

UC San Diego

UC San Diego Previously Published Works

Title

No Heightened Musical Pitch Weighting For Tone Language Speakers in Early Childhood

Permalink

<https://escholarship.org/uc/item/1vv0s7g7>

Journal

Music Perception, 40(3)

ISSN

0730-7829

Authors

Creel, Sarah C
Mizrahi, Reina
Escobedo, Alicia G
[et al.](#)

Publication Date

2023-02-01

DOI

10.1525/mp.2023.40.3.193

Peer reviewed

No Heightened Musical Pitch Weighting For Tone Language Speakers in Early Childhood

Sarah C. Creel, Reina Mizrahi, & Alicia G. Escobedo

University of California San Diego

Li Zhao

Hangzhou Normal University, Hangzhou, China

Gail D. Heyman

University of California San Diego

Abstract

Numerous studies suggest that speakers of some tone languages show advantages in musical pitch processing compared to non-tone language speakers. A recent study in adults (Jasmin et al., 2021) suggests that in addition to heightened pitch sensitivity, tone language speakers weight pitch information more strongly than other auditory cues (amplitude, duration) in both linguistic and nonlinguistic settings compared to non-tone language speakers. The current study asks whether pitch upweighting is evident in early childhood. To test this, two groups of 3- to 5-year-old children—tone-language speakers ($n = 48$), a group previously shown to have a perceptual advantage in musical pitch tasks (Creel et al., 2018), and non-tone-language speakers ($n = 48$)—took part in a musical “word learning” task. Children associated two cartoon characters with two brief musical phrases differing in both musical instrument and contour. If tone language speakers weight pitch more strongly, cue conflict trials should show stronger pitch responding than for non-tone speakers. In contrast to both adult speakers’ stronger pitch weighting and child and adult pitch perception advantages, tone-language-speaking children did not show greater weighting of pitch information than non-tone-language speaking children. This suggests a slow developmental course for pitch reweighting, contrasting with apparent early emergence of pitch sensitivity.

Received August 5, 2022, accepted November 10, 2022.

Key words: pitch, language and music, child development, auditory perception, perception

A major question in cognition is the degree to which cognitive domains overlap. Is the brain a collection of distinct tools specialized for particular tasks, or a set of related abilities that share neural resources? Evidence for overlap comes from *transfer effects*, where experience in one domain (such as language) influences performance in a different domain (such as music). Less well understood is what exactly is transferred, and the developmental time scale over which transfer effects arise. Here we consider transfer effects from *tone languages*, which use pitch as an integral feature, to musical pitch processing. In the current study, we ask whether tone language experience alters the weighting of pitch cues compared to other auditory cues (in our case instrument sound quality differences or *timbre*) at an early point in development.

Tone languages—for example, Mandarin, Cantonese, Thai, Yoruba, and Akan—are languages in which changing the pitch pattern of a word results in another word with a different meaning. For example, in Mandarin, *ma* spoken at a high, level pitch means “mother,” while *ma* with a dipping pitch (fall-rise pattern) means “horse.” A recurring finding (Bidelman et al., 2013; Bradley, 2016; Hove et al., 2010; Hutka et al., 2015; Liu et al., 2021; Pfordresher & Brown, 2009; though see Bent et al., 2006) is that tone-language speakers excel in musical pitch processing tasks relative to non-tone-language speakers, in particular melodic change detection.

One recent study suggests that pitch processing is not only heightened, but that it also comes to dominate other auditory cues in tone language speakers. Jasmin et al. (2021) examined *weighting* of pitch relative to duration information. They presented English words or musical note sequences where pitch and duration sometimes suggested different patterns. For example, a sequence of notes might contain the pitches [A B A B A B] suggests 3 pairs of 2 beats, but the

durations [200ms 50ms 50ms 200ms 50ms 50ms], suggesting 2 pairs of 3 beats. They found that adult speakers of the tone language Mandarin used pitch information more strongly than speakers of non-tone languages English and Spanish. They also found that Mandarin listeners had more difficulty than non-tone speakers in *ignoring* pitch variation when it is irrelevant to the task (loudness judgment). These findings suggest that speaking a tone language leads to greater perceptual dominance of pitch over other information (duration, loudness).

Development of Language-to-Music Transfer

Developmental studies are fewer but suggest that tone-language-speaking children as young as 4-6 years show facilitated musical pitch processing (Creel et al., 2018; Deroche et al., 2019; though see Peretz et al., 2013, for a different interpretation). Deroche et al. (2019) tested children ages 6-19 years and found that tone language speakers outperformed non-tone speakers in identifying the direction of pitch sweeps and discriminating direction of pitch sweeps. Creel et al. (2018) tested 3-5-year-old Mandarin speakers and non-tone speakers on a same-different discrimination test (data replotted in Figure 1). Creel et al. (2018) found that non-tone speaking children (leftmost panel) were much less accurate on pitch contour change detection than they were on control trials testing musical-instrument (*timbre*) change detection. By contrast, child tone language speakers (right three panels) showed pitch contour change detection that was just as good as their timbre change detection. This suggests that facilitation is particular to pitch information, not just heightened task performance. These studies together suggest that language properties may affect nonlinguistic processing early in life, age 4 years or earlier.

Insert Figure 1 about here

One might think that greater accuracy of pitch perception in tone speakers would naturally translate to stronger cue weighting of pitch, and this is consistent with Jasmin et al.'s (2021) findings in adults. It is not clear whether this is also true of children, as no studies to date have examined tone language effects on pitch cue weighting in children. One developmental study suggests that children do not weight pitch as strongly as timbre, at least in an English-speaking country. Creel (2014; [see related work by Creel, 2016](#)) conducted a developmental study of pitch cue weighting in the United States, where the non-tone language English is the dominant language. In a learning task, children associated one cartoon character with one brief melody (example: *rising pitch-trumpet*), and a different cartoon character with another brief melody differing in both timbre and contour (*falling pitch-vibraphone*). After several learning trials, children were tested: they saw both characters at once, heard a melody, and were asked to point to the character that went with that melody. On some test trials, pitch contour and timbre cues conflicted (example: *rising pitch-vibraphone*). Children appeared mostly insensitive to contour, responding strongly to timbre differences. Interestingly, adult controls *did* show sensitivity to pitch contour, and when timbres were similar, adults responded almost completely to pitch contour. A parallel experiment pitting pitch contour against absolute pitch *height* (*high pitch-falling* vs. *low pitch-rising*) found similar developmental shifts, with children showing more sensitivity to pitch height than to contour but adults showing more balanced sensitivity. Adults' greater use of pitch contour is consistent with long-term developmental increases in pitch sensitivity (Fancourt et al., 2013; Stalinski & Schellenberg, 2008).¹

¹ It may be relevant that both of these studies were conducted in English-speaking countries.

Thus, both language experience (Jasmin et al., 2021) as well as development and/or concomitant music (or auditory) experience (Creel, 2014) may influence weighting of pitch relative to other cues. It remains unknown whether developmental upweighting of pitch contour is hastened by tone language input since no developmental study of tone language effects on pitch processing has assessed combinations of cues.

The Current Study

We tested whether nonspeech cue weighting differences appear in childhood for tone vs. non-tone language speakers. A learning task, modeled closely on Creel (2014), allowed assessment of cue weighting (Figure 2). Children learned that two melodies were the “favorite songs” of two cartoon characters. Each melody was distinguished by both pitch contour and timbre (example: *falling pitch-vibraphone* vs. *rising pitch-trumpet*). At test, children saw both characters and heard one melody at a time. *Combined-cue* trials presented one of the original melodies (example: *rising pitch-trumpet*), assessing learning of the original mapping. On *cue-conflict* trials, children heard the timbre of one and the pitch contour of the other (example: *rising pitch-vibraphone*; Figure 2, lower right). Conflict trials assess the relative weights assigned to cues. To replicate Creel (2014) and also to span a range of timbre cue strengths, both a similar-timbre pairing and a distinct-timbre pairing were used.

Insert Figure 2 about here

If pitch has a higher weighting for tone language-speaking children, then this should be reflected particularly in the cue-conflict responses. As in Creel (2014), we expect non-tone speaking children to respond predominantly to timbre, essentially inverting their contour-recognition accuracy relative to combined-cue trials. However, tone-speaking children should respond by using pitch contour over timbre (no inversion of responses) or some intermediate value, such as performance falling to chance (indicating approximately equal cue weights for timbre and pitch contour). If pitch does *not* have a higher weighting, then the two groups should look similar. An additional expectation is that tone language speakers' better access to pitch cues should yield higher performance overall on combined cue trials.

Method

Ethics Statement

In the United States, children were run under an IRB protocol approved by the UCSD Human Research Protections Program, which obtained parental informed consent and child assent. In China, the study was approved by the Hangzhou Normal University Scientific Research Ethics Committee, and children were run with their guardian's informed consent and child assent.

Participants

The 96 participants included 48 tone-language speakers in China (24 females) and 48 non-tone language speakers in the United States (20 females), with 16 3-year-olds, 16 4-year-olds, and 16 5-year-olds in each country. Sample sizes were determined in advance based on Creel et al. (2018), including 16 children in each of three age groups. Seven additional children tested in the US were excluded prior to analysis due to the experimental program crashing (3), unwillingness

to continue (2), unwillingness to point (1), exposure to a pitch-accent language (Japanese) which has been linked to enhanced pitch perception (1; see Henthorn & Deutsch, 2007).

Stimuli

We used stimuli similar to Creel (2014). Creel (2014) had omitted the second onset in the falling melody in an ultimately unsuccessful attempt to make the two melodies more distinguishable.

We amended those stimuli slightly to equate number of notes across the two melodies (Figure 3).

All melodies occurred in both similar-timbre pairs (bassoon vs. saxophone) and distinct-timbre pairs (muted trumpet vs. vibraphone).

Insert Figure 3 about here

Procedure

Each child completed two cycles of audiovisual association learning and testing (learn, test, learn, test) as outlined in Table 1. At the beginning of the study, each child heard the following instructions (in English in the United States, in Mandarin in China): *You're going to meet some creatures. Each creature has a FAVORITE SONG. Every time you see that creature, you're going to hear his FAVORITE SONG play. Are you ready to find out their favorite songs? We checked instruction accuracy by backtranslating Mandarin instructions into English.*

On each learning trial, one cartoon character moved to screen center, paused, and its “favorite song” played. Then it moved offscreen, and the next trial began. Each melody+cartoon pairing

(rising+Timbre1 or falling+Timbre2) occurred 8 times during a learning phase (total 16 learning trials), with the two characters alternating in random order. Children were encouraged to watch the display but did not have to make a response. Periodically, distractor trials appeared (moving pictures of animals and engaging non-musical sounds) to maintain interest (see Table 1).

Insert Table 1 about here

Next was the test phase (24 trials). First, children heard instructions: *Now you get to pick whose favorite song you hear. When you hear a creature's favorite song, just point to that creature, okay?* Each test trial presented both cartoons on the computer screen (one on left, one on right; position counterbalanced across trials) as children heard one melody. The first 8 test trials (*combined-cue* trials only) presented one of the two exact melodies heard during learning (4 trials per melody). This allowed us to verify that stimuli had been learned successfully prior to the presentation of cue-conflict trials that might be confusing. The next 16 trials were split between *combined-cue trials* and *cue-conflict trials*.

Each child completed *two* sequences of learning + testing trials. In one sequence, children heard melodies with two similar timbres (bassoon-saxophone; see Iverson & Krumhansl, 1993) and saw two characters; in the other sequence, they heard melodies with distinct timbres (muted trumpet-vibraphone) and saw two other characters. Timbre pair and cartoon characters were counterbalanced across melodies and order of occurrence (first vs. second learning-test sequence). A total of 16 lists counterbalanced the order of cartoons, order of timbres, and cartoon-to-timbre mappings. Each list was run once within each age x language group.

Equipment and Software

The design was implemented in PsychoPy experimental presentation software (Peirce et al., 2019). KidzGear child-sized headphones were used to present sounds at a comfortable listening level. In the United States, children were tested in a quiet area in their preschool or daycare facility, on one of two Mac laptops running PsychoPy (v1.82.1, v1.85.0, or v1.85.2). In China, children were tested in a private, quiet room of their preschool, on a PC laptop running PsychoPy v1.84.2-win32.

Predictions

Predictions are framed with pitch contour responses as “correct” responses. If tone speakers weight pitch more strongly than non-tone speakers, there should be an accuracy interaction between Language Group (tone, non-tone) and Cue Conflict (combined cues, conflicting cues), with a larger drop in accuracy for non-tone speakers when the timbre cue conflicts. For combined-cue trials, we predicted that tone speakers would be more accurate because they have stronger access to pitch cues than non-tone listeners.

Results

Counter to predictions, the two groups of children performed very similarly (Figure 4a). In original test trials (Figure 4a, left), accuracy was comparable between groups, and was higher for distinct timbres. In conflict test trials (Figure 4a, right), accuracy dropped for both groups. To

verify these observations analytically, we conducted mixed-effects logistic regressions using package lme4 (Bates et al., 2015) in R 3.3.2 (R Core Team, 2016) with accuracy (melody match) as the dependent variable. Logistic regression takes into account the underlyingly binomial nature of accuracy data. For initial analyses, contour match responses were counted as correct. Language (tone, non-tone) was a between-groups factor, and Timbre Similarity (similar, distinct) and Cue Conflict were within-subjects factors. All possible interactions were included, plus random participant intercepts and slopes. Predictors were converted to numeric values and mean-centered prior to analysis, allowing ANOVA-like interpretation of effects.

Insert Figure 4 about here (two panels, A and B)

Accuracy in Block 1 (Original Stimuli).

In Block 1, conflict trials had not been presented yet, assessing initial learning in a Language x Timbre Similarity design. Only Timbre Similarity was significant (estimate = 1.45, $SE = 0.31$, $z = 4.67$, $p < .0001$): with higher accuracy in selecting the matching cartoon for distinct timbres (75.3%) than for similar timbres (58.6%). Neither Language (estimate = -0.36, $SE = 0.35$, $z = -1.03$, $p = .30$), nor the Language x Timbre Similarity interaction (estimate = 0.24, $SE = 0.55$, $z = 0.43$, $p = .67$) approached significance. Thus tone-language speakers did *not* display an advantage from greater access to pitch cues. Logistic regression tests at each level of Timbre Similarity, which assessed whether the intercept differed from chance, were significant, suggesting that both similar timbres (estimate = 0.51, $SE = 0.18$, $z = 2.83$, $p = .005$) and distinct timbres (estimate = 1.94, $SE = 0.25$, $z = 6.73$, $p < .0001$) exceeded chance accuracy.

Accuracy in Block 2 (Combined-Cue vs. Cue Conflict Stimuli)

Next, we analyzed performance in the second block of trials (Figure 4a, middle, and left), where half of the trials were the combined-cue stimuli and half were cue-conflict trials. Accuracy was again scored as match to the pitch contour, so timbre responding on conflict trials would yield below-chance accuracy. Language, Timbre Similarity, Conflict, and all interactions were included as fixed effects. A main effect of Conflict (estimate = -2.31, $SE = 0.26$, $z = -8.86$, $p < .0001$) indicated lower accuracy on conflict trials than combined-cue trials. A Timbre Similarity x Conflict interaction (estimate = 3.51, $SE = 0.48$, $z = 7.32$, $p < .0001$) indicated that the drop in accuracy from original to conflict stimuli was larger in the distinct-timbre case. No other effects approached significance (all $p \geq .17$), including Language and its interactions, suggesting children of the two language backgrounds performed similarly.

Accuracy in Block 2 Rescored as Timbre Matching

The above analyses of Block 2 accuracy suggest that timbre has a strong influence in identifying character-associated melodies. However, this does not directly gauge influence of pitch contour, which should show up as a decrease in timbre-based classification on conflict trials. To test this, we re-ran the second analysis but redefined accuracy as matching *timbre* (Figure 4b). Only Timbre Similarity was significant (estimate = 1.66, $SE = 0.21$, $z = 7.76$, $p < .0001$), with no change in accuracy apparent between combined-cue and cue-conflict trials (estimate = -0.14, $SE = 0.12$, $z = -1.17$, $p = .24$). Neither Language (estimate = 0.14, $SE = 0.24$, $z = 0.58$, $p = .56$), nor its interactions (all $p \geq .22$) approached significance. This suggests that participants of both language backgrounds responded based on timbre, not pitch contour.

Analyses Including Age

Expanded exploratory analyses included the continuous factor of Age and all of its possible interactions. For Block 1 accuracy, there was an effect of Timbre Similarity (estimate = 1.48, $SE = 0.31$, $z = 4.80$, $p < .0001$), and an effect of Age (estimate = 0.47, $SE = 0.17$, $z = 2.77$, $p = .006$), such that older children were more accurate. A Timbre Similarity x Age interaction (estimate = 0.67, $SE = 0.28$, $z = 2.40$, $p = .02$) resulted from age improvement for the distinct timbre (estimate = 0.77, $SE = 0.25$, $z = 3.07$, $p = .002$) but not the similar timbre (estimate = 0.11, $SE = 0.18$, $z = 0.64$, $p = .52$). No effects or interactions involving Language approached significance ($p \geq .12$).

For the second block of trials, effects of Conflict (estimate = 2.10, $SE = 0.21$, $z = 9.86$, $p < .0001$) and Conflict x Timbre Similarity interaction (estimate = 3.00, $SE = 0.39$, $z = 7.60$, $p < .0001$) both reached significance. There was an interaction of Conflict x Age (estimate = 0.89, $SE = 0.21$, $z = 4.33$, $p < .0001$), which was qualified by a 3-way Timbre Similarity x Conflict x Age interaction (estimate = 1.36, $SE = 0.38$, $z = 3.54$, $p = .0004$). The apparent source of the 3-way interaction is that children at all ages show the Timbre Similarity x Conflict interaction pattern, but the interaction is larger in magnitude for 4-5-year-olds. No effects of or interactions with Language approached significance ($p \geq .14$).

For Block 2 accuracy rescored as timbre match, there was a Timbre Similarity effect (estimate = 1.79, $SE = 0.25$, $z = 7.27$, $p < .0001$). There was also an Age effect (estimate = 0.51, $SE = 0.12$, $z = 4.24$, $p < .0001$), consistent with improvement over age, and a Timbre Similarity x Age

interaction (estimate = 0.84, $SE = 0.23$, $z = 3.68$, $p = .0002$). The interaction appears to result from large improvements in distinct-timbre accuracy (estimate = 0.92, $SE = 0.21$, $z = 4.29$, $p < .0001$) with age, but no age improvement in similar-timbre accuracy (estimate = 0.08, $SE = 0.08$, $z = 0.93$, $p = .35$). Neither Language nor interactions with Language approached significance ($p \geq .11$).

Exploratory Analysis of High Accuracy Participants

Creel et al.'s (2018) discrimination data (see Figure 1) suggest that in the similar-timbre condition, contour should outrank timbre for tone speakers. Was low accuracy in this condition masking group differences in cue weighting? We assessed high-scoring children ($\geq .75$) on combined-cue trials in Block 2. In the similar-timbre condition, 18 tone and 15 non-tone participants met this inclusion threshold. In that group, did conflict trial performance look different between tone and non-tone participants? Using a logistic regression to test for a Language x Conflict interaction did not show a significant result (estimate = 0.90, $SE = .67$, $z = 1.35$, $p = .18$). This suggests that, even among children who learned well, there was not a strong effect of tone vs. non-tone language on conflict trial performance.

Support for the Null Hypothesis

When one observes a lack of effect, as we did for language (tone vs. non-tone), it is often unclear whether the lack of effect is due to evidence in favor of the null hypothesis or insufficient power. To assess this, we conducted a Bayes factor analysis. The Bayes factor is a ratio of the evidence for the experimental hypothesis vs. the evidence for the null. Conventionally, a Bayes factor of 3 or more reflects greater evidence for the alternative hypothesis, while a Bayes factor of 1/3 or

less reflects greater evidence for the null hypothesis. A Bayes factor (ratio) of 1 indicates that the evidence is equivocal.

We calculated Bayes factors for two Language effects: overall accuracy in Block 1 (combined-cue trials only); and interaction scores in Block 2 (combined-cue minus cue-conflict). Recall that tone language speakers were hypothesized to perform better in Block 1 simply because they had access to more cues (tone-language: both pitch and timbre; non-tone: mostly timbre alone). We averaged all Block 1 trials for each participant, collapsing across timbre. The difference between group means was -0.052, with slightly higher accuracy for the non-tone group; the standard error was 0.046. We used Tattan-Birch's Bayes factor calculator at BayesFactor.info, which implements Dienes' calculator (see Dienes, 2014). To characterize the alternative hypothesis mean difference, we assumed that the two groups' difference (tone minus non-tone) would fall between 0 and .2 (uniform distribution). Note that assuming a wider interval, such as 0 and .5, represents a *less* conservative test of the null (that is, it would favor the null). The derived Bayes factor was .14, which is less than .33. We can interpret this as some support for a *lack* of tone-language advantage. For the Language x Conflict interaction, we calculated differences between the two language groups' Conflict scores, which should be larger for non-tone speakers as they should respond more unequivocally to timbre. The difference was .017, numerically larger for non-tone speakers, with $SE = .065$. Again, the null was formulated as varying between 0 and .2. In this case, the Bayes factor was .51. This is greater than .33, meaning that while it nominally favors the null hypothesis, it points to the need for more sensitive tests in the future.

Discussion

We asked whether children who speak a tone language show heightened weighting of pitch cues as do adults who speak a tone language. To address this question, we compared children who did vs. did not speak a tone language on their relative weighting of pitch contour and timbre cues. Specifically, children learned to associate different cartoon characters with two different pitch contour-timbre combinations, and then contour and timbre cues were put in conflict. Tone-language speaking children did not show heightened weighting of pitch information—and in fact showed little evidence of using pitch contour at all, performing similarly to non-tone language children. This contrasts with findings of robust contour change detection advantages over non-tone language speaking children (Creel et al., 2018; see also Deroche et al., 2019), and heightened pitch weighting in adult tone speakers (Jasmin et al., 2021). These findings combined suggest that increased pitch cue weighting develops later than heightened pitch *sensitivity* in tone language speakers.

Another way to view these findings is that they provide cross-cultural evidence of a general timbre bias in childhood. Recall that this “extreme” timbre responding pattern shown by both child groups here and by children in Creel (2014) is not evident in adults. It is especially interesting that this timbre bias occurs even in children learning a language in which pitch is critical to comprehension. This implies that language-specific features may take some time during development to shift initial perceptual weightings.

Open Questions and Future Directions

An obvious question is why Jasmin et al. (2021) found stronger pitch weighting in adult tone speakers, but we did not find it in children. We suggest that pitch weighting may develop slowly

via not just language exposure but also general auditory exposure. This would be consistent with Creel (2014)'s findings of stronger pitch weighting (vs. timbre and vs. absolute pitch height) in adults than 3-5-year-old children. It is also consistent with recent word-learning and word-recognition studies of 2-3-year-old Mandarin speakers (Ma et al., 2017) which suggest that children give stronger weight to vowels than to lexical tones. Thus, even for tone-language speakers, pitch may not yet figure prominently in recognition during the preschool years.

However, an alternative explanation is that timbre (tested here) is not downweighted in the same way as the loudness and duration cues used by Jasmin et al. (2021). This could be tested in the future in adult tone and non-tone language speakers.

Another question is why Creel et al. (2018) showed robust differences between tone and non-tone children's contour change detection, but contour responding was largely absent here. One possibility is that the exact contour stimuli used here (5 rising vs. 5 falling notes) were less discriminable than those in Creel et al. (2018) (rising vs. falling major thirds or perfect fifths). This seems unlikely given that additional studies (Creel, under review) suggest that the rising vs. falling patterns are approximately as distinguishable as rising vs. falling perfect fifths (the dissimilar contours in Creel et al., 2018), at least for non-tone speaking children. Future work should test discrimination and cue weighting within subjects using matched stimuli to rule out inadvertent population and stimulus differences.

A third question is whether differences might emerge at higher accuracy levels. That is, perhaps floor effects on similar-timbre trials obscured differences between tone and non-tone groups. If so, a prediction for future research is that language-specific performance patterns may emerge if

overall performance improves, such as at later child ages or with pitch patterns that are more familiar.

Conclusion

Children whose tone language experience gives them an advantage in musical pitch perception do not show evidence of stronger weighting of pitch information in a learning task. This differs from recent evidence that tone-language speaking adults weight pitch more strongly than non-tone speakers do (Jasmin et al., 2021). Future work should track the developmental course of reweighting of pitch information and its interaction with language(s) spoken.

Author Note

The authors declare no conflicts of interest.

Thanks to Creel lab research assistants for their work on data collection in the United States, and Xiaoxing He and Xinyan Jia for data collection in China.

Correspondence concerning this article should be addressed to Sarah C. Creel, Social Sciences Research Building, Room 236, 7835 Trade St. Suite 100, San Diego, CA 92121. E-mail: screel@ucsd.edu

References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. DOI:10.18637/jss.v067.i01
- Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 97–103.
<https://doi.org/10.1037/0096-1523.32.1.97>
- Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLOS ONE*, *8*(4).
<http://doi.org/10.1371/journal.pone.0060676>
- Bradley, E. D. (2016). Phonetic dimensions of tone language effects on musical melody perception. *Psychomusicology*, *26*(4), 337–345.
- Creel, S. C. (2014). Tipping the scales: Auditory cue weighting changes over development. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1146–1160. <http://doi.org/10.1037/a0036057>
- Creel, S. C. (2016). Ups and downs in auditory development: Preschoolers' sensitivity to pitch contour and timbre. *Cognitive Science*, *40*, 373–403. <http://doi.org/10.1111/cogs.12237>
- Creel, S. C., Weng, M., Fu, G., Heyman, G. D., & Lee, K. (2018). Speaking a tone language enhances musical pitch perception in 3–5-year-olds. *Developmental Science*, *21*(1), e12503.
<http://doi.org/10.1111/desc.12503>
- Deroche, M. L. D., Lu, H. P., Kulkarni, A. M., Caldwell, M., Barrett, K. C., Peng, S. C., et al.

- (2019). A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children. *Scientific Reports*, 9(1), 1–12. <https://doi.org/10.1038/s41598-018-36393-1>
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5(July), 1–17. <https://doi.org/10.3389/fpsyg.2014.00781>
- Fancourt, A., Dick, F. K., & Stewart, L. (2013). Pitch-change detection and pitch-direction discrimination in children. *Psychomusicology: Music, Mind. and Brain*, 23(2), 73–81. <http://doi.org/10.1037/a0033301>
- Henthorn, T., & Deutsch, D. (2007). Ethnicity versus early environment: Comment on “early childhood music education and predisposition to absolute pitch: teasing apart genes and environment.” *American Journal of Medical Genetics*, 282(2000), 102–103. <http://doi.org/10.1002/ajmg.a>
- Hove, M. J., Sutherland, M. E., & Krumhansl, C. L. (2010). Ethnicity effects in relative pitch. *Psychonomic Bulletin and Review*, 17(3), 310–316. <http://doi.org/10.3758/PBR.17.3.310>
- Hutka, S., Bidelman, G. M., & Moreno, S. (2015). Pitch expertise is not created equal: Cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. *Neuropsychologia*, 71(August), 52–63. <http://doi.org/10.1016/j.neuropsychologia.2015.03.019>
- Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, 94(5), 2595–2603. <http://www.ncbi.nlm.nih.gov/pubmed/8270737>
- Jasmin, K., Sun, H., & Tierney, A. T. (2021). Effects of language experience on domain-general perceptual strategies. *Cognition*, 206(September 2020), 104481.

<https://doi.org/10.1016/j.cognition.2020.104481>

Liu, J., Hilton, C. B., Bergelson, E., & Mehr, S. A. (2021). Language experience shapes music processing across 40 tonal, pitch-accented, and non-tonal languages. *BioRxiv*.

<https://doi.org/https://doi.org/10.1101/2021.10.18.464888>

Ma, W., Zhou, P., Singh, L., & Gao, L. (2017). Spoken word recognition in young tone language learners: Age-dependent effects of segmental and suprasegmental variation. *Cognition*, *159*, 139–155. <https://doi.org/10.1016/j.cognition.2016.11.011>

~~Pajak, B., Creel, S. C., & Levy, R. (2016). Difficulty in learning similar-sounding words: A developmental stage or a general property of learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1377–1399. <http://doi.org/10.1037/xlm0000247>~~

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. <http://doi.org/10.3758/s13428-018-01193-y>

Peretz, I., Gosselin, N., Nan, Y., Caron-Caplette, E., Trehub, S. E., & Béland, R. (2013). A novel tool for evaluating children's musical abilities across age and culture. *Frontiers in Systems Neuroscience*, *7*(July), 30. <http://doi.org/10.3389/fnsys.2013.00030>

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, and Psychophysics*, *71*(6), 1385–1398. <http://doi.org/10.3758/APP>

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

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*(3), 1759–1763. <http://doi.org/10.1121/1.2956470>

Table 1.*Trial Orders*

Phase	N Trials	Pitch patterns
First timbre pair		
Learning	8	Original
<i>Distractors</i>	2	
Learning	8	Original
Testing Block 1	8	Original
Testing Block 2	16	half original, half conflict
<i>Distractors</i>	2	
Second timbre pair		
Learning	8	original
<i>Distractors</i>	2	
Learning	8	original
Testing Block 1	8	original
Testing Block 2	16	half original, half conflict

Figure Captions

Figure 1. Data replotted from Creel et al. (2018) to show performance on both similar and distinct contours and timbres. Left two plots show children in Experiment 1 between ages 3 and 5 years, right two plots show younger samples of tone-language speaking children in Experiment 2. Non-tone speaking children were less accurate at distinguishing pitch contours than distinguishing timbres, while tone-speaking children showed equal performance on contour and timbre discrimination. At a finer grain, tone speakers showed somewhat stronger discrimination of distinct timbres (triangles) than of contours, but stronger discrimination of contours than of similar timbres (circles). This held across three different sets of tone-language participants.

Figure 2. Learning task used in Creel (2014) and the current study. Rounded rectangles indicate visual displays; filled dot patterns approximate the pitch contours in heard musical stimuli.

Figure 3. Rising (left) and falling (right) melodic stimuli in musical notation.

Figure 4. Accuracy for each group of participants, scored both as (A) selecting the stimulus with the associated contour and (B) selecting the stimulus with the associated timbre. Error bars represent standard errors.