Effects of language experience on domain-general perceptual strategies

Kyle Jasmin*, Hui Sun*, and Adam T. Tierney

Department of Psychological Sciences, Birkbeck University of London

* equal contributions

Correspondence to: Kyle Jasmin (k.jasmin@bbk.ac.uk)

Abstract

Acoustic dimensions important in a person's native language have been shown to influence second language perception. Here we show that such effects can extend beyond language. In two experiments, native speakers of Mandarin (N=45)—where pitch is a crucial cue to word identity—placed more importance on pitch and less importance on other dimensions compared to native speakers of non-tonal languages English (N=46) and Spanish (N=27), during the perception of both second language speech and musical beats. In a third experiment, we further show that Mandarin speakers are better able to attend to pitch and ignore irrelevant variation in other dimensions compared to English and Spanish speakers, and even struggle to ignore pitch when asked to attend to other dimensions. Thus, an individual's idiosyncratic perceptual system reflects a complex mixture of congenital predispositions and biases instilled by extensive experience in directing attention to important dimensions in their native language.

Introduction

Acoustic dimensions play different roles across language. For example, in Mandarin, a tonal language, pitch is very important: each syllable is assigned one of four pitch contours which help determine which word was spoken. In non-tonal languages such as English, on the other hand, pitch plays a more secondary role: it helps convey phrase structure and pragmatic characteristics of language—features that are also redundantly conveyed by other dimensions, such as duration and amplitude (Mattys, 2000; Chrabasz, et al, 2014; Streeter, 1978). Pitch is therefore not a primary determiner of word meaning in non-tonal languages. A consequence of this is a marked difference in what acoustic dimensions of speech are attended to during first language acquisition—a child acquiring a tonal language will need to attend much more closely to pitch than one acquiring a non-tonal language.

What effects does the relative importance of different acoustic dimensions in one's native language have on other aspects of perception? One possibility is that effects of language experience on perception are limited to speech. For example, native speakers of tone languages, compared to native English speakers, have been shown to place more importance on pitch, and less importance on other cues, in languages they acquire later in life. For instance, they rely more on pitch when they categorize and produce English stress (Nguyen et al. 2008, Wang 2008, Zhang et al. 2008, Yu and Andruski 2010, Zhang and Francis 2010; but see Chrabaszcz et al. 2014) and phrase boundaries (Zhang 2012). According to *perceptual interference models* of second language speech perception (Iverson and Kuhl 1994, Flege 1995, Best et al. 2001), the explanation for this is that first language speech categories (knowledge about lexical tones) interfere with the perception of second language speech. For example, according to these models, when native Mandarin speakers need to decide whether an English syllable is stressed or

unstressed, they are unable to avoid referring to lexical tone categories in Mandarin, and so end up placing more importance on pitch information (and less on other cues) compared to native English speakers. One prediction generated by these models is that effects of language experience on perceptual strategies should be limited to speech, and should not extend to auditory perception in other domains.

An alternative account of how language experience affects perception is through differential salience of acoustic dimensions—the *dimension-selective attention account*. Attentional theories of how acoustic cues (e.g. pitch, duration, and amplitude) are weighted in speech perception suggest that dimensions that are particularly informative or task-relevant receive more attention (Gordon et al. 1993, Francis and Nusbaum 2002, Holt et al. 2018). In support of this theory, listeners have been shown to alter their perceptual strategies in response to short-term changes in the usefulness of different dimensions (Idemaru and Holt 2011, Winn et al. 2013). Dimension-selective attention models of cue weighting can account for effects of language experience on perceptual strategies by suggesting that repeatedly directing attention to a particular dimension leads to an increase in perceptual salience of that dimension. For example, because native Mandarin speakers have had to rely on pitch to learn words during thousands of hours of Mandarin acquisition, pitch may have become more salient for them compared to speakers of languages where pitch is less important.

These two accounts— perceptual interference and dimension-selective attention—make different predictions about how language experience should affect auditory perception. For instance, pitch is not only a speech cue: it is also important for perception of musical features such as phase boundaries (Palmer & Krumhansl, 1987; Tierney, Russo, & Patel, 2011) and the location of musical ‘beats’ (Hannon, Snyder, Eerola, & Krumhansl, 2004; Ellis & Jones, 2009;

Prince, 2014). According to the perceptual interference model, increased experience with pitch in a linguistic context should not affect perception of music, because lexical tone units cannot be mapped onto musical pitches (Patel, 2010, pp. 39-50). In contrast, the dimension-selective attention model would predict that greater experience attending to pitch would lead to increased sensitivity to pitch *domain generally*, and would therefore affect perception of music as well as speech. Furthermore, the dimension-selective attention account predicts that tone language speakers perceive pitch more saliently. They should therefore have difficulty ignoring pitch cues even when they are irrelevant to the task at hand.

To test these accounts, here we investigated perception of English phrase boundaries and musical beats in native speakers of a tonal language (Mandarin Chinese) and two comparison groups—native speakers of English and of Spanish (non-tonal languages). (The native Spanish speakers were included as a second comparison group to ensure that differences in English proficiency were not the primary determinant of any differences between the Mandarin and English speakers.) For both the music and speech tasks, a two-dimensional stimulus space was created by orthogonally varying the extent to which pitch versus duration patterns implied the existence of a particular structural feature. Participants were presented with each stimulus multiple times throughout the experiment and asked to categorize it as having an early versus late intonational phrase boundary (prosody perception test) or as having duple meter (strong-weak) versus triple meter (strong-weak-weak) (music perception test). We then used logistic regression to calculate cue weights—the extent to which participants' categorizations were influenced by pitch versus durational information. The dimension-selective attention account predicts that the Mandarin speakers would rely more on pitch and less on other dimensions (i.e. in this case, duration) when perceiving and categorizing speech, as well as music.

In addition, to further confirm that pitch was more salient for Mandarin speakers, we asked participants to judge whether one of two words was either higher in pitch or greater in amplitude, while ignoring task-irrelevant changes from trial to trial along the unattended dimension. According to the dimension-selective attention account, Mandarin speakers should show increased performance when attending to pitch as well as a stronger influence of pitch cues on behavior. Crucially, they should also have difficulty ignoring pitch, exhibiting lower performance when asked to attend to another dimension (amplitude), and showing a heightened sensitivity to pitch cues even when attending to another orthogonal dimension.

Methods

Participants

Fifty (50) native speakers of English were recruited from the Prolific online participant recruitment service (prolific.co). An initial automated screening accepted only participants who spoke English as a native language, and an initial questionnaire at the outset of the study confirmed that this was the case. Fifty (50) native speakers of Mandarin were recruited from an ongoing longitudinal study. The Mandarin speakers all had resided within the United Kingdom for around five months and had not previously lived in an English-speaking country. A second non-tonal language group was recruited, consisting of 30 speakers of Spanish who reside in the UK. All participants gave informed consent and ethical approval was obtained from the ethics committee of the Department of Psychological Sciences at Birkbeck, University of London.

We used the Gorilla Experiment Builder (www.gorilla.sc) to create and host our experiment (Anwyl-Irvine, Massonnié, Flitton, Kirkham & Evershed, 2018). Participants were

asked to wear headphones, and automated procedures ensured that participants were all using the Google Chrome browser on a desktop computer. One drawback of online testing is that it can be somewhat more difficult to ensure that participants are fully engaged with the task. In an attempt to minimize spurious data points we only included data from participants for whom, in both the music and prosody categorization tasks, there was a significant relationship ($p < 0.05$) between at least one of the stimulus dimensions (pitch or duration) and categorization responses. (See the task descriptions below for more details.) This criterion caused the exclusion of four Mandarin-speaking participants, five English-speaking participants and three Spanish speakers, resulting in final group totals of 46 Mandarin speakers (mean age 23.7 ± 2.0, 43 F), 45 English speakers (mean age 25.6±5.2, 21 F), and 27 Spanish speakers (mean age 29.5±6.1, 18 F). All the Mandarin speakers arrived late in a second language environment after the age of 21 (mean age of arrival 23.1±1.9) and had only a short length of residence in the UK (mean years 0.4±0.02). However, they had received an extensive amount of English class training in China (mean years 13.6±2.0). The other non-English group, Spanish speakers, showed greater individual variability in their age of arrival (mean age 26.4±6.1), length of residence in the UK (mean years 3.0±2.0) and English class training (mean years 12.1±4.5). Both groups also reported varied music training backgrounds (mean years 2.6±4.2 for Mandarin speakers; mean years 0.4±1.6 for Spanish speakers).

Language task

Stimuli

First, recordings were made of a Standard Southern British English-speaking voice actor reading aloud two different sentences: “If Barbara gives up, the ship will be plundered” and “If Barbara gives up the ship, it will be plundered”. The first six words of each recording were extracted; these recordings were identical lexically but differed in the placement of a phrase boundary, i.e. after “up” in the first recording (henceforth “early closure”) and after “ship” in the second recording (“late closure”). The speech morphing software STRAIGHT (Kawahara & Irino 2005) was then used to morph the early closure and late closure recordings onto one another so that the extent to which acoustic cues imply the existence of a phrase boundary either at the middle or at the end of the phrase could be precisely controlled. (For more details see Jasmin et al. 2019). Pitch and duration were then set to vary across five morphing levels, expressed as percentages, which included 0% (identical to the acoustic pattern for the early closure recording), 25% (a greater contribution of early closure than late closure recording), 50% (equal contribution from both recordings, and therefore ambiguous with respect to the placement of the phrase boundary), 75% (greater contribution of late closure than early closure), and 100% (identical to the pattern for the late closure recording). In total, therefore, there were 25 stimuli (one stimulus for every unique combination of five pitch and duration levels).

Procedure

Participants read instruction slides and completed practice trials in order to get familiarized with the task. During instructions, participants were presented with a clear example of early versus late closure (with original, unaltered pitch and duration cues) and were asked to listen to each

example three times before proceeding. During practice trials they were then presented with these examples, asked to categorize them as early or late closure, and given feedback as to whether they answered correctly. During the test itself, on each trial participants were presented with an auditory stimulus, then asked to click a button to indicate if it sounded more as if the phrase boundary was in the middle ("If Barbara gives up, the ship") or at the end ("If Barbara gives up the ship,"). Each item was presented 10 times, for a total of 250 trials.

Analysis

For each participant logistic regression was conducted to examine the extent to which pitch versus duration influenced their categorization judgments. The outcome variable was the categorization decision for a given trial, with pitch (5 levels) and duration (5 levels) as predictors. The resulting coefficients were then normalized so that they summed to 1 using the following equation:

$$\frac{|\text{PitchCueWeight}|}{|\text{PitchCueWeight}| + |\text{DurationCueWeight}|}$$

Musical beat categorization test

Stimuli

In each trial participants heard 18 tones (a group of 6 tones repeated three times) that varied in pitch and duration patterning. Tones were four-harmonic complex tones with equal amplitude across harmonics and a 15-ms cosine ramp at note onset and offset to avoid transients. Pitch and duration patterns each varied across five levels, which differed in the extent to which the cues

implied a three-note grouping (STRONG weak weak STRONG weak weak) versus a two-note grouping (STRONG weak STRONG weak STRONG weak). The strength of these groupings was conveyed by varying the pitch of the first note of the 2-note or 3-note groupings relative to the other notes in the grouping. An increase in the pitch of a note implied the existence of a strong beat at that location. Similarly, an increase in the duration of a note implied the existence of a strong beat there.

The five pitch levels were [B A A B A A] (strongly indicating a groups of three), [Bflat A A Bflat A A], [A A A A A A] (no grouping structure indicated by pitch), [Bflat A Bflat A Bflat A], and [B A B A B A] (strongly indicating groups of two), where "A" was equal to A440, i.e. 440 Hz. The duration levels manipulated the duration of notes (not the inter-onset intervals, which were always 250 ms). So, the five duration levels (in ms) were [200 50 50 200 50 50] (strongly indicating groups of three), [100 50 50 100 50 50], [50 50 50 50 50 50] (no grouping conveyed by duration), [100 50 100 50 100 50], and [200 50 200 50 200 50] (strongly indicating groups of two). Crucially, the five pitch levels and 5 duration levels were varied orthogonally, for a total of 25 conditions -- thus pitch and duration sometimes conveyed the same grouping pattern (a group of two or a group of three), and for other stimuli conveyed different, competing patterns. Note also that the size of the cues was kept large enough that they should be detectable by most listeners. Psychophysical thresholds were collected in an attempt to confirm whether our participants could hear all cue differences; see below for details.

Procedure

On each trial participants were presented with a sequence, then asked to click a button to indicate if it sounded more as if the beat was on every other note ("STRONG weak STRONG weak STRONG weak") or every third note ("STRONG weak weak STRONG weak weak"). Each of the sequences was presented 10 times, for a total of 250 trials. The experiment began with two practice trials.

Analysis

For each participant logistic regression was conducted to examine the extent to which pitch versus duration influenced their categorization judgments. The outcome variable was the categorization decision for a given trial, with pitch (5 levels) and duration (5 levels) as predictors. The resulting coefficients were then normalized so that they summed to 1 as above.

Dimension selective attention test

Stimuli

First, a recording was made of a voice actor reading aloud two difference sentences: "Dave likes to STUDY music, but he doesn't like to PLAY music" and "Dave likes to study MUSIC, but he doesn't like to study HISTORY". The fourth and fifth words of each recording—"study music"—were extracted; these recordings were, then, identical lexically but differed in the placement of word emphasis (i.e. on "STUDY music" versus "study MUSIC"). The speech morphing software STRAIGHT (Kawahara & Irino 2005) was then used to morph these recordings onto one another so that the extent to which acoustic cues imply the existence of

emphasis on one or the other word could be precisely controlled. (For more details see Jasmin et al. 2019). Pitch and amplitude were then set to vary across four levels, from 0% (identical to the acoustic pattern for the recording with emphasis on the first word) to 33%, to 67%, to 100% (identical to the acoustic pattern for the recording with emphasis on the second word).

Procedure

For the "attend amplitude" condition, on each trial participants were presented with a single two-word phrase, then asked to say which word was louder. If the first word was louder, they clicked on a button marked "1"; if the second word was louder, they clicked on a button marked "2". For the "attend pitch" condition, the procedure was the same, except that participants were asked to indicate which word was higher in pitch. Feedback was presented immediately after each trial in the form of a green check mark for correct responses and a red "X" for incorrect responses. Trial order was randomized. The "attend amplitude" condition was presented in its entirety first, followed by the "attend pitch" condition, to minimize task-switching effects. Each of the 16 stimuli was presented 3 times per condition, for a total of 48 trials per condition and 96 trials overall.

Analysis

Portion correct was calculated separately for "attend amplitude" and "attend pitch" conditions. These values for the Mandarin, English and Spanish groups were compared with a two-sample t-test. To investigate the effects of pitch and amplitude levels on responses, cue weights were calculated. For each participant logistic regression was conducted, with the outcome variable being the categorization decision for a given trial, with pitch (4 levels) and amplitude (4 levels)

as predictors. The coefficients for pitch and amplitude from these regressions were taken as measure of cue weights.

Psychophysical discrimination tests

Stimuli

Participants completed two auditory discrimination tests, one of which assessed their ability to tell sounds apart on the basis of their fundamental frequency (pitch) and the other of which assesed their ability to tell sounds apart on the basis of stimulus duration. For both tests we created a linear continuum of 100 complex tones which varied on the basis of a single dimension. Across both tests stimuli were constructed from four-harmonic complex tones (equal amplitude across harmonics) with initial and final amplitude rise time of 15 ms (linear ramps) to avoid perception of clicks. For the pitch discrimination test, the baseline sound had a fundamental frequency of 330 Hz and the comparison sounds had fundamental frequencies which varied from 330.3 to 360 Hz, while the duration of the sounds was fixed at 500 ms. For the duration discrimination test, the baseline sound had a duration of 250 ms and the comparison sounds had durations which varied from 252.5 to 500 ms, while the fundamental frequency of the sounds was fixed at 330 Hz.

Procedure

Psychophysical thresholds were recorded using a three-down one-up adaptive staircase procedure (Levitt 1971). On each trial participants were presented with three sounds with a constant inter-stimulus-interval of 500 ms, with either the first sound or the last sound different from the other two. Participants were told to press either the “1” key or the “3” key on the keyboard to indicate which sounds was different. No feedback was presented. The comparison

stimulus level was initially set at step 50. The change in comparison stimulus level after each trial was initially set at 10 steps; in other words, the test became easier by 10 steps after every incorrect response and became more difficult by 10 steps after every third correct response. This step size changed to 5 steps after the first reversal, to 2 steps after the second reversal, and to 1 step after the third reversal and for the remainder of the test thereafter. Stimulus presentation continued until either 50 stimuli were presented or eight reversals were reached. Performance was calculated as the mean stimulus levels across all reversals from the second through the end of the test.

Statistical analysis

The cue weights calculated for the prosody experiment and the dimension-selective attention experiment as well as the psychophysical thresholds were not normally distributed. For this reason, comparisons among all three language groups are reported with the non-parametric Kruskal-Wallis H statistic. Comparisons of two language groups are reported with Mann-Whitney *U* tests. As an effect size measure, we report Vargha-Delaney A, which is the probability that a randomly selected value from one group will be greater than a randomly-selected value from another group.

Results

Language and Music Cue Weights

Normalized cue weights for language and music were calculated for each participant. These measures reflect the degree to which participants relied on duration or pitch to perceive speech and musical beats (see Methods). Cue weights differed across the three language groups for both speech ($H(2) = 22.5$, $p < .001$) and musical beats ($H(2) = 34.9$, $p < .001$). For the speech task,

native Mandarin speakers had larger normalized pitch cue weights than both native English speakers (U = 1608, $p < .001$, A = 0.78) and native Spanish speakers (U = 869, $p = .004$, A = 0.70), indicating that they relied on pitch to a greater degree than these groups (Fig. 1a). English and Spanish speakers's cue weights did not differ (U = 468, $p = 0.11$, A = 0.39). Results of the music beat task indicated that native Mandarin speaking participants had larger normalized pitch cue weights than both native English speakers (U = 1697, $p < .001$, A = 0.82) and Spanish speakers (U = 1029, $p < .001$, A = 0.83), indicating that Chinese native speakers also relied on pitch to a greater degree when perceiving musical beats (Fig. 1c). English and Spanish speakers' cue weights did not differ (U = 595, $p = .89$, A = 0.49).

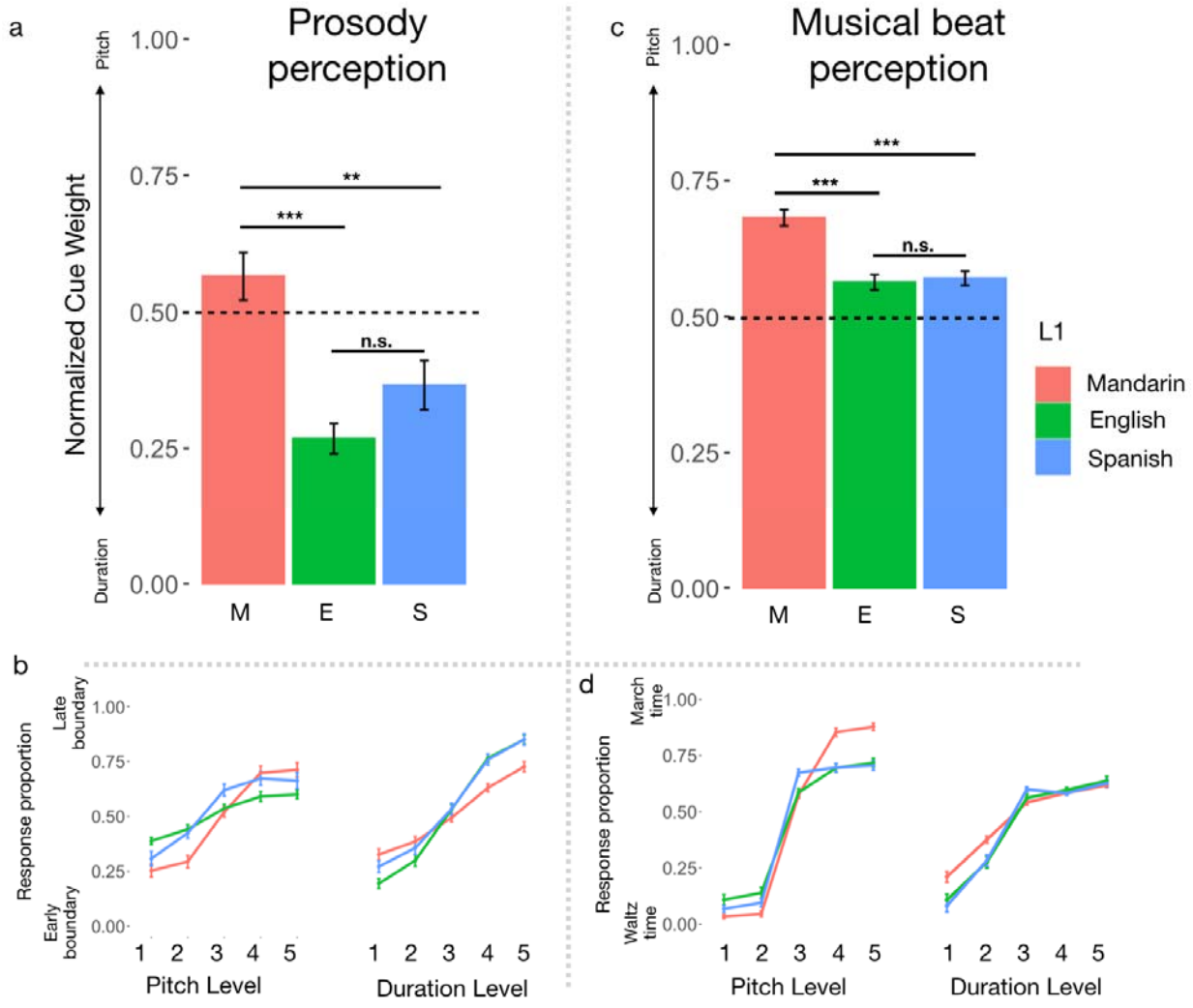

Figure 1. Mandarin speakers rely more on pitch and less on duration when categorizing features in speech and music compared to English and Spanish speakers. **a,b)** Plots of normalized cue weights by language group for the speech task and musical rhythm task**.** Greater values (approaching 1) indicate greater reliance on pitch, and lower values (approaching 0) reflect greater reliance on duration. **c,d)** Plots of responses during the speech task, by pitch level (collapsed over duration) and duration level (collapsed over pitch), for the speech task and music task**.** Values reflect means and error bars represent standard error of the mean.

Dimension selective attention

Attend pitch condition

The dimension selective attention task measured participants' ability to attend to one dimension of speech (pitch or loudness) while simultaneously ignoring the other, independently varying dimension (loudness or pitch). The three groups differed in task performance during the attend-to-pitch condition ($H(2) = 33.8$, $p < .001$). The pitch cue weights also differed across groups ($H(2) = 33.6$, $p < .001$), but the amplitude weights did not ($H(2) = 1.4$, $p = 0.5$). Comparing the groups pair-wise revealed that, when asked to attend to pitch, Mandarin speakers had a greater proportion of correct judgments compared to speakers of English ($U = 1603$, $p < .001$, $A = 0.77$) and Spanish ($U = 1080$, $p < .001$, $A = 0.87$). In line with this result, when asked to attend to pitch, pitch cues exhibited a stronger effect on judgments for speakers of Mandarin than for speakers of English ($U = 1625$, $p < .001$, $A = 0.79$) or Spanish ($U = 1067$, $p < .001$, $A = 0.86$; Fig. 2a). Performance did not differ for the English and Spanish groups ($U = 715$, $p = 0.21$, $A = 0.59$), and neither did pitch cue weights ($U = 686$, $p = 0.37$ $A = 0.56$). Amplitude cue weights did not differ across groups in the attend-pitch condition ($H(2) = 1.4$, $p = 0.5$).

Attend amplitude condition

In the attend-amplitude condition, the groups also differed in task performance ($H(2) = 26.6$, $p < .001$). Pair-wise group comparisons showed that when asked to attend to amplitude, Mandarin speakers showed lower performance than native speakers of English ($U = 407$, $p < .001$, $A = 0.20$) or Spanish ($U = 340$, $p = .001$, $A = 0.27$). English and Spanish speakers did not differ in performance ($U = 711$, $p = 0.23$, $A = 0.58$). The decreased performance in the Mandarin group appears to be driven by Mandarin speakers' difficulty with ignoring task-irrelevant cues

from pitch. Indeed, the three groups differed in the degree to which they responded to task-irrelevant pitch cues when attending to amplitude ($H(2) = 48.5$, $p < .001$). Mandarin speakers exhibited significantly greater task-irrelevant pitch cue weights than speakers of English ($U = 1838$, $p < .001$, $A = 0.89$) and Spanish ($U = 1057$, $p < .001$, $A = 0.85$; Fig. 2a). English and Spanish speakers' pitch cue weights did not differ ($U = 451$, $p = 0.07$, $A = 0.37$). Conversely, Mandarin speakers relied less on the task-relevant amplitude cues compared to English ($U = 483$, $p < .001$, $A = 0.23$) and Spanish ($U = 386$, $p < .001$, $A = 0.31$) speakers, while English and Spanish speakers' amplitude weights did not differ ($U = 697$, $p = 0.30$, $A = 0.57$; comparison across all three groups: $H(2) = 20.0$, $p < .001$).

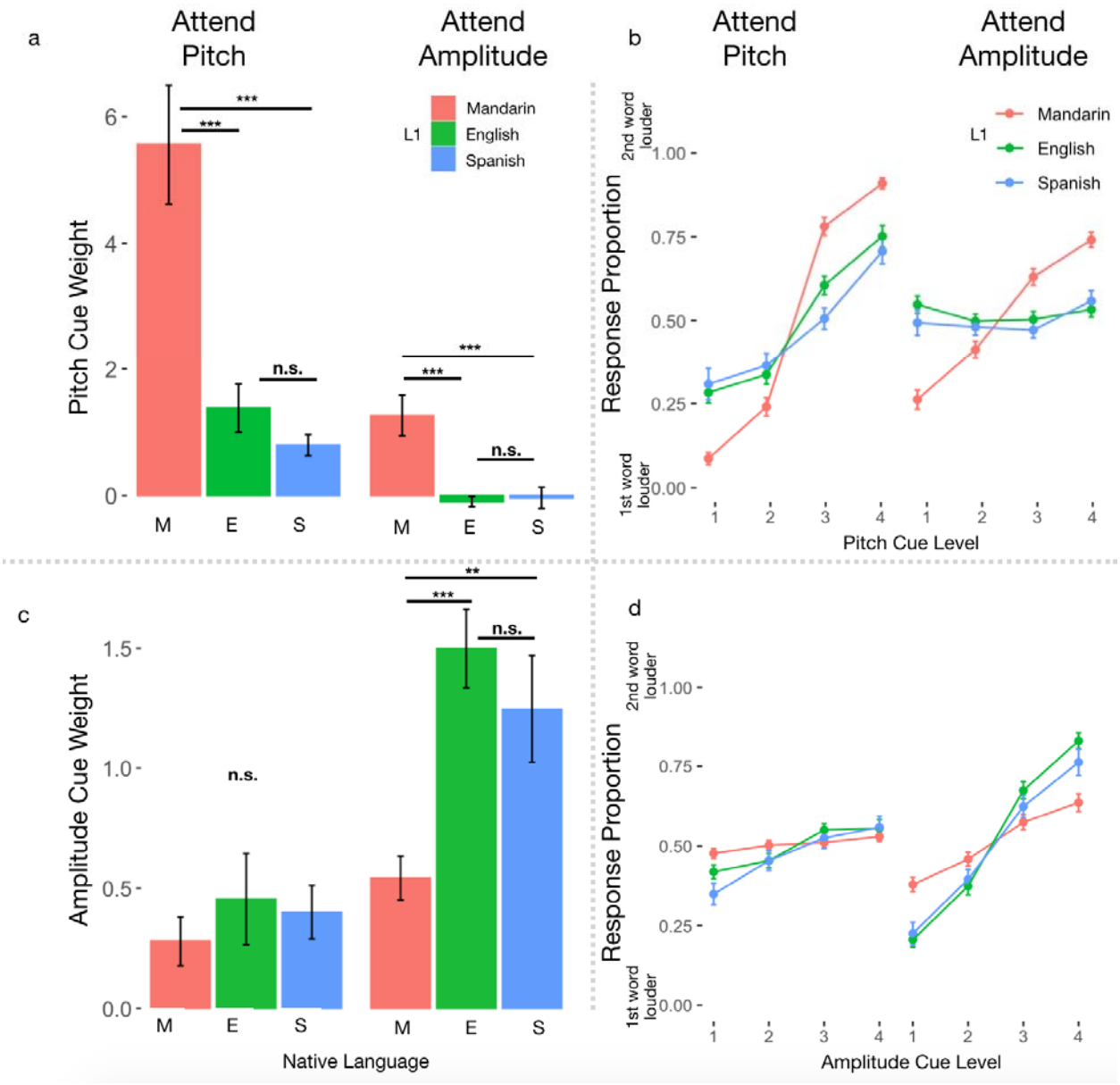

Figure 2. Mandarin speakers rely more on pitch than English and Spanish regardless of its task-relevance. a-b) Pitch cue weights when attending to pitch and to amplitude. **c-d)** Amplitude cue weights when attending to pitch and amplitude. Values reflect means and error bars represent standard error of the mean.

Psychophysics

Finally, we examined pitch and duration thresholds. Duration thresholds did not differ across groups ($H(2) = 1.7$, $p = 0.44$). Pitch thresholds, however, did vary ($H(2) = 15.7$, $p < .001$). Mandarin speakers had lower thresholds than speakers of English ($U = 715.5$, $p = .01$, $A = .34$) and Spanish ($U = 284.5$, $p < .001$, $A = 0.23$). Spanish and English speakers' thresholds did not differ ($U = 473.5$, $p = .12$). Importantly, all pitch thresholds were less than 1 semitone, ensuring that all pitch differences between stimuli were detectable by all participants. Six English and two Spanish speakers had duration thresholds which exceeded 50 ms, the size of the difference between stimuli in the musical beat categorization test. However, the most likely effect of the high duration thresholds for these non-tonal language speaking participants is a down-weighting of duration and up-weighting of pitch during categorization, working against our hypothesis.

To ensure that the Mandarin speakers' lower pitch detection thresholds were not driving the results of our experiments, we matched all three language groups for pitch thresholds by excluding 10 Mandarin speakers with the lowest thresholds, 10 English speakers' with the highest thresholds, and 10 Spanish speakers with the highest thresholds (pitch thresholds $H(2) = 0.97$, $p=0.62$). The results of these analyses were qualitatively the same as those reported in the manuscript, such that all significant results were still significant in the same, and all non-significant results still non-significant.

Discussion

Here we show that native speakers of Mandarin, compared to native speakers of English and Spanish, have different perceptual strategies that are not limited to speech perception but extend to music perception as well. Mandarin speakers place more importance on pitch cues and less

emphasis on durational cues compared to speakers of non-tonal languages, both when judging the locations of linguistic phrase boundaries and of musical beats. This suggests that Mandarin speakers' extensive experience relying on pitch in the course of listening to speech has led to an increase in pitch salience that is not limited to speech perception but extends to other domains. We also find that Mandarin speakers are better able to attend to pitch while ignoring amplitude changes in speech, but are impaired at attending to amplitude while ignoring pitch changes. These results support the dimension-selective attention account of how language experience shapes auditory perception.

Our finding of a difference in perceptual strategies during musical beat perception in Mandarin speakers compared to non-tonal language speakers would not be predicted by several of the existing perceptual interference models which attempt to explain effects of language experience on perception. According to the perceptual magnet model (Iverson and Kuhl 1994), for example, language experience can distort perception, leading to changes in discrimination ability within certain portions of perceptual space. This model, however, cannot account for the Mandarin speakers' up-weighting of pitch and down-weighting of duration during music perception, because we used simple, abstract acoustic stimuli which were likely to be relatively novel to listeners and highly distinct from speech stimuli. Thus, to perform this task, listeners needed to create a new perceptual space, which should, according to the perceptual magnet model, be undistorted by prior language experience. According to the Speech Learning Model (Flege, 1995) and Perceptual Assimilation Model (Best et al., 2001), L2 stimuli are heard as L1 categories, resulting in errant categorization and discrimination and differences in cue weighting. However, again, the stimuli from the music test were so different from speech that they were unlikely to have been assimilated in this manner, i.e. to be heard as L1 categories.

On the other hand, our finding of up-weighting of pitch and down-weighting of duration in Mandarin speakers during both prosodic and musical perception is consistent with a dimension-selective-attention account of cue weighting (Gordon et al. 1993, Francis and Nusbaum 2002, Holt et al. 2018). We suggest that Mandarin speakers' vast experience with relying on pitch in the course of acquiring language has increased pitch salience for them across domains, giving them a characteristic perceptual strategy which extends to music perception as well as to speech perception. This domain-general account makes several additional predictions regarding effects of learning to speak a tonal versus a non-tonal language which could be tested by future work. First, the model predicts that Mandarin speakers will show greater relative pitch weighting across all auditory tasks; if so, this strategy should extend to perception of environmental sounds as well. Second, Mandarin speakers should show greater pitch weighting even when learning a new acoustic category which they have not previously encountered (Holt and Lotto 2006). Third, Mandarin speakers should have difficulty ignoring pitch and attending to other dimensions even in complex non-verbal stimuli. And fourth, it should be possible, by training Mandarin speakers to attend to duration and ignore pitch while categorizing a variety of complex non-verbal sounds, to change cue-weighting strategies during speech perception.

Our finding of up-weighting of pitch during non-verbal perceptual categorization in Mandarin speakers is consistent with prior evidence that speaking a tone language can affect non-verbal auditory processing. In particular, some prior work has shown that, compared with speakers of non-tonal languages, tone language speakers show more precise pitch discrimination in non-verbal sounds, although the literature is somewhat inconsistent on this point (Pfordresher and Brown 2009, Giuliano et al. 2011, Wong et al. 2012, Bidelman et al. 2013, Hutka et al. 2015, Creel et al. 2018; but see Burns and Sampat 1980, Stagray and Downs 1993, Bent et al. 2006,

Peretz et al. 2011). Tone language speakers also show enhanced brainstem and cortical responses to pitch contours in verbal and non-verbal sounds (Krishnan et al. 2005, Chandrasekaran et al. 2007, Swaminathan et al. 2008, Chandrasekaran et al. 2009, Krishnan et al. 2009, Bidelman et al. 2010, Bidelman et al. 2011, Krishnan et al. 2019; but see Xu et al. 2006). Here we find more precise discrimination of the pitch of non-verbal tones in Mandarin speakers compared to speakers of non-tonal languages. However, we would argue for several reasons that the group difference in perceptual strategies cannot simply be a consequence of more precise pitch perception in the Mandarin speakers. First, we took care to ensure that across all three tasks, the size of the pitch differences between stimuli were greater than two semitones, well above participants' discrimination thresholds. Second, increased pitch sensitivity cannot explain our finding that Mandarin speakers perform worse than non-tonal language speakers when asked to ignore pitch and attend to the amplitude of sounds. Lastly, we tested all effects reported in the paper in a subset of participants for which pitch thresholds were equivalent in the three groups: all results persisted. We suggest therefore that Mandarin speakers, compared to non-tonal language speakers, are better able to attend to the pitch of sounds (and less able to direct their attention away from pitch and towards other cues). Although this greater pitch salience cannot simply be reduced to increased pitch sensitivity, there may be a relationship between the two advantages: decades of experience directing attention to pitch may lead to low-level enhancements of the precision of pitch representations (consistent with the Reverse Hierarchy theory of perceptual learning, Ahissar and Hochstein 2004).

In conclusion, here we show that native language experience shapes auditory perception in highly specific ways, not only in perception of other languages, but also for perception of other domains such as music. The results highlight a novel form of linguistic relativity: learning

a language in which words are distinguished by a particular acoustic dimension affects perceptual strategies more generally.

Author Contributions

All authors developed the study concept and contributed to the design. Testing and data collection were performed by H. Sun and A. Tierney. K Jasmin and A. Tierney performed the data analysis and interpretation. All authors drafted the manuscript and approved the final version of the manuscript for submission.

Acknowledgments

We thank Aniruddh Patel for comments on an earlier version of this manuscript.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in cognitive sciences*, *8*(10), 457-464.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our Midst: An online behavioral experiment builder. Behavior research methods, 1-20.

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.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775-794.

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), e60676.

Bidelman G, Gandour J, Krishnan A (2010) Cross-domain effects of music and language experience on the representation of pitch in the auditory brainstem. Journal of Cognitive Neuroscience 23, 425-434.

Bidelman G, Gandour J, Krishnan A (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. Brain and Cognition 77, 1-10.

Burns, E. M., & Sampat, K. S. (1980). A note on possible culture-bound effects in frequency discrimination. *The Journal of the Acoustical Society of America*, *68*(6), 1886-1888.

Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research, 57*, 1468–1479. http://dx.doi.org/10.1044/2014_JSLHR-L-13-0279

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and language*, *108*(1), 1-9.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain research*, *1128*, 148-156.

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.

Ellis, R. J., & Jones, M. R. (2009). The role of accent salience and joint accent structure in meter perception. *Journal of Experimental Psychol- ogy: Human Perception and Performance, 35*, 264–280. http://dx.doi .org/10.1037/a0013482

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*, 233-277.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human perception and performance*, *28*(2), 349.

Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., & Wicha, N. Y. (2011). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Frontiers in psychology*, *2*, 146.

Gordon P, Eberhardt J, Rueckl J (1993) Attentional modulation of the phonetic significance of acoustic cues. Cognitive Psychology 25, 1-42.

Hannon, E. E., Snyder, J. S., Eerola, T., & Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 956–974. http://dx.doi.org/10.1037/0096-1523.30.5.956

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categoriza- tion: Implications for first and second language acquisition. *Journal of the Acoustical Society of America, 119,* 3059–3071. http://dx.doi.org/10 .1121/1.2188377

Holt, L. L., Tierney, A. T., Guerra, G., Laffere, A., & Dick, F. (2018). Dimension-selective attention as a possible driver of dynamic, context- dependent re-weighting in speech processing. *Hearing Research, 366,* 50–64. http://dx.doi.org/10.1016/j.heares.2018.06.014

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*, 52-63.

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension- based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 1939 – 1956. http://dx.doi.org/10.1037/ a0025641

Iverson, P., & Kuhl, P. K. (1994). Tests of the perceptual magnet effect for American English/r/and/l. *The Journal of the Acoustical Society of America*, *95*(5), 2976-2976.

Jasmin, K., Dick, F., Holt, L. L., & Tierney, A. (2019, October 7). Tailored Perception: Individuals' Speech and Music Perception Strategies Fit Their Perceptual Abilities. *Journal of Experimental Psychology: General*. Advance online publication. http://dx.doi.org/10.1037/xge0000688

Kawahara, H., & Irino, T. (2005). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In P. Divenyi (Ed.), *Speech separation by humans and machines* (pp. 167–180). Boston, MA: Kluwer Academic Publishers. http://dx.doi.org/10.1007/0-387-22794-6_11

Krishnan A, Swaminathan J, Gandour J (2009) Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. Journal of Cognitive Neuroscience 21, 1092-1105.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal*

Krishnan A, Suresh C, Gandour J (2019) Tone language experience-dependent advantage in pitch representation in brainstem and auditory cortex is maintained under reverberation. Hearing Research 377, 61-71.

Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics, 62,* 253–265. http://dx.doi.org/10 .3758/BF03205547

Nguyễn, T. A. T., Ingram, C. J., & Pensalfini, J. R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, *36*(1), 158-190.

Palmer, C., & Krumhansl, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 116–126. http:// dx.doi.org/10.1037/0096-1523.13.1.116

Patel, A. D. (2010). *Music, language, and the brain*. Oxford University Press.

Peretz, I., Nguyen, S., & Cummings, S. (2011). Tone language fluency impairs pitch discrimination. *Frontiers in psychology*, *2*, 145.

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, perception, & psychophysics*, *71*(6), 1385-1398.

Prince, J. B. (2014). Pitch structure, but not selective attention, affects accent weightings in metrical grouping. *Journal of Experimental Psy- chology: Human Perception and Performance, 40,* 2073–2090. http:// dx.doi.org/10.1037/a0037730

Stagray, J. R., & Downs, D. (1993). DIFFERENTIAL SENSITIVITY FOR FREQUENCY AMONG SPEAKERS OF A TONE AND A NONTONE LANGUAGE/使用声调语言和非声调语言为母语的人对声音频率的分辨能力. *Journal of Chinese Linguistics*, 143-163.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary percep- tion. *Journal of the Acoustical Society of America, 64,* 1582–1592. http://dx.doi.org/10.1121/1.382142

Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport*, *19*(11), 1163.

Tierney, A. T., Russo, F. A., & Patel, A. D. (2011). The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 15510–15515. http:// dx.doi.org/10.1073/pnas.1103882108

Wang, Q. (2008). *Perception of English stress by Mandarin Chinese learners of English: An acoustic study* (Doctoral dissertation).

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). Roles of voice onset time and F0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering. *Journal of Speech, Language, and Hearing Research, 56*, 1097–1107. http://dx.doi.org/10.1044/1092-4388(2012/ 12-0086)

Wong, P. C., Ciocca, V., Chan, A. H., Ha, L. Y., Tan, L. H., & Peretz, I. (2012). Effects of culture on musical pitch perception. *PloS one*, *7*(4), e33424.

Xu Y, Krishnan A, Gandour J (2006) Specificity of experience-dependent pitch representation in the brainstem. Neuroreport 17, 1601-1605.

Yu, V.Y. & Andruski, J. E. (2010). A cross-language study of perception of lexical stress in English. *Journal of psycholinguistic research*, *39*(4), 323-344.

Zhang, Y., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *The Journal of the Acoustical Society of America*, *123*(6), 4498-4513.

Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics*, *38*(2), 260-271.