nature of participants’ responses, a positive beta value thus reflects the log odds of systematically responding higher (sharp), whereas a negative beta value would reflect the log odds of systematically responding lower (flat). To assess significance, we computed bootstrapped 95% confidence intervals of the fixed effects with 10,000 simulations. For interpretability, we converted the log odds into odds ratios.

Results

AP Prescreening Measure

Given that participants were recruited from a database of AP possessors, it is perhaps not surprising that all participants performed sufficiently well to pass the prescreening measure. Participants correctly identified 95.34% (*SE:* 1.28%) of notes, with the lowest-scoring participant correctly identifying 83.33% of notes. Moreover, participants' average response time was 3.53 (*SE*: 0.14) seconds, with the slowest participant responding on average in 4.53 seconds.

Intonation Judgments

In our model, we created dummy variables for flat and sharp lead instrument intonation, which reflected the extent to which flat and sharp lead instruments differentially influenced intonation ratings compared to in-tune lead instruments. We also included dummy variables that specified whether the secondary instrument was flat or sharp, as these reflected the extent to which flat and sharp secondary instruments were judged to differ in intonation relative to in-tune instruments. We included random intercepts for both participant and duet, and we additionally included participant random slopes for lead instrument intonation. This random effects structure, while not maximal (Barr, Levy, Scheepers, & Tily, 2013), allowed the model to converge and was thus selected over a maximal random effects structure (Bates, Kliegl, Vasishth, & Baayen, 2015).

Results from the model (Table 1) clearly demonstrated that flat secondary instruments were associated with greater predicted odds of responding flat, while sharp secondary instruments were associated with greater predicted odds of responding sharp. This means participants were able to accurately differentiate flat, in-tune, and sharp melodies (according to an A4 = 440 Hz standard), which is expected of AP possessors (e.g., Levitin & Rogers, 2005). More importantly for our hypothesis, we found evidence that the intonation of the lead instrument significantly influenced intonation judgments contrastively, in that flat lead instruments resulted in greater predicted odds of responding to the target secondary as sharp, while sharp lead instruments resulted in greater predicted odds of responding to the target secondary as flat. Figure 3 plots mean intonation responses as a function of both the lead and secondary instrument's intonation.

Discussion

Experiment 1 provides clear evidence that intonation judgments among AP possessors are relativized against the immediate listening context within a duet. When the immediate contextual listening environment is conventionally “in-tune,” participants have little trouble accurately judging the intonation of flat, in-tune, and sharp melodies. When the immediate contextual listening environment deviates from this canonical tuning, however, participants’ intonation judgments are shifted such that the preceding intonation context appears to serve, de facto, as an “in-tune” standard and judgments are relativized against this standard. More specifically, a flat secondary instrument in the context of a flat lead instrument sounded in-tune, whereas an in-tune secondary instrument sounded sharp, and a sharp secondary instrument in the context of a sharp lead instrument sounded in-tune, whereas an in-tune secondary instrument sounded flat. In this sense, the present results conceptually replicate Hedger et al. (2013) albeit on a trial-by-trial basis and suggest the way in which context influences AP intonation judgments – specifically, through demonstrating (1) that the amount of intonation context necessary to shift intonation perceptions can operate on the order of seconds (rather than 10s of minutes), (2) that both sharp and flat contexts systematically shift intonation perception, (3) that intonation judgments are systematically relativized (rather than simply destabilized) against the preceding intonation context, and (4) that intonation context effects can be observed across different instrumental timbres, at least when presented simultaneously.

There are several potential concerns, however, with the design of Experiment 1. First, given that the target secondary instrument was always played simultaneously with the context lead instrument, it is possible that our results were confounded with the extent to which participants could successfully selectively attend to the secondary instrument. Second, despite the instructions stating that participants should solely judge the secondary instrument based on an “A4 = 440Hz” standard, it is possible that this objective was particularly confusing to participants, especially since the lead and secondary instruments formed a cohesive musical duet. Third, without having participants identify the category identity of the duets (e.g., through providing an explicit key signature), it is not clear to what extent category identity accuracy influenced intonation judgments, particularly for trials in which the lead and secondary instrument were maximally contrastive in intonation (e.g., flat lead and sharp secondary). Fourth, the overall length of the duets and the amount of intonation context provided by the lead instrument was highly variable, making it difficult to draw conclusions regarding the relative amount of context necessary to shift intonation judgments. These concerns are directly addressed in Experiment 2.

Experiment 2

Experiment 2 was designed to further clarify the role of the immediate listening context on AP category judgments. Specifically, we controlled the amount of preceding intonation context (always 12 notes, 6 seconds), and we had participants provide both a note name and intonation judgment for a subsequently presented single musical target note. These changes were implemented for three primary reasons. First, temporally separating the intonation context from the to-be-judged target stimulus will help us interpret whether the results from Experiment 1 were driven, in part, by the simultaneous presentation of melodies, which could have been difficult for participants to distinguish and thus cause confusion regarding the task, or encourage the direct use of intervals between the instruments. Second, by having participants provide both a note name and intonation judgment, we can better assess whether intonation judgments are sufficiently plastic that they can influence note category determinations. Put another way, for conditions in which the difference between the context and to-be-judged note is greater than 50 cents (flat context with sharp note and sharp context with flat note), it is unclear whether participants would provide a “correct” note label, compared to systematically misclassifying a note as belonging to an adjacent category. Third, divorcing the intonation context from the to-be-judged stimulus allows for greater flexibility in manipulating the tonal relationship between the two events. Thus, in the present experiment, we tested whether the preceding intonation context differentially influenced to-be-judged notes based on whether the notes had been heard as a part of the preceding context.

Methods

Participants

We recruited 17 AP participants for Experiment 2 (M_{age} ± *SD*: 25.00 ± 8.28 years, range: 18 to 51 years, 5 male, 12 female). Participants were recruited from the same database of AP possessors as was used in Experiment 1, though most (15 of 17) had not participated in Experiment 1. Participants had an average of 18 years (*SD:* 10.33, range: 3 to 48 years) of training on their primary musical instruments. The mean age of beginning musical instruction was 5.17 years old (*SD:* 1.91, range: 2 to 9 years old), and 9 of the 17 participants completed (or were working towards completing) a college degree in music. We set the sample size based on Experiment 1. All participants were compensated for their participation in the experiment and were treated in accordance with the NIH guidelines for interacting with human participants. The experimental protocol was approved by the University of Chicago Social and Behavioral Sciences Institutional Review Board.

Materials

The experiment was coded in Inquisit 4 (Millisecond Software: Seattle, WA) and was administered online. The 54 melodies were composed by the first author and two undergraduate research assistants. Each melody consisted of exactly 12 quarter notes played at 120 BPM, resulting in melodies that were six seconds in duration. The melodies were composed in either C major (n = 27) or A natural minor (n = 27) and did not contain any accidentals (e.g., G#). We chose these particular keys because we wanted to assess whether intonation context would influence these unheard note categories that were not implied by the key signature (i.e., “black key” notes), despite participants never hearing these note categories as sharp or flat (cf. Hedger et al., 2013). The melodies were digitally notated and converted to MIDI files, where we then used the sound library of Reason 4 (Propellerhead: Stockholm, Sweden) to export each melody as an audio file with a violin timbre. The melodies were processed in a similar manner as in Experiment 1, with the exception that we did not need to split and recombine each melody (as there was only a single instrument). All melodies were shifted by 16.5 cents in Audacity twice (16.5c + 16.5c for sharp melodies, -16.5 + -16.5c for flat melodies, and either 16.5c – 16.5c or -16.5c + 16.5c for in-tune melodies).

The target notes for which participants had to provide both a note category and intonation label were created using the same violin timbre as the melodies and were 500ms in duration. We tested six notes (F 349.23 Hz to B-flat 466.16 Hz, inclusive). We chose to test this subset of notes because they included three white-key notes (F, G, and A) as well as three black-key notes (F#, G#, and A#), allowing us to test for influences of intonation on heard as well as unheard note categories. The isolated notes were pitch shifted in an identical manner as the melodies. The white noise and intermediary microtones, which were played between trials to minimize between-trial intonation influences, were identical to those used in Experiment 1. Similar to Experiment 1, all audio files were RMS normalized to the same digital level (-15 dBFS). However, given the online nature of the experiment, participants were ultimately responsible for adjusting their computer volume to a comfortable listening level to complete the experiment. All auditory files had a 44.1 kHz sampling rate with 16-bit depth.

Procedure

Given that participants provided note category labels in the present experiment as a part of the main task, we did not administer an AP prescreening measure (as AP ability was a necessary component of successfully completing the task). After providing informed consent, participants were instructed that they would hear short violin melodies, which would be followed by a single note. Participants were further instructed that they would be asked to provide both a note name and intonation judgment for this final note. On each trial, participants heard one of the 54 melodies (randomly selected), which could either be flat, in-tune, or sharp (pseudo-randomly selected to yield exactly 18 flat, 18 in-tune, and 18 sharp melodies). Each melody was followed by a 250ms pause, after which participants heard the isolated musical note. Participants were asked to provide intonation and note category judgments (in this order). Intonation judgments were made on the same scale as Experiment 1 (-1 for “flat,” 0 for “in-tune,” and 1 “sharp”). Note category judgments were freely typed by participants. After participants made both of these judgments, we played 10 seconds of inter-trial noise (5s of scrambled microtones, 5s of white noise) to minimize between-trial intonation influences (in an identical manner as Experiment 1). Figure 2b provides a sketch of the experimental design.

After participants had completed all trials, we collected basic demographic and musical information (age, sex, handedness, tonal language proficiency, age of music onset, primary instrument, number of years on primary instrument, self-reported AP ability, and self-reported musical ability). Similar to Experiment 1, this information was not specifically collected because of any a priori hypotheses related to intonation context – rather, we included this component of the experiment to provide a more detailed description of our AP participants. After the questionnaire, participants were given a unique identifier to ensure that they were paid for their participation.

Results

AP Assessment

We first assessed whether our participants were sufficiently accurate in labeling notes to be considered AP possessors and therefore included in further analyses. Given that the intonation of the preceding melody could potentially influence AP accuracy, we only considered the 18 trials that had an in-tune preceding melody. Participants accurately labeled an average of 91.83% of these notes (*SE*: 1.27%, range: 83.33% to 100%) and never missed by more than one semitone. As such, all participants were included in further analyses.

Intonation Judgments

Similar to Experiment 1, we constructed cumulative link mixed models (CLMM) to assess whether intonation judgments of the isolated notes were influenced by the intonation of the preceding melodic context. We only included trials in which participants correctly identified the note, as incorrect note classifications would systematically influence intonation responses. Unlike Experiment 1, we constructed three models, splitting target notes into three categories based on their relationship with the preceding melody. The first model (Heard Model) contained target notes that were physically present in the preceding melody (e.g., if the melody contained a “G” and “G” was the target note). The second model (Near Transfer Model) contained target notes not physically present in the preceding melody, but implied by the key signature (e.g., if the melody did not contain a “G” and “G” was the target note). The third model (Far Transfer Model) contained target notes that were not only not physically present in the preceding melody but also not implied by the key signature (which, given our key signatures of C major and A natural minor, were target notes F#, G#, and A#). In terms of the relative proportions of target note categories, 24.6% of trials contained target notes that were physically heard in the preceding melody, 25.4% of trials contained target notes that were not physically heard in the preceding melody but were implied by the key signature, and 50% of trials contained target notes that were both not physically heard in the preceding melody and not implied by the key signature.

Similar to Experiment 1, for all three models participants’ intonation responses were the dependent variable (-1, 0, or 1, treated as ordinal). Dummy variables for flat melodic context and sharp melodic context (coded as 1 or 0) reflected the extent to which flat and sharp melodies differentially influenced intonation ratings compared to in-tune melodies. We also included dummy variables that specified whether the to-be-judged note was flat or sharp (coded as 1 or 0), as these reflected the extent to which flat and sharp notes were judged to differ in intonation relative to the in-tune notes (which, given task instructions, would be expected). We included random intercepts for participants and melodies, as well as random slopes of flat and sharp melodic intonation context for participants.

In the Heard Model (Table 1, top), flat notes had significantly increased odds for a lower (flat) rating compared to in-tune notes, while sharp notes has significantly increased odds for a higher (sharp) rating compared to in-tune notes, meaning participants were able to accurately differentiate flat, in-tune, and sharp notes. Crucially, we also found evidence that the intonation context of the preceding melody significantly influenced intonation judgments contrastively, in that flat melodies resulted in significantly increased odds for a higher rating compared to in-tune melodies, while sharp melodies resulted in significantly increased odds for a lower rating relative to in-tune melodies. As a whole, these results are consistent with Experiment 1 in that the preceding intonation appears to have relativized intonation judgments of the subsequent, to-be-judged note. The condition means are plotted in Figure 4A.

In the Near Transfer Model (Table 2, middle), flat notes had significantly increased odds for a lower rating compared to in-tune notes, while sharp notes had significantly increased odds for a higher rating compared to in-tune notes, once again suggesting a general ability to differentiate flat and sharp notes from in-tune notes. In terms of preceding melodic intonation context influencing subsequent intonation judgments, flat melodic contexts resulted in increased odds for a higher rating, and sharp melodic contexts resulted in increased odds for a lower rating, with both confidence intervals not including one. As such, the intonation of the preceding melody appeared to relativize intonation judgments for unheard but implied notes. The condition means are plotted in Figure 4B.

In the Far Transfer Model (Table 2, bottom), flat notes had significantly increased odds for a lower rating compared to in-tune notes, while sharp notes had significantly increased odds for a higher rating compared to in-tune notes, once again suggesting a general ability to differentiate flat and sharp notes from in-tune notes. However, unlike the Heard and Near Transfer Models, we found no reliable evidence that the intonation of the preceding melody relativized intonation judgments. While flat melodic contexts resulted in slightly increased odds for higher ratings and sharp melodic contexts resulted in slightly increased odds for lower ratings, both confidence intervals included one (i.e., the null hypothesis) and both confidence intervals overlapped with each other. The condition means are plotted in Figure 4C.

Note Accuracy

We also examined participants’ accuracy as a function of both the preceding melodic context and note intonation. While intonation judgments were influenced by the context of the preceding melody (at least for heard and near transfer notes), it is unclear if the preceding melody also influenced the ability to accurately provide a note name for the isolated notes. In particular, if participants relativize their responses against the intonation of the melody, we would predict that the trial types of “flat melody/sharp note” and “sharp melody/flat note” would systematically engender more misclassification errors than any other trial type, given that the absolute distance between stimuli in these trials is greater than 50 cents. For example, if a flat melody systematically makes subsequent notes sound sharp, an already sharp note might begin to sound like a flat member of the adjacent category. On the other hand, a sharp melody preceding a sharp note may result in a lower probability of misclassification, as the intonation context in this case would make the sharp note sound more “in-tune.”

To assess how melodic intonation influenced semitone errors for these particular intonation combinations we constructed linear mixed-effects models using the {lme4} package in R (Bates, Mächler, Bolker, & Walker, 2014). We used a linear mixed model as opposed to the cumulative link mixed model because participants were not inherently range limited given the dependent measure of accuracy.[1] The dependent measure was coded in a manner that preserved the directionality of misclassification errors (semitone too high = 1, semitone too low = -1, correct = 0). This approach is particularly useful for modeling note classification accuracy because we had specific hypotheses about the directionality of note misclassifications within particular conditions. This approach helps rule out alternative explanations of note misclassification errors (e.g., certain trial types causing more semitone errors, but not in a systematic direction). Given the hypotheses regarding note classification accuracy, we analyzed (objectively) sharp and flat notes in separate models. In both models, we included dummy variables for sharp melodic context and flat melodic intonation. Similar to the previous section, we separately analyzed notes that were heard as a part of the preceding melody (Heard Model), notes that were not heard but implied by the key signature (Near Transfer Model), and notes that were not heard and also not implied by the key signature (Far Transfer Model). All models included random intercepts for participants.

The sharp and flat Heard Models both supported the idea that misclassifications were related to conditions in which the intonation between the melody and the to-be-judged note were maximally contrastive. The results of the models are reported in Table 3 (top), and misclassification errors across all conditions for heard notes are plotted in Figure 5A. Flat intonation preceding a sharp, to-be-judged note caused a sharp misclassification error around 68% of the time, while sharp intonation preceding a flat, to-be-judged note caused a flat misclassification error around 48% of the time. As a comparison, sharp and flat notes were misclassified in these directions 17% and 14% of the time, respectively, when preceded by an in-tune melody.

---

[1] It should be noted that all but one note misclassification error was ± 1 semitone from the correct note.

The sharp and flat Near Transfer Models similarly provided evidence that maximal intonation differences between the melody and the to-be-judged note were related to note misclassifications. The results of the models are reported in Table 3 (middle), and misclassification errors across all conditions for near transfer notes are plotted in Figure 5B. Flat intonation preceding a sharp, to-be-judged note caused a sharp misclassification error around 69% of the time, while sharp intonation preceding a flat, to-be-judged note caused a flat misclassification error around 36% of the time. As a comparison, both sharp and flat notes were misclassified in these directions 7% of the time when preceded by an in-tune melody.

Finally, the sharp and flat Far Transfer Models similarly provided evidence that maximal intonation differences between the melody and the to-be-judged note were related to note misclassifications, which is perhaps surprising as the Far Transfer Model for intonation ratings (previous section) did not find evidence that the melodic intonation influenced subsequent intonation ratings of isolated notes. The results of the models are reported in Table 3 (bottom), and misclassification errors across all conditions for far transfer notes are plotted in Figure 5C. Flat intonation preceding a sharp, to-be-judged note caused a sharp misclassification error around 72% of the time, while sharp intonation preceding a flat, to-be-judged note caused a flat misclassification error around 51% of the time. As a comparison, sharp and flat notes were misclassified in these directions 20% and 6% of the time, respectively, when preceded by an in-tune melody.

Discussion

There are several important conclusions to draw from Experiment 2. First, similar to the previous experiment, AP participants' note and intonation judgments were clearly affected by the intonation of a melodic context, despite this intonation context occurring separately in time from the to-be-judged stimulus. This suggests that our results in Experiment 1 could not be entirely explained by difficulty selectively attending to a single instrument in the duets or any effect of presenting simultaneous intervals.

Second, Experiment 2 provides the first demonstration that, beyond intonation ratings, AP category accuracy can be shifted on a relatively short timescale. For trials in which the intonation of the melody and the note were maximally separated – namely, flat melodies paired with sharp notes and sharp melodies paired with flat notes - participants systematically misclassified notes as the adjacent higher or lower note category, respectively. This systematic misclassification was not observed for other trials comprised of sharp and flat melodies, suggesting that the mere act of hearing altered intonation does not always hurt the accurate category identification of a subsequent note, even if it does influence how "in-tune" the note is judged. These error analyses provide particularly strong evidence for the context-dependent nature of AP perception, as prior intonation context in these specific cases can make a typically easy-to-identify note that is canonically thought to be unambiguous sound like an entirely different note.

Third, our results have important implications for how individual AP note categories may be systematically linked to one another as a part of a larger cultural system of Western music. In prior work, flattened intonation context with a limited set of notes was shown to generalize to *all* notes, even those that were never experienced as out-of-tune (Hedger et al., 2013). The present experiment generally replicated this effect, though our results suggest that generalization may be limited to notes that are conceptually related to the preceding melodic context (e.g., part of the same key signature), at least for intonation judgments. In this sense, the present results provide an important clarification on how tonal relatedness and psychological distance between notes (e.g., Krumhansl, 1979) influences the extent to which intonation context generalizes. Moreover, these findings provide an interesting perspective on our previous work (Hedger et al., 2013), in which we tested intonation judgments on all notes after hearing music that only consisted of the first five notes of a minor scale. Given that the first five notes of a minor scale do not clearly distinguish between natural, harmonic, or melodic minor varieties, four of the seven unheard notes were actually conceptually related to the intonation context. As such, the generalization reported in Hedger et al. (2013) may have been largely driven by these four conceptually related notes.

It is also possible, however, that the reason we did not observe influences of intonation context for far-transfer notes is because they were "black key" notes, which are identified more slowly and less accurately compared to white key notes (e.g., Miyazaki, 1989). Given these prior findings, it is possible that black key notes are also harder to accurately identify in terms of intonation, which could have made it more difficult to observe any effects of intonation context. Indeed, the differentiation of sharp and flat target notes (regardless of intonation context), while significant for far-transfer notes, was attenuated relative to heard and near-transfer notes (Table 2). In fact, the confidence intervals for far-transfer flat and sharp notes did not overlap with the confidence intervals for heard notes, suggesting that participants were significantly worse at differentiating sharp, in-tune, and flat notes for far transfer notes compared to heard notes *regardless of intonation context*. Future research could disambiguate the “tonal relationship” and the “black key” explanations through designing a study in which white key target notes could also serve as far-transfer notes. If intonation context remains non-significant, this would strengthen the possibility that the effect is driven by the tonal relationship between the context and the target note, rather than general difficulties making intonation judgments for particular pitch classes.

Regardless of the reason why we did not find contextual effects on intonation ratings for far-transfer notes, these notes (F#, G#, and A#) were still subject to systematic note misclassifications based on the prior intonation context, in that a flat intonation context pushed sharp versions of these notes to be misclassified as their adjacent higher neighbors (G, A, and B) 72% of the time, while a sharp intonation context pushed flat versions of these notes to be misclassified as their adjacent lower neighbors (F, G, and A) 51% of the time. Why would we find no evidence for far transfer when assessing intonation judgments, but strong evidence for far transfer when assessing systematic note classifications? There are two likely possibilities. First, given that these notes are “black key” notes, they might have been overall less stable and more prone to misclassification (cf. Miyazaki, 1989). Second, given the key signature of the prior melodic context, these notes may have been more likely to have been misclassified as a note that was consistent with the preceding melody – a possibility which conceptually supports the near transfer findings in intonation judgments.

General Discussion

The present results clearly demonstrate that, among AP possessors, any particular pitch is not heard in an absolute way, isolated from a listening context, for an identical pitch may be interpreted differently depending on prior musical context. Rather than representing fixed pitch-label mappings that are established early in life, AP categories thus appear to be sensitive to the immediate listening environment. When the typical listening environment reflects dominant cultural tuning norms, AP possessors judge the intonation of melodies and notes to be consistent with these norms. When the listening environment deviates from these cultural tuning norms, AP possessors' intonation judgments are relativized against this listening experience, systematically shifting intonation judgments sharp or flat. These relativized intonation judgments, moreover, have strong influences on note classification accuracy. Sharp notes are consistently heard as members of the adjacent higher note category when preceded by a flat intonation context, and flat notes are consistently heard as members of the adjacent lower note category when preceded by a sharp intonation context. This is the first demonstration that "objective" AP classification accuracy can be systematically disrupted on a rapid timescale.

Our results also have implications for the degree to which AP possessors' relative pitch processing influences absolute identification. While prior research has debated the extent to which AP possessors have intact versus diminished relative pitch abilities (e.g., see Dooley & Deutsch, 2011; Miyazaki, 1995), our results suggest that relative intonation processing was strong enough to reliably influence putatively stable absolute pitch representations. In this sense, our results align with prior literature demonstrating that AP possessors may perceive relative pitches directly (i.e., not mediated through absolute pitch labels), even when they conflict with absolute pitch labels (Benguerel & Westdal, 1991). More specifically, Benguerel and Westdal (1991) used interval stimuli that were tuned in a similar manner as our "flat/sharp" and "sharp/flat" intonation pairs (e.g., a 40-cent-flat "C" could be paired with an adjacent 40-cent-sharp "C-sharp"). From an absolute labeling perspective, the interval between these notes should be predicted by the note labels (i.e., C to C-sharp, or "minor second"). From a relative labeling perspective, the interval between these notes should be predicted by the relative difference between the pitches (i.e., 180 cents, or "major second"). Despite finding that AP possessors could accurately label notes in isolation, the authors found overwhelming evidence that AP possessors were much more likely to make interval judgments based on relative intonation differences compared to absolute note labels. Not only do our present results align with these findings, but they also demonstrate the influence of relative intonation in tasks that specifically instruct participants to discount relative pitch. In this sense, our results confirm a larger than previously believed role for relative pitch processing in absolute pitch perception.

The present experiments, however, are not able to definitively address whether our results are truly perceptually based. This is because we did not have participants provide any kind of intonation judgments for the lead instrument (Experiment 1) or context melody (Experiment 2). Thus, given the relatively small intonation deviations used in our experiments, it is possible that participants simply assumed that the context was always “in-tune” and based their answers accordingly. This would mean that participants might have perceived the objectively correct note, but then responded with the contextually correct note from assuming that the context was always in-tune. The strongest evidence against this explanation, however, is our semitone error analyses from Experiment 2. It is difficult to believe that, as a post-perceptual strategy, an AP possessor would willingly mislabel a note simply to maintain constancy with the preceding context. Moreover, in an exploratory analysis from our own data (see Supplementary Figures 1 and 2), we found evidence that participants tended to respond faster for contextually-based (versus objectively-based) judgments, which tentatively supports the idea that responding in an objective manner (i.e., disregarding the context) may require additional processing. Nevertheless, future research could address this issue by having participants also judge the intonation of the context, assessing whether it is always perceived as “in-tune,” or even simply alerting participants to the fact that the preceding context will not always be in-tune.

The current results are consistent with a Bayesian approach to understanding context effects, recently put forth by Snyder, Schwiedrzik, Vitela, and Mellon (2015). From the perspective of Bayesian models, the inclusion of preceding context serves to further minimize error that stems from either a fading sensory experience or from systematic biases introduced from temporally more long-term priors. Importantly, the Bayesian framework put forth by Snyder et al. (2015) provides a functional and neurally plausible account of both assimilative and contrastive effect in perception. Assimilative effects are argued to occur when the likelihood distributions between sensory and long-term experience are aligned with one another, while contrastive effects are argued to occur when the likelihood distributions between sensory and long-term experience are misaligned. This framework is consistent with the current data, in which flat or sharp preceding intonation context can be viewed as producing a misalignment in the likelihood distributions between sensory and long-term experience, giving rise to our contrastive context effects. The model by Snyder et al. (2015) also offers insights into the neurophysiological mechanisms that may underlie the current effects. Specifically, contrastive context effects are thought to arise from changes in sensory cortex, while assimilative effects are thought to be supported by higher-level, frontoparietal areas.

Given that the percept associated with a given stimulus was dependent on the immediately preceding context, the current results are also consistent with adaptation level theory as proposed by

Helson (1964). In adaptation level theory, the percept associated with a given stimulus depends on the immediately preceding context. For example, the perceived weight of a given object will be subject to the weight of objects most recently experienced. According to adaptation level theory, the contrastive effects observed in the present set of experiments arise from residual traces of the recently experienced melodic line that are presumably maintained in auditory working memory. Together, these residual traces form an adaptation level, combined with some influence from relevant long-term categories, against which subsequent isolated notes are compared. Overall, then, the data here are consistent with an adaptation level account of anchoring that emerges from the contribution of recent experience against long-term established categories. This is reminiscent of plasticity-inducing adaptation effects described in speech (e.g., Lotto & Holt, 2006), as well as those found in other domains, such as vision (Kohn & Kohn, 2007) and weight estimation (von Wright & Mikkonen, 1964), suggesting that such context effects in perception reflect attributes of general sensory processing (Snyder, Schwiedrzik, Vitela, & Mellon, 2015).

Given the interpretation of our results under adaptation level theory, it is possible that individual differences in auditory working memory may determine the strength of contextual effects in perception. Some of our prior work has established relationships between auditory working memory and absolute pitch representations (Van Hedger, Heald, Koch, & Nusbaum, 2015; Van Hedger, Heald, & Nusbaum, in press), though the specific question of how working memory may modulate the influence of context on perceptual judgments remains unanswered. If auditory working memory is the mechanism through which context is maintained (to form the adaptation level), then an individual with *lower* auditory working memory may demonstrate *less* sensitivity to context effects. To better understand the mechanisms that underlie adaptation, future work should therefore examine how the strength of contextual effects may relate to individual differences in auditory working memory. Examining this relationship in both music and speech may provide a more general understanding of how recent contextual information is integrated with perceptual representations in audition.

The results in this study also help situate how and when context may be used more broadly in audition. In speech, context effects are traditionally thought to be limited to target sounds that are ambiguous in nature. For example, a phoneme that falls half way between /p/ and /b/ will be heard more as /b/ when presented in the context of a sentence such as, “The gambler took the (bet, pet)”, and more as /p/ when it is presented in the context of a sentence such as, “The man wanted company so he took the (bet, pet)” (Borsky, Tuller, & Shapiro, 1998; Connine & Clifton, 1987). This is even true of low-level context effects found in speech perception that occur outside of semantic sources of information (Lotto & Kluender, 1998; Lotto, Kluender, & Holt, 1997). In the current study, however, target notes are in a definitional sense *unambiguous*, yet we observed clear context effects that influenced both intonation judgments and note category accuracy. As such, these results indicate that that context effects in perception are not simply determined by the ambiguity of a given stimulus. One explanation for why this might be the case is that in music, an altered intonation context *that preserves relative pitch relationships* may encourage listeners to generally attend to relative pitch cues in order to achieve perceptual coherence. Ecologically speaking, this would allow AP possessors to adapt to music that is not canonically tuned (e.g., orchestras that tune slightly higher or lower than the standard A4 = 440Hz), and it also may allow for adaptation of music that slowly drifts from an initial tuning over time, such as the drift in intonation that sometimes occurs in unaccompanied vocal performances (cf. Ward, 1963).

Taken together, our results provide strong empirical support for the notion that AP possessors' internal category standards are not firmly fixed (Siegel, 1972), as they can be influenced by the intonation of recent listening experience. The present results, however, cannot speak to the limits of this kind of flexibility within AP, as our intonation deviations were relatively small. Our results also cannot speak to the length of time intonation context may influence note representations, as our test stimuli were either played simultaneously with the context (Experiment 1) or within a quarter-second after the context ended (Experiment 2). Nevertheless, the present results clearly demonstrate that AP representations are flexibly and rapidly updated based on recent listening experience. Not only do these results further our understanding of the nature of AP categories, but they also help elucidate the principles that guide perceptual category plasticity more generally.

Acknowledgements

The authors would like to acknowledge Alex Huang and Salina Wu for their assistance in stimuli creation. This research was supported by the Multidisciplinary University Research Initiatives (MURI) Program of the Office of Naval Research though grant, DOD/ONR N00014-13-1-0205.The stimuli and data can be accessed through the Open Science Framework (https://osf.io/zqgy4/).

References

Asendorpf, J. B., Conner, M., Fruyt, F. D. E., Houwer, J. A. N. D. E., Denissen, J. J. A., Fiedler, K., … Vanaken, M. A. G. (2013). Recommendations for Increasing Replicability in Psychology. *European Journal of Personality*, *27*, 108–119.

Athos, E. A., Levinson, B., Kistler, A., Zemansky, J., Bostrom, A., Freimer, N., & Gitschier, J. (2007). Dichotomy and perceptual distortions in absolute pitch ability. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(37), 14795–14800. https://doi.org/10.1073/pnas.0703868104

Bahr, N., Christensen, C. A., & Bahr, M. (2005). Diversity of accuracy profiles for absolute pitch recognition. *Psychology of Music*, *33*(1), 58–93. https://doi.org/10.1177/0305735605048014

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious Mixed Models. *arXiv*, *1506.04967*, 1–27. https://doi.org/arXiv:1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*(1), 51. https://doi.org/10.18637/jss.v067.i01

Benguerel, A.-P., & Westdal, C. (1991). Absolute Pitch and the Perception of Sequential Musical Intervals. *Music Perception*, *9*(1), 105–120. https://doi.org/10.2307/40286161

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). “How to milk a coat:” the effects of semantic and acoustic information on phoneme categorization. *The Journal of the Acoustical Society of America*, *103*(5 Pt 1), 2670–2676. https://doi.org/10.1121/1.422787

Christensen, R. H. B. (2011). Analysis of ordinal data with cumulative link models—estimation with the R-package “ordinal”. Retrieved June 1, 2017, from cran.r-project.org/web/packages/ordinal/vignettes/clm_intro.pdf

Connine, C. M., & Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology. Human Perception and Performance*, *13*(2), 291–299. https://doi.org/10.1037/h0090425

Crozier, J. B. (1997). Absolute pitch: Practice makes perfect, the earlier the better. *Psychology of Music*, *25*(2), 110–119. https://doi.org/0803973233

Deutsch, D. (2013). Absolute Pitch. In D. Deutsch (Ed.), *The Psychology of Music* (3rd ed., pp. 141–182). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-381460-9.00005-5

Dohn, A., Garza-Villarreal, E. A., Riisgaard Ribe, L., Wallentin, M., & Vuust, P. (2014). Musical activity tunes up absolute pitch ability. *Music Perception*, *31*(4), 359–371. https://doi.org/10.1525/rep.2008.104.1.92.This

Dooley, K., & Deutsch, D. (2011). Absolute pitch correlates with high performance on interval naming tasks. *The Journal of the Acoustical Society of America*, *130*(6), 4097. https://doi.org/10.1121/1.3652861

Gur-Ozmen, S., Nirmalananthan, N., & Von Oertzen, T. J. (2013). Change of pitch due to carbamazepine and oxcarbazepine independently. *Seizure*, *22*(2), 162–163. https://doi.org/10.1016/j.seizure.2012.11.010

Heald, S. L. M., Van Hedger, S. C., & Nusbaum, H. C. (2017a). Perceptual plasticity for auditory object recognition. *Frontiers in Psychology*, *8*(May), 1–16. https://doi.org/10.3389/fpsyg.2017.00781

Heald, S. L. M., Van Hedger, S. C., & Nusbaum, H. C. (2017b). *Understanding Sound: Auditory Skill Acquisition*. *Psychology of Learning and Motivation* (Vol. 67). Elsevier Ltd. https://doi.org/10.1016/bs.plm.2017.03.003

Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2013). Absolute Pitch May Not Be So Absolute. *Psychological Science*, *24*(8), 1496–1502. https://doi.org/10.1177/0956797612473310

Helson, H. (1964). Current trends and issues in adaptation-level theory. *American Psychologist*, *19*(1), 26–38. https://doi.org/http://dx.doi.org/10.1037/h0040013

Holt, L. L., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *The Journal of the Acoustical Society of America*, *108*(2), 710–722.

Kobayashi, T., Nisijima, K., Ehara, Y., Otsuka, K., & Kato, S. (2001). Pitch perception shift: A rare side-effect of carbamazepine. *Psychiatry and Clinical Neurosciences*, *55*(4), 415–417. https://doi.org/10.1046/j.1440-1819.2001.00883.x

Kohn, A., & Kohn, A. (2007). Visual Adaptation: Physiology, Mechanisms, and Functional Benefits. *Journal of Neurophysiology*, *10461*, 3155–3164. https://doi.org/10.1152/jn.00086.2007.

Konno, S., Yamazaki, E., Kudoh, M., Abe, T., & Tohgi, H. (2003). Half pitch lower sound perception caused by carbamazepine. *Internal Medicine*, *42*(9), 880–883. https://doi.org/10.2169/naika.94.759

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*(3), 346–374. https://doi.org/10.1016/0010-0285(79)90016-1

Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences*, *9*(1), 26–33. https://doi.org/10.1016/j.tics.2004.11.007

Levitin, D. J., & Zatorre, R. J. (2003). On the Nature of Early Music Training and Absolute Pitch: A Reply to Brown, Sachs, Cammuso, and Folstein. *Music Perception*, *21*(1), 105–110. https://doi.org/10.1525/mp.2003.21.1.105

Lockhead, G., & Byrd, R. (1981). Practically perfect pitch. *The Journal of the Acoustical Society of America*, *70*(2), 387–389.

Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: a commentary on Fowler (2006). *Perception & Psychophysics*, *68*(2), 178–183. https://doi.org/10.3758/BF03193667

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, *60*(4), 602–619. https://doi.org/10.3758/BF03206049

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). *The Journal of the Acoustical Society of America*, *102*(2 Pt 1), 1134–40. https://doi.org/10.1121/1.419865

Miyazaki, K. (1990). The Speed of Musical Pitch Identification Possessors by Absolute-Pitch Possessors. *Music Perception*, *8*(2), 177–188.

Miyazaki, K. (1995). Perception of relative pitch with different references: some absolute-pitch listeners can’t tell musical interval names. *Perception & Psychophysics*, *57*(7), 962–970. https://doi.org/10.3758/BF03205455

Ross, D. a., Olson, I. R., & Gore, J. C. (2003). Absolute Pitch Does Not Depend on Early Musical Training. *Annals of the New York Academy of Sciences*, *999*, 522–526. https://doi.org/10.1196/annals.1284.065

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Sawusch, J. R., & Nusbaum, H. C. (1979). Contextual effects in vowel perception I: Anchor-induced contrast effects. *Perception & Psychophysics*, *25*(4), 292–302. https://doi.org/10.3758/BF03198808

Sawusch, J. R., Nusbaum, H. C., & Schwab, E. C. (1980). Contextual effects in vowel perception II: Evidence for two processing mechanisms. *Perception & Psychophysics*, *27*(5), 421–434. https://doi.org/10.3758/BF03204460

Siegel, J. A. (1972). The nature of absolute pitch. In E. Gordon (Ed.), *Experimental research in the psychology of music: VIII. Studies in the psychology of music* (pp. 65–89). Iowa City: Iowa University Press.

Snyder, J. S., Schwiedrzik, C. M., Vitela, A. D., & Melloni, L. (2015). How previous experience shapes perception in different sensory modalities. *Frontiers in Human Neuroscience*, *9*(October), 1–8. https://doi.org/10.3389/fnhum.2015.00594

Van Hedger, S. C., Heald, S. L. M., Koch, R., & Nusbaum, H. C. (2015). Auditory working memory predicts individual differences in absolute pitch learning. *Cognition*, *140*, 95–110. https://doi.org/10.1016/j.cognition.2015.03.012

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (in press). Long-term pitch memory for music recordings is related to auditory working memory precision. *The Quarterly Journal of Experimental Psychology*. https://doi.org/10.1080/17470218.2017.1307427

Vanzella, P., & Schellenberg, E. G. (2010). Absolute pitch: Effects of timbre on note-naming ability. *PLoS ONE*, *5*(11). https://doi.org/10.1371/journal.pone.0015449

Vernon, P. E. (1977). Absolute pitch: a case study. *British Journal of Psychology*, *68*(4), 485–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/588883

von Wright, J. M., & Mikkonen, V. (1964). Changes in repeated reproduction of weight as a function of adaptation level. *Scandinavian Journal of Psychology*, *5*(1), 239–248.

Ward, W. D. (1963). Absolute pitch. Part II. *Sound*, *2*(4), 33–41.

Ward, W. D., & Burns, E. M. (1982). Absolute pitch. In D. Deutsch (Ed.), *The Psychology of Music* (1st ed., pp. 431–451). San Diego, CA: Academic Press.

Watkins, A. J. (2005). Perceptual compensation for effects of echo and of reverberation on speech identification. *The Journal of the Acoustical Society of America*, *118*(1), 249–262. https://doi.org/10.1121/1.1923369

Watkins, A. J., & Makin, S. J. (1994). Perceptual compensation for speaker differences and for spectral-envelope distortion. *The Journal of the Acoustical Society of America*, *96*(3), 1263–1282. https://doi.org/10.1121/1.410275

Wilson, S. J., Lusher, D., Martin, C. L., Rayner, G., & McLachlan, N. (2012). Intersecting factors lead to absolute pitch acquisition that is maintained in a “Fixed do” Environment. *Music Perception*, *29*(3), 285–296.

Wynn, V. T. (1973). Absolute pitch in humans, its variations and possible connections with other known rhythmic phenomena. *Progress in Neurobiology*, *1*(2), 111–149. https://doi.org/10.1016/0301-0082(73)90018-X

Zatorre, R. J. (2003). Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience*, *6*(7), 692–695. https://doi.org/10.1038/nn1085

Tables and Table Captions

Term	Log Odds	Standard Error	Odds Ratio	**95% C.I. of Odds Ratio**
Flat Lead	1.157	0.188	3.181	**(2.201, 4.598)***
Sharp Lead	-0.924	0.232	0.397	**(0.252, 0.626)***
Flat Secondary	-1.252	0.167	0.286	**(0.206, 0.396)***
Sharp Secondary	1.598	0.168	4.941	**(3.553, 6.871)***

Table 1: Fixed effects from the cumulative link mixed model. Significance (denoted with *) was assessed based on whether the 95% confidence interval of the odds ratio (calculated from the log odds) included 1. For the main effects, an odds ratio less than 1 denotes a lower odds of responding sharp, whereas an odds ratio higher than 1 denotes a greater odds of responding sharp. The model adhered to the equation of “Intonation Judgment ~ Flat Lead + Sharp Lead + Flat Secondary + Sharp Secondary + (1 + Flat Lead + Sharp Lead | Participant) + (1 | Melody)” using the “clmm” model from the {ordinal} package in R.

Heard				
Term	*Log Odds*	*Standard Error*	*Odds Ratio*	***95% C.I. of Odds Ratio***
Flat Context	2.335	0.473	10.328	**(4.084, 26.118)***
Sharp Context	-1.445	0.415	0.236	**(0.105, 0.532)***
Flat Note	-2.765	0.475	0.063	**(0.025, 0.160)***
Sharp Note	2.647	0.464	14.106	**(5.682, 35.020)***
Near Transfer				
Term	*Log Odds*	*Standard Error*	*Odds Ratio*	***95% C.I. of Odds Ratio***
Flat Context	1.008	0.452	2.739	**(1.129, 6.643)***
Sharp Context	-1.176	0.446	0.308	**(0.129, 0.739)***
Flat Note	-1.632	0.383	0.195	**(0.092, 0.414)***
Sharp Note	1.929	0.413	6.882	**(3.065, 15.455)***
Far Transfer				
Term	*Log Odds*	*Standard Error*	*Odds Ratio*	***95% C.I. of Odds Ratio***
Flat Context	0.154	0.273	1.167	**(0.683, 1.992)**
Sharp Context	0.106	0.316	1.112	**(0.599, 2.065)**
Flat Note	-0.633	0.256	0.531	**(0.322, 0.876)***
Sharp Note	1.140	0.283	3.126	**(1.797, 5.439)***

Table 2: Fixed effects from the Heard, Near Transfer, and Far Transfer mixed models. Significance (denoted with *) was assessed based on whether the 95% confidence interval of the odds ratio (calculated from the log odds) included 1. For the main effects, an odds ratio less than 1 denotes a lower odds of responding sharp, whereas an odds ratio higher than 1 denotes a greater odds of responding sharp. All models adhered to the general equation of "Intonation Judgment ~ Flat Context + Sharp Context +_Flat Note + Sharp Note + (1 + Flat Context + Sharp Context | Participant) + (1 | Melody)" using the "clmm" model from the {ordinal} package in R.

Heard

Sharp Notes

Term	*Estimate*	*Standard Error*	*95% C.I. of Estimate*
Intercept	0.16723	0.07853	(0.013, 0.323)*
Sharp Context	-0.09010	0.10867	(-0.304, 0.127)
Flat Context	0.51264	0.10966	(0.297, 0.727)*

Flat Notes

Term	*Estimate*	*Standard Error*	*95% C.I. of Estimate*
Intercept	-0.1461	0.0877	(-0.318, 0.023)
Sharp Context	-0.2946	0.1109	(-0.512, -0.077)*
Flat Context	0.1567	0.1084	(-0.054, 0.366)

Near Transfer

Sharp Notes

Term	*Estimate*	*Standard Error*	*95% C.I. of Estimate*
Intercept	0.07209	0.07887	(-0.080, 0.242)
Sharp Context	0.12366	0.10556	(-0.084, 0.327)
Flat Context	0.63223	0.10338	(0.430, 0.834)*

Flat Notes

Term	*Estimate*	*Standard Error*	*95% C.I. of Estimate*
Intercept	-0.05592	0.06350	(-0.182, 0.068)
Sharp Context	-0.31125	0.08033	(-0.464, -0.160)*
Flat Context	0.03914	0.08033	(-0.120, 0.198)

Far Transfer

Sharp Notes

Term	*Estimate*	*Standard Error*	*95% C.I. of Estimate*
Intercept	0.17647	0.06867	(0.043, 0.311)*
Sharp Context	0.04470	0.08076	(-0.113, 0.202)
Flat Context	0.54902	0.08033	(0.391, 0.707)*

Flat Notes

Term	*Estimate*	*Standard Error*	*95% C.I. of Estimate*
Intercept	-0.05882	0.05641	(-0.169, 0.053)
Sharp Context	-0.43137	0.06707	(-0.563, -0.300)*
Flat Context	0.01961	0.06707	(-0.112, 0.152)

Table 3: Fixed effects from the sharp and flat Heard, Near Transfer, and Far Transfer models. In all cases, conditions in which intonation was maximally contrastive (flat context/sharp note and sharp context/flat note) resulted in a significantly increased predicted number of errors in the expected directions. Significance (denoted with *) was assessed based on whether the 95% confidence interval of estimate included zero. All models adhered to the general equation of "Semitones (Mis)classification~ Flat Context + Sharp Context + (1 | Participant)" using the {lme4} package in R.

Figures and Figure Captions

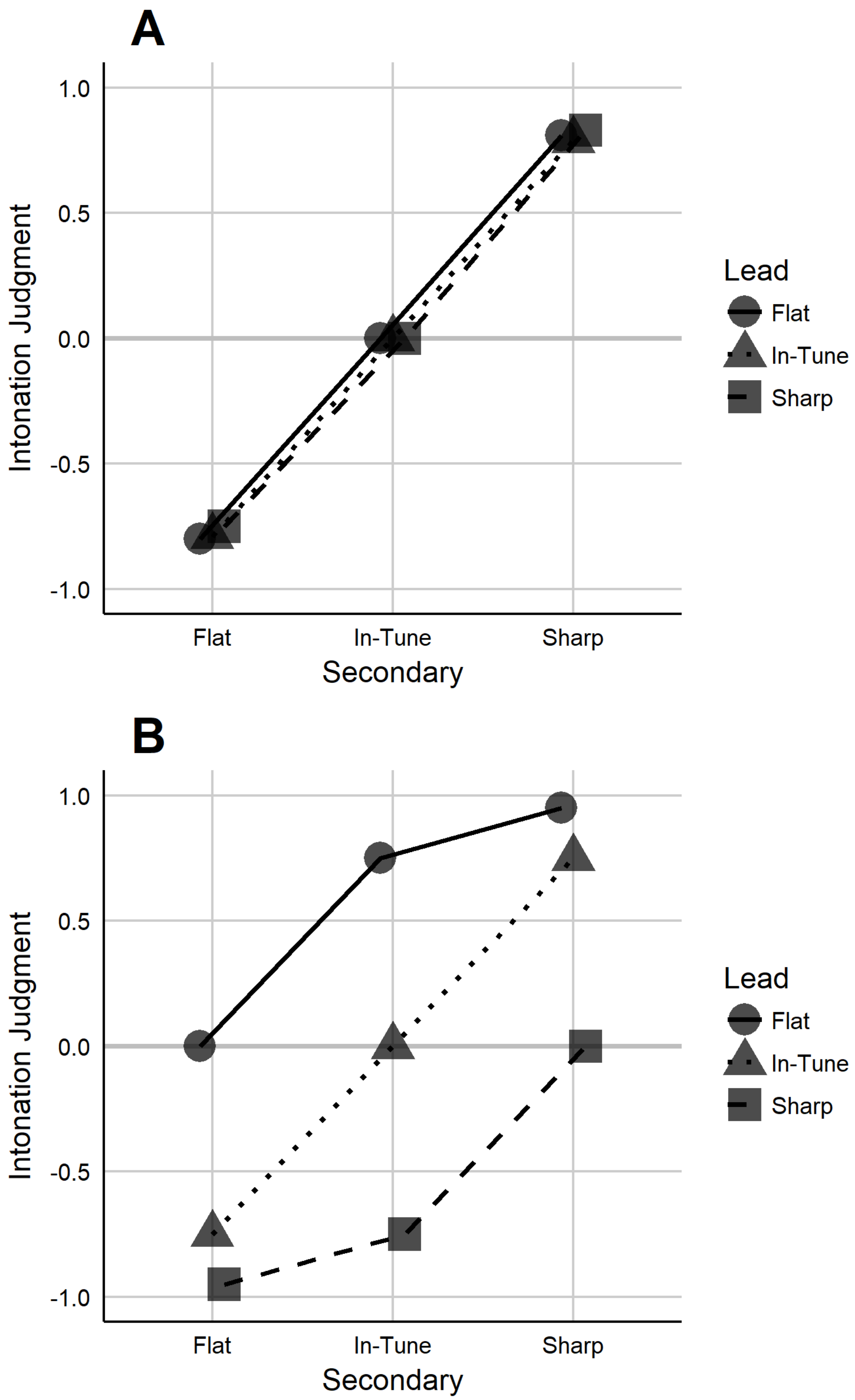

Figure 1: Hypothesized intonation judgments under a context-independent (a) and context-dependent (b) framework. In this schematic, values of zero represent judgments that the note is “in-tune,” whereas values higher and lower than zero represent judgments that the note is “sharp” or “flat,” respectively.

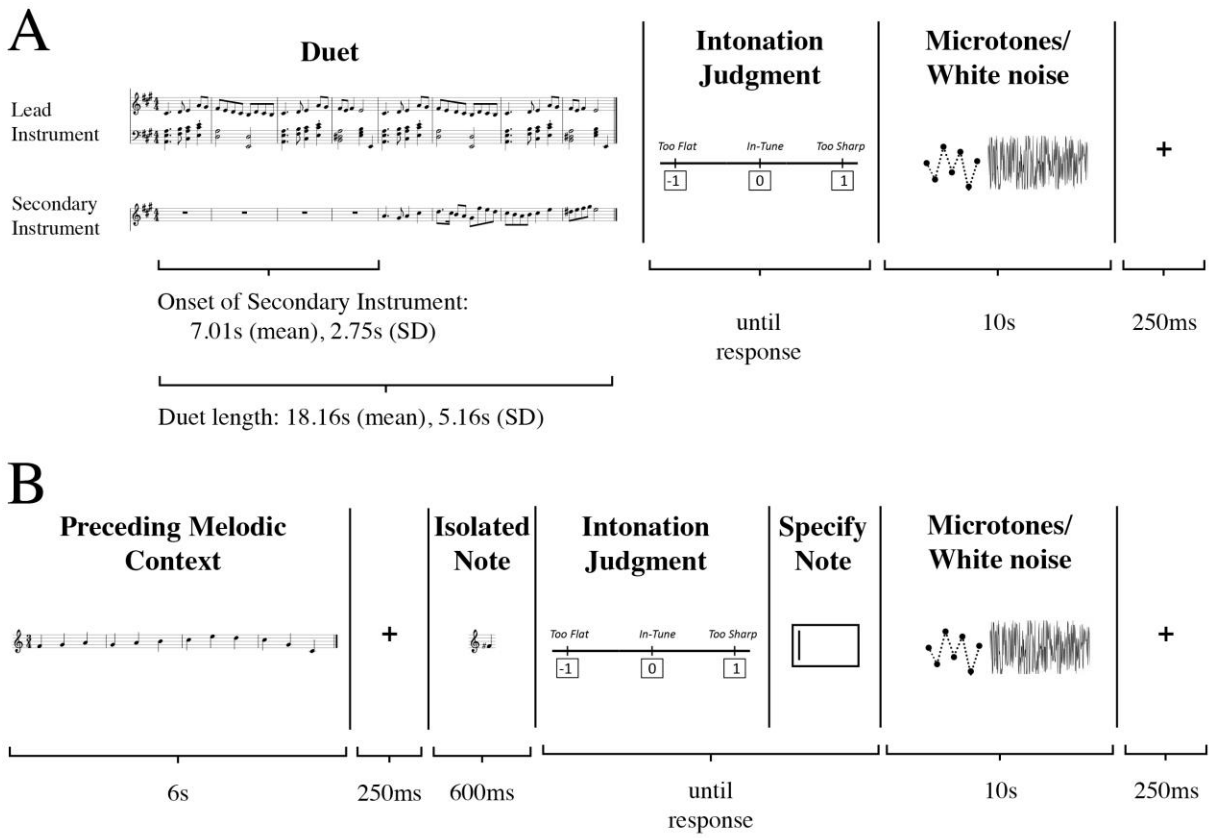

Figure 2: Sketch of the experimental design for Experiment 1 (A) and Experiment 2 (B).

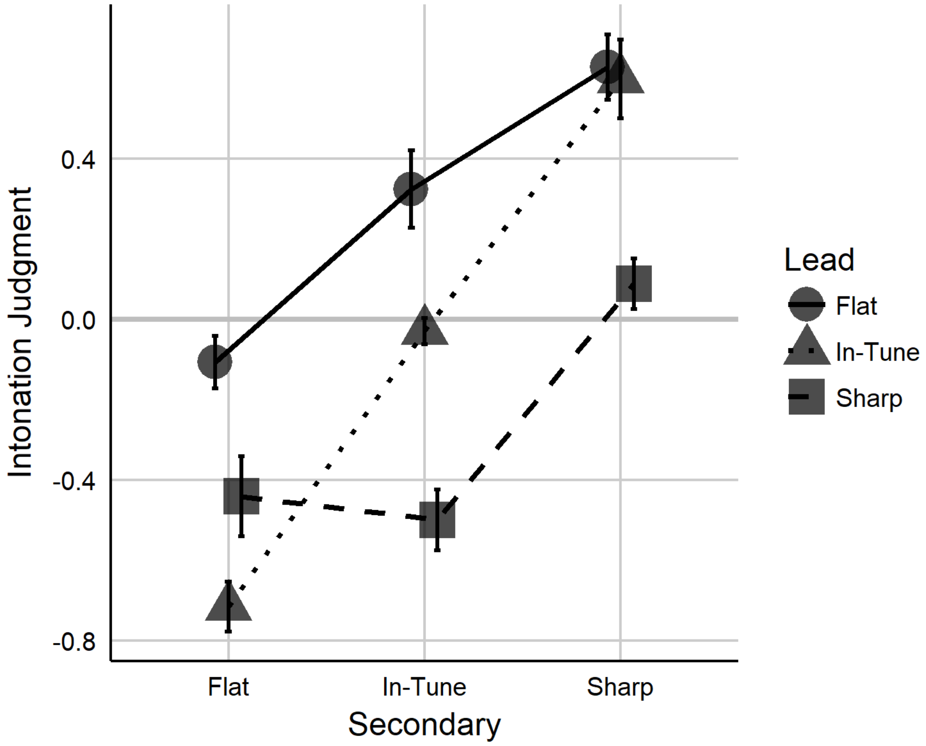

Figure 3: Mean intonation judgments plotted as a function of both the lead instrument's intonation (separate lines) and the secondary instrument's intonation (x-axis). While participants responded on each trial in a discrete manner (with either -1, 0, or 1), points on the graph represent averaged values across condition and participant. Given the response scale, mean responses greater than zero thus represent a tendency to respond "sharp," while mean responses less than zero represent a tendency to respond "flat." Error bars represent ± 1 SEM.

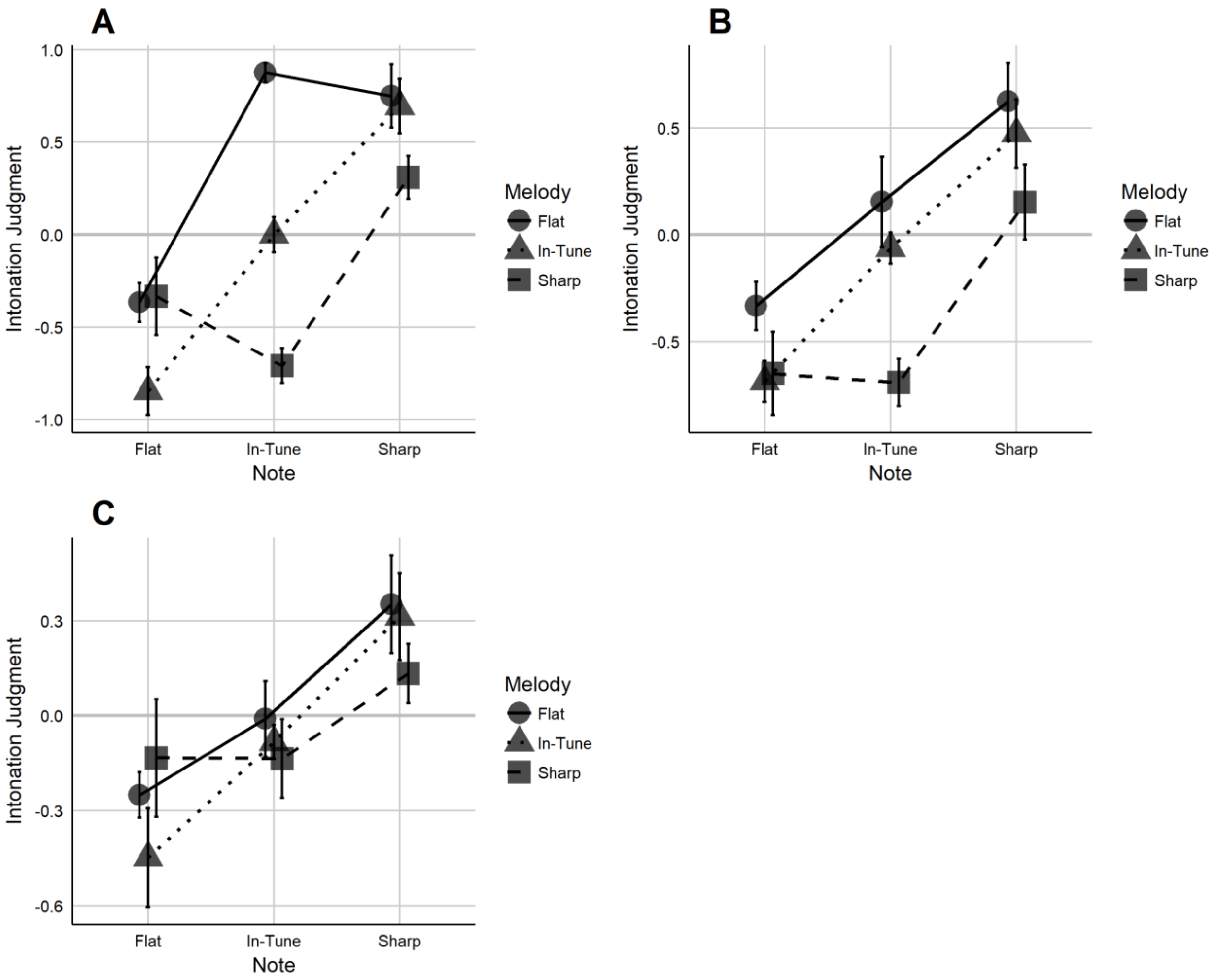

Figure 4: Mean intonation responses as a function of both the intonation of the preceding melody (separate lines) and the to-be-judged note (x-axis) for notes in the Heard Model (A), notes in the Near Transfer Model (B), and notes in the Far Transfer Model (C). Points represent mean responses for each condition. Given the response scale, mean responses greater than zero represent a tendency to respond "sharp," while mean responses less than zero represent a tendency to respond "flat." Error bars represent ± 1 SEM.

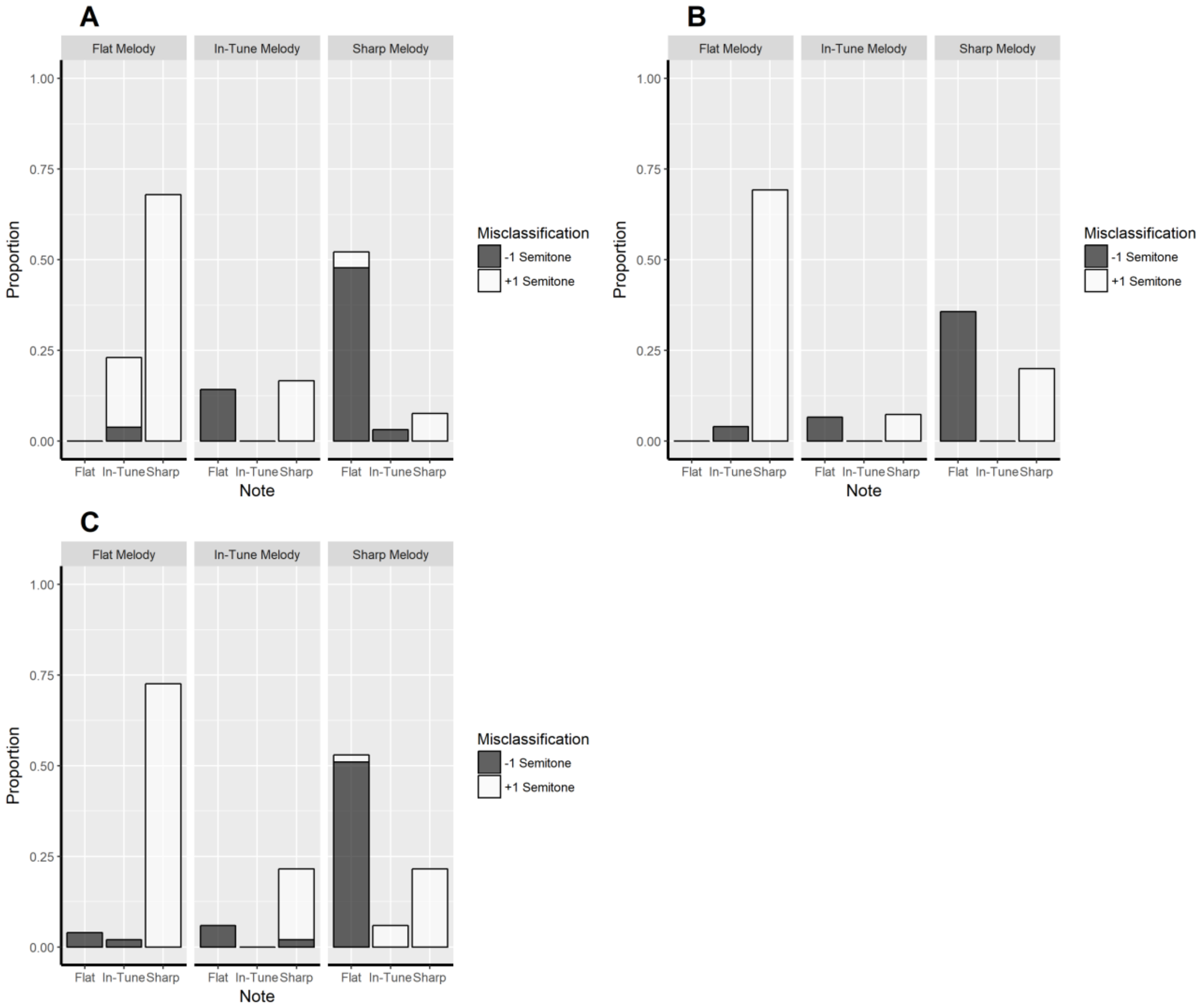

Figure 5: Semitone errors plotted as a function of trial type for heard (A), near transfer (B), and far transfer (C) notes.