Symons, Ashley and Tierney, Adam (2024) Musical experience is linked to enhanced dimension-selective attention to pitch and increased primary weighting during suprasegmental categorization. Journal of Experimental Psychology: Learning, Memory, and Cognition 50 (2), pp. 189-203. ISSN 0278-7393.

Journal of Experimental Psychology: Learning, Memory, and Cognition

Musical Experience Is Linked to Enhanced Dimension-Selective Attention to Pitch and Increased Primary Weighting During Suprasegmental Categorization

Ashley E. Symons and Adam T. Tierney

Online First Publication, June 8, 2023. https://dx.doi.org/10.1037/xlm0001217

CITATION

Symons, A. E., & Tierney, A. T. (2023, June 8). Musical Experience Is Linked to Enhanced Dimension-Selective Attention to Pitch and Increased Primary Weighting During Suprasegmental Categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://dx.doi.org/10.1037/xlm0001217

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

https://doi.org/10.1037/xlm0001217

Musical Experience Is Linked to Enhanced Dimension-Selective Attention to Pitch and Increased Primary Weighting During Suprasegmental Categorization

Ashley E. Symons and Adam T. Tierney
Department of Psychological Sciences, Birkbeck College, University of London

Speech perception requires the integration of evidence from acoustic cues across multiple dimensions. Individuals differ in their cue weighting strategies, that is, the weight they assign to different dimensions during speech categorization. In two experiments, we investigate musical training as one potential predictor of individual differences in prosodic cue weighting strategies. Attentional theories of speech categorization suggest that prior experience with the task-relevance of a particular dimension leads that dimension to attract attention. Experiment 1 tested whether musicians and nonmusicians differed in their ability to selectively attend to pitch and loudness in speech. Compared to nonmusicians, musicians showed enhanced dimension-selective attention to pitch but not loudness. Experiment 2 tested the hypothesis that musicians would show greater pitch weighting during prosodic categorization due to prior experience with the task-relevance of pitch cues in music. Listeners categorized phrases that varied in the extent to which pitch and duration signaled the location of linguistic focus and phrase boundaries. During linguistic focus categorization, musicians upweighted pitch compared to nonmusicians. During phrase boundary categorization, musicians upweighted duration relative to nonmusicians. These results suggest that musical experience is linked with domain-general enhancements in the ability to selectively attend to certain acoustic dimensions in speech. As a result, musicians may place greater perceptual weight on a single primary dimension during prosodic categorization, while nonmusicians may be more likely to choose a perceptual strategy that integrates across multiple dimensions. These findings support attentional theories of cue weighting, which suggest attention influences listeners' perceptual weighting of acoustic dimensions during categorization.

Keywords: speech, music, prosody, attention, pitch

Supplemental materials: https://doi.org/10.1037/xlm0001217.supp

Speech perception requires the listener to integrate information from multiple acoustic dimensions and map this information onto discrete linguistic categories. In natural environments, variation within and between speakers as well as masking of speech by environmental noise pose a challenge for perception. Fortunately, speech is a highly redundant signal, with multiple acoustic dimensions providing overlapping information that can be used for categorization (Winter, 2014). For example, at least 16 acoustic dimensions distinguish the English voiced stop consonant /b/ from the voiceless consonant /p/ (Lisker, 1986). This type of redundancy in speech has been observed across a wide range of linguistic features including focus (Breen et al., 2010), phrase structure (de Pijper & Sanderman, 1994; Streeter, 1978), and syllable stress (Fear et al., 1995; Mattys, 2000).

However, not all acoustic dimensions are equally reliable. The reliability of an acoustic dimension reflects, in part, the learned distributional statistics of the input (Toscano & McMurray, 2010). In the /b/-/p/ example mentioned above, the distributions of the two categories show little overlap in voice onset time (VOT), making VOT a reliable dimension for distinguishing voicing categories in English (Keating, 1984; Lisker & Abramson, 1964). As a result, VOT is often considered to be the "primary" cue to voicing for native English speakers. That is, VOT tends to provide the most reliable information and is weighted more highly than other cues during categorization. Other "secondary" dimensions such as fundamental frequency (F0) may provide less reliable information but can still be used, especially when VOT is ambiguous (Holt et al., 2018). However, there is substantial individual variability in listeners' cue

---

Ashley E. Symons https://orcid.org/0000-0001-5980-6752

A previous version of this work appeared on the website psyArXiv.

Adam Tierney developed the study concept and design. Ashley Symons performed formal data analysis and drafted the manuscript. Adam Tierney provided critical revisions. All authors approved the final version of the manuscript for submission. This work was supported by the Leverhulme Trust [RPG-2019-107] and Economic and Social Research Council [ES/V007955/1].

The processed data that support this study are available at https://osf.io/da6pv/. This study was not preregistered.

Correspondence concerning this article should be addressed to Ashley E. Symons, Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, United Kingdom. Email: a.symons@bbk.ac.uk

weighting strategies (A. C. L. Yu & Zellou, 2019 for review). These individual differences appear to be stable across multiple testing sessions (Idemaru et al., 2012; Kim et al., 2018; Kong & Edwards, 2016), suggesting that they reflect consistent cue weighting strategies rather than random variation. The underlying factors explaining individual variability in cue weighting have only recently begun to be investigated. Recent work on prosodic cue weighting points toward two potential predictors of individual variability in cue weighting strategies: auditory perceptual abilities (Jasmin, Dick, Holt, & Tierney, 2020) and prior experience (Jasmin et al., 2021).

Jasmin, Dick, Holt, & Tierney (2020) tested the hypothesis that differences in cue weighting strategies reflect listeners' auditory perceptual abilities. Compared to controls, listeners with amusia, who have severe difficulties with pitch perception and memory, weighted durational cues more highly relative to pitch across both speech and music categorization tasks. Importantly, this pattern was observed even when pitch differences were large enough for those with amusia to detect. Placing greater weight on duration appeared to help those with amusia overcome pitch perception deficits; controls outperformed amusics when pitch was the only available cue for categorization, but no differences between groups were observed when duration cues were available. These findings suggest that differences in cue weighting strategies reflect, in part, individual differences in auditory perceptual ability.

Another factor that might help predict cue weighting strategies is the degree to which an individual has experience making use of a given acoustic dimension for the purposes of perceptual categorization. For example, speakers of tonal languages, in which pitch variations help determine lexical meaning, weight pitch more highly compared to nontonal language speakers during perception of prosodic features in a nontonal second language (Jasmin et al., 2021; Nguyễn et al., 2008; V. Y. Yu & Andruski, 2010; Y. Zhang & Francis, 2010). Moreover, the effects of tone language experience on cue weighting strategies seem to not be strictly limited to the language domain. As Jasmin et al. (2021) observed, native Mandarin speakers weighted pitch more highly than native English and Spanish speakers across both speech prosody and musical beat categorization tasks. Furthermore, Mandarin speakers were better at selectively attending to pitch and ignoring variations in an irrelevant dimension but showed difficulty ignoring pitch when selectively attending to other dimensions. This suggests that pitch is particularly salient for Mandarin speakers, with the result that they can easily attend to pitch when it is task-relevant but have difficulty ignoring pitch when it is irrelevant. These findings lend support to attentional accounts of cue weighting (Francis & Nusbaum, 2002; Gordon et al., 1993; Holt et al., 2018), which suggest that experience directing attention to a given acoustic dimension increases its salience, or tendency to capture attention, leading to greater perceptual weighting of that dimension during categorization.

Relationship Between Musical Experience and Speech Perception

Another possible source of specialized perceptual experience that could lead to shifts in cue weighting strategies is a musical experience. Many studies now suggest that musicians demonstrate enhanced precision and robustness of auditory processing across multiple domains (Kraus & Chandrasekaran, 2010). This includes speech perception, where musicians show enhanced neural encoding of speech compared to nonmusicians (Musacchia et al., 2007) as well as better behavioral performance on a variety of linguistic measures, including phonological processing (Linnavalli et al., 2018), speech prosody perception (Thompson et al., 2004), and speech-in-noise perception (Du & Zatorre, 2017; Parbery-Clark et al., 2009; Swaminathan et al., 2015; Zendel et al., 2015).

However, not all studies have observed a link between musical experience and enhanced speech perception. In particular, evidence for a relationship between musical experience and speech-in-noise perception has been mixed, with some studies showing no difference between musicians and nonmusicians (e.g., Boebinger et al., 2015; Madsen et al., 2017; Ruggles et al., 2014). One potential explanation for the discrepancy between these findings is that enhanced domain-general cognitive abilities that covary with musical experience may improve speech perception only under certain conditions. Consistent with this explanation, performance on the Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence (WASI), rather than musical experience, predicted speech-in-noise thresholds (Boebinger et al., 2015). There also is evidence that the musician advantage for speech-in-noise perception is only present in conditions where informational masking is present, in other words, when listeners must resist distraction by irrelevant information (Morse-Fortier et al., 2017; Swaminathan et al., 2015).

Given the mixed findings in the literature, whether and under what conditions musical experience may be linked to cross-domain effects in speech perception has been the topic of considerable debate (Besson et al., 2011; Patel, 2011, 2014; Sala & Gobet, 2017; Strait & Kraus, 2011). According to the OPERA hypothesis (Patel, 2011, 2014), musical experience can enhance neural encoding of an acoustic dimension in speech when five conditions are met: (a) There is overlap in the neural foundations of acoustic dimension processing across speech and music domains, (b) Greater precision is required in the processing of acoustic dimensions in music compared to speech, (c) Strong positive emotion is experienced during musical activities, (d) Musical activities that involve the acoustic dimension are frequently repeated, and (e) Musical experience requires focused attention to the dimension. This theory focuses on neural encoding/processing of acoustic dimensions in speech. Here, we hypothesize that any dimension that satisfies the OPERA conditions may also, in addition to being encoded more precisely in the brain, become more salient, or likely to capture attention regardless of task. We test this hypothesis by examining selective attention to and weighting of acoustic dimensions during speech perception in musicians and nonmusicians.

One acoustic dimension that plays an important role in both speech and music is pitch. In English speech, pitch plays an important role in prosodic aspects of speech such as conveying syllable stress, word emphasis, and phrase structure (Breen et al., 2010; de Pijper & Sanderman, 1994; Fear et al., 1995; Streeter, 1978). However, information about prosodic features is also conveyed by multiple redundant acoustic cues. By comparison, music is an informationally brittle signal: information carried by pitch is crucial for conveying melody and harmony (McDermott & Oxenham, 2008) and has no acoustic substitute. For this reason, pitch may serve an especially vital function in music. Throughout the course of musical training, musicians acquire experience directing selective attention to pitch, potentially leading to an increase in the salience of pitch cues. Although pitch salience in musicians and nonmusicians has not to our knowledge been previously investigated, there is substantial prior evidence for enhanced neural encoding and precise perception of pitch in musicians versus nonmusicians. For example, musicians show enhanced pitch discrimination for both nonverbal (Carey et al., 2015; Kishon-Rabin et al., 2001; Micheyl et al., 2006; Nikjeh et al., 2009; Spiegel & Watson, 1984) and verbal stimuli (Besson et al., 2007; Choi, 2020; Magne et al., 2006; Marques et al., 2007; Schön et al., 2004) as well as more robust neural encoding of pitch compared to nonmusicians during speech and music perception (Musacchia et al., 2007; Wong et al., 2007). In addition to enhancing the robustness and precision of pitch representations, experience directing attention to pitch may also increase its tendency to capture attention in a bottom-up fashion, regardless of task. This increased perceptual salience could, in turn, lead musicians to weight pitch information more highly than other acoustic cues during speech perception.

Overview of the Present Studies

Together, these findings suggest that musical experience may be linked to a domain-general enhancement in pitch salience. In line with attentional theories of cue weighting (Francis & Nusbaum, 2002; Gordon et al., 1993; Holt et al., 2018), this enhanced pitch salience may be linked to an upweighting of pitch cues during speech categorization. In two experiments, we examine musical experience as one potential predictor of individual differences in dimension-selective attention and prosodic cue weighting strategies.

Experiment 1 tested the hypothesis that musical experience is linked to cross-domain enhancements in pitch salience. In this experiment, participants with varying degrees of musical experience completed two dimension-selective attention tasks in which they judged which word in a two-word phrase was louder (attend loudness) or higher in pitch (attend pitch), ignoring task-irrelevant variation in the unattended dimension. Following J. D. Zhang et al. (2020), we compared performance between a subset of musicians ($\geq$6 years of musical experience) and nonmusicians (0 years of musical experience). Since music requires greater selective attention to particular dimensions of sound, including pitch, compared to speech (Patel, 2014), we predicted that musicians would show better dimension-selective attention to pitch compared to nonmusicians. Importantly, we also predicted that musicians would show *worse* dimension-selective attention to loudness compared to nonmusicians, due to pitch capturing attention even when task-irrelevant.

If musicians experience greater pitch salience, they may also place greater weight on pitch cues during perceptual categorization. In Experiment 2, we tested this prediction in the same sample of participants who completed Experiment 1. Participants in this experiment completed two prosodic categorization tasks. In each task, a two-dimensional stimulus space was created with stimuli varying in the extent to which pitch and duration signaled category identity. Importantly, this stimulus space contained stimuli whose categorization was fundamentally ambiguous, because information from pitch suggested one interpretation while information from duration suggested another; the manner in which participants categorize these stimuli can reveal the source of information on which participants prefer to rely. Upon hearing each stimulus, participants were asked to categorize the location of linguistic focus or phrase boundary. Based on prior work, we predicted that, on average, listeners would place greater weight on the pitch dimension during linguistic focus categorization and place greater weight on the duration dimension during phrase boundary categorization. However, because of the specific demands that musical experience places on pitch processing, we predicted that musical experience would be associated with increased pitch weighting during both prosodic categorization tasks.

Experiment 1

Introduction

In Experiment 1, we test the hypothesis that musical experience is linked to increased pitch salience during speech perception. Participants completed two dimension-selective attention tasks in which the pitch height and amplitude of two words within a short phrase were varied orthogonally. On each trial, participants decided which word was higher in pitch or louder while ignoring variations in the other dimension. Accuracy on each trial provided a measure of task performance. Based on prior work showing enhanced pitch processing in musicians compared to nonmusicians, we predicted that musicians would show superior dimension-selective attention to pitch but inferior dimension-selective attention to loudness compared to nonmusicians.

Methods

Participants

Ninety-five native English speakers (54 female, 38 male, 3 other gender, $M_{\text{age}} = 33.76$, $SD = 11.81$) were recruited from the Prolific online participant recruitment service (prolific.co). An automated screening procedure accepted only participants who reported speaking English as a native language. This was confirmed by responses to an additional questionnaire. One participant was excluded for reporting that English was not their native language on this questionnaire.

The experiment was conducted via the online experiment platform Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Automated procedures ensured that participants completed the experiment on a desktop or laptop using the Google Chrome browser. All participants were asked to wear headphones throughout the experiment.

To minimize spurious data points, we excluded participants based on categorization responses in Experiment 2. This was done by constructing a logistic regression for each participant and task, with pitch and duration (levels 1 through 5) as continuous predictors and categorization response as the outcome variable. Only data from participants for whom there was a significant relationship ($p < .001$) between at least one of the stimulus dimensions and categorization responses in both the linguistic focus and phrase boundary tasks were included. This exclusion criterion was chosen to eliminate data from participants who were simply responding randomly. However, to ensure that these exclusion criteria did not bias our results in any way, we also ran all analyses on all musicians and nonmusicians and observed an identical pattern of results (see Supplementary Materials).

The Ethics Committee in the Department of Psychological Sciences at Birkbeck, University of London approved all experimental procedures. Informed consent was obtained from all participants. Participants were compensated for their participation in the form of payment at a standard rate.

The final sample consisted of 82 participants (47 female, 32 male, 3 other gender) between the ages of 18 and 66 ($M_{\text{age}} = 34.17$ $SD = 12.29$). From this sample, 40 participants reported regularly engaging in a daily practice of a musical instrument (including voice) for 1–46 years (mean years = 12.20, *SD* = 12). The relationship between dimension-selective attention and musical experience was analyzed by dividing participants into two groups, with musicians defined as those who reported at least 6 years of musical experience (J. D. Zhang et al., 2020) and nonmusicians as those reporting 0 years of musical experience. Any participants who reported 0 years of musical experience but reported an age in response to the question "At what age did you start playing music?" or reported an instrument in response to the question "What instruments can you play (including voice)?" were not included in either group. According to this criterion, 24 musicians and 34 nonmusicians were identified (Table 1). The remaining participants with between 1 and 6 years of musical experience were not included in this analysis. Of the participants in the music group, four reported experience with voice only, 10 reported instrumental musical experience only, eight reported voice and instrumental experience, and two did not provide a response.

To determine whether this study had sufficient power to detect a difference between musicians and nonmusicians in attend pitch versus attend loudness tasks, we conducted an observed power analysis using the *simr* package (Green & MacLeod, 2016) in R (version 4.1.3). To do this, we constructed a mixed-effects logistic regression model with group (musicians, nonmusicians), task (attend pitch, attend loudness), and their interaction as fixed effects and accuracy on each trial as the dependent variable. This model with the group × task interaction was compared to a model without the interaction term using a likelihood ratio test. Based on 100 simulations, a sample size of 58 participants (24 musicians, 33 nonmusicians) provided over 80% power (95% confidence intervals = 96.38%–100%) to detect a significant interaction between group and task.

Stimuli

The stimuli for these tasks were derived from recordings from the MBOPP database (Jasmin, 2020). The initial recordings were made by a Southern British English-speaking voice actor reading aloud two different sentences (capitalization indicating contrastive focus): "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" (Jasmin et al., 2021; Jasmin, 2020). The original stimuli were recorded at a sampling rate of 44.1 kHz, and all morphed stimuli were subsequently presented at this same rate. From the initial recordings, the words "study music" were extracted and morphed along two dimensions (F0 and amplitude) using STRAIGHT software (Kawahara et al., 2013; Kawahara & Irino, 2005) with the standard procedure. First, F0 was extracted from the voiced segments of each phrase. Then the aperiodicity of the signal and filter characteristics were analyzed. This resulted in two morphing substrates, which represent the speech signal decomposed into F0, aperiodic components, and filter characteristics. The morphing substrates were manually time-aligned by marking corresponding "anchor points" (e.g., onset of key words or phonemes) in each recording. Pitch (F0) and loudness (amplitude) were then morphed along four levels: 1 (100% contribution of "STUDY music," 0% contribution of "study MUSIC"), 2 (67% contribution of "STUDY music," 33% contribution of "study MUSIC"), 3 (33% contribution of "STUDY music," 67% contribution of "study MUSIC"), and 4 (0% contribution of "STUDY music," 100% contribution of "study MUSIC"). The use of four levels meant that there was a correct response for each item, allowing us to determine whether musicianship was associated with better selective attention to pitch versus loudness. The four levels of pitch were crossed with four levels of loudness, resulting in 16 unique stimuli, reflecting a combination of each pitch and loudness level. Each stimulus was presented 3 times for a total of 48 stimuli. This decision on the number of trials to include was based on prior work comparing differences between Mandarin and English speakers (Jasmin et al., 2021).

To ensure this morphing procedure produced the desired changes in pitch and loudness across morphing levels and that the different morphing levels were evenly spaced perceptually, we computed differences in mean F0, maximum F0, and mean dB (rms) between "study" and "music" were computed for each of the pitch and loudness levels using Praat (version 6.1.08, Boersma & Weenink, 2019). These features were selected based on prior work showing that differences in mean F0, maximum F0, and intensity signal the location of linguistic focus (Breen et al., 2010). Differences in the mean F0 at each of the four pitch levels were +7.60, 4.29, 0.51, and −3.46 semitones, with negative values indicating a higher mean F0 for "music" compared to "study". Differences in maximum F0 at each of the four pitch levels were +8.71, +2.44, −2.09, and −5.58 semitones, with negative values indicating a higher maximum F0 for "music" compared to "study." Differences in dB (rms) at each of the four loudness levels were 5.23, 1.52, −2.80, and −7.29 dB, with negative values indicating higher dB levels for "music" compared to "study." Examining these differences also shows that the four levels were evenly spaced in semitone and dB space. For example, the differences between successive mean F0 levels were: 3.31 (level 2–1), 3.78 (level 3–2), and 3.97 (level 4–3) semitones. Similarly, the differences between successive loudness levels were: 3.71 (level 2–1), 4.32 (level 3–2), and 4.49 (level 4–3) dB. Tables summarizing the acoustic features of these stimuli can be found in the online supplemental materials.

Table 1

Summary of Age, Gender, and Language Background for Musician (6+ Years of Experience) and Nonmusician (0 Years of Experience) Groups

N	Musicians 24	Nonmusicians 34
Gender	Female: 14 Male: 7 Other: 3	Female: 20 Male: 14 Other: 0
Age	Mean: 34.58 SD: 11.42	Mean = 34.03 SD = 11.94
Languages	# Speak L2: 7 L2 Age Range: 5–25	# Speak L2: 2 L2 Age Range: 5–16

Procedure

Upon signing up to the study, participants were presented with a link to the experiment. After providing informed consent, participants completed a short demographic questionnaire in which they provided information about their age, gender, language background (native and second languages), and musical experience (years of training, age at which training began, instrument(s) on which they were instructed, and hours spent listening to music).

On each trial, participants were presented with a single auditory stimulus and asked to indicate which of the two words was louder (attend loudness task) or higher in pitch (attend pitch task). Participants made their responses by clicking an on-screen button labeled “1” for the first word or “2” for the second word. Feedback was provided on each trial in the form of a green check mark if the response was correct and a red “x” if the response was incorrect. Attend loudness and attend pitch tasks were completed in separate blocks to minimize switching costs, with the order of the blocks counterbalanced across participants.

Data Processing and Analysis

The data were analyzed with a mixed-effects logistic regression model using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2022). The dependent variable was accuracy on each trial (1 = correct, 0 = incorrect). The categorical variables group (musician, nonmusician) and task (attend pitch, attend loudness) were centered (−0.5 and 0.5), and the continuous predictors pitch level (1–4) and loudness level (1–4) were standardized by centering and dividing by 2 *SD*s using the “rescale” function in the *arm* package (Gelman et al., 2021). By standardizing each variable, the beta coefficients from the model represent the change in log odds given an increase of 1 *SD* of that variable. Participant was included a random intercept with random slopes for pitch level, loudness level, and their interaction. Processed data files and data processing scripts for Experiments 1 and 2 are available at: https://osf.io/da6pv/.

Results

Results of the mixed-effects logistic regression model are summarized in Table 2. There was a significant effect of task ($\beta = -0.506$, $p < .001$) as well as a group × task interaction ($\beta = 1.050$, $p < .001$), suggesting that performance was better in the attend pitch compared to the attend loudness task for musicians compared to nonmusicians (Figure 1). To resolve this interaction, we constructed separate models for attend pitch and attend loudness tasks. Compared to nonmusicians, musicians showed better dimension-selective attention performance in the attend pitch task ($\beta = 1.040$, $p < .001$), but not the attend loudness task ($\beta = -0.122$, $p = .486$).

Table 2

Summary of Fixed Effects in a Mixed-Effects Logistic Regression Model for the Dimension-Selective Attention Tasks. Reference Level is Provided in Parentheses for Categorical Predictors

Predictor	Estimate	*SE*	*z*	*p*
Intercept	1.127	0.101	11.180	<.001
Task (attend loudness)	−0.506	0.079	−6.442	<.001
Group (nonmusicians)	−0.119	0.156	−0.760	.447
Pitch	0.081	0.117	0.689	.491
Loudness	0.117	0.124	0.951	.341
Task × Group	1.050	0.130	8.090	<.001
Task × Pitch	−0.214	0.157	−1.359	.174
Group × Pitch	−0.171	0.179	−0.952	.341
Task × Loudness	−0.048	0.157	−0.308	.758
Group × Loudness	−0.210	0.189	−1.110	.267
Pitch × Loudness	0.149	0.251	0.592	.554
Task × Group × Pitch	0.056	0.260	0.216	.892
Task × Group × Loudness	−0.099	0.260	−0.380	.704
Task × Pitch × Loudness	0.358	0.315	1.137	.255
Group × Pitch × Loudness	0.026	0.385	0.066	.947
Task × Group × Pitch × Loudness	0.273	0.520	0.524	.600

To determine whether there was evidence for the null hypothesis, we computed the proportion of correct responses in the attend loudness task for each group and compared performance between musicians and nonmusicians using Bayesian independent samples *t*-tests (two-tailed) using (JASP). The resulting Bayes Factor was 0.359, providing anecdotal evidence in favor of the null hypothesis. By contrast, in the attend pitch task, the Bayes Factor was 283.415, providing strong support for the alternative hypothesis.

Discussion

Experiment 1 tested whether musical experience is linked to an increase in pitch salience during speech perception. In this experiment, listeners judged the pitch and loudness of phrases that varied orthogonally in fundamental frequency and amplitude. Musicians out-performed nonmusicians when attending to pitch and ignoring variations in loudness. However, musicians and nonmusicians showed equivalent levels of performance when attending to loudness and ignoring pitch.

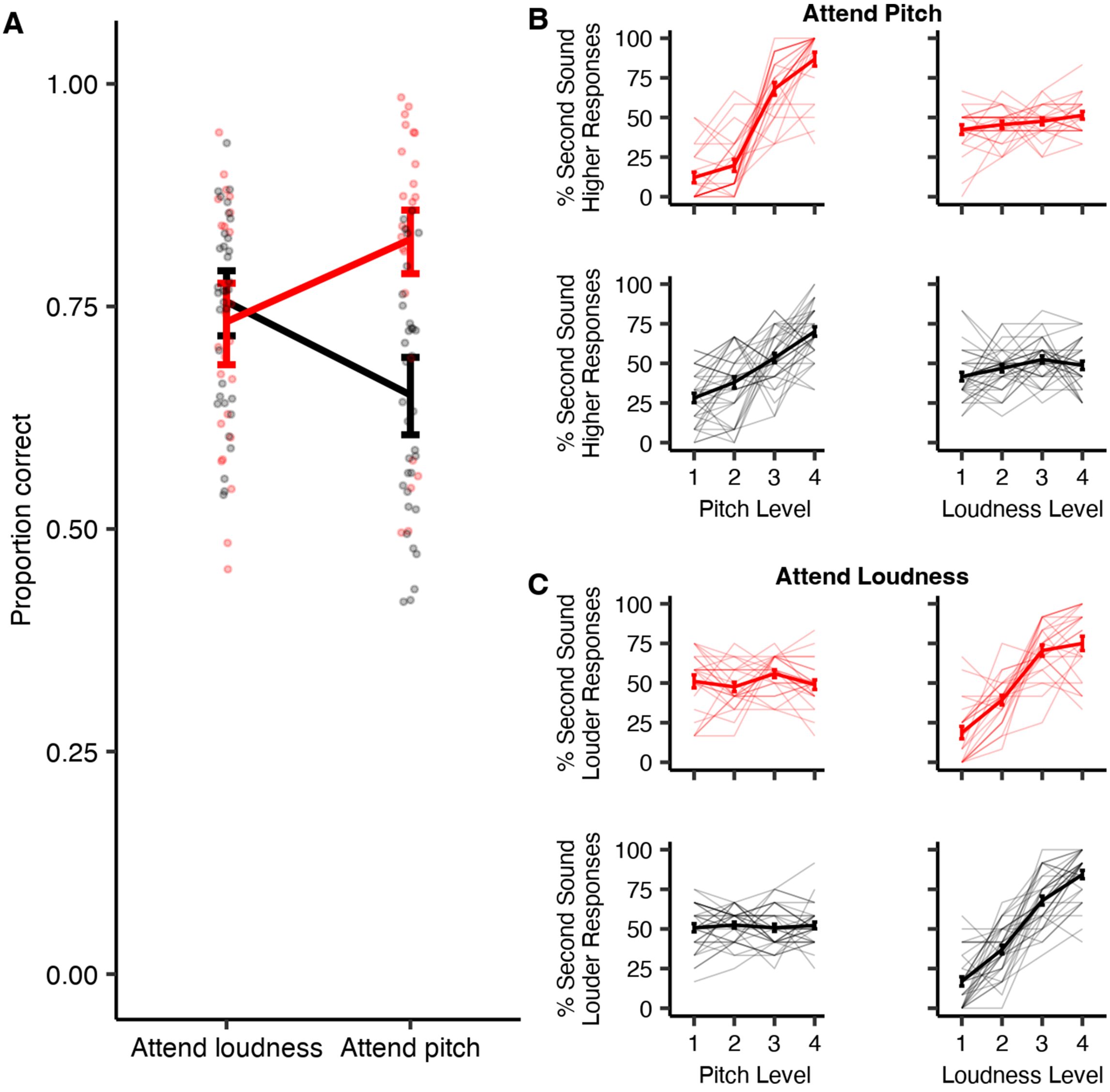
Importantly, these results do not support our hypothesis that musical training is linked to an increase in the salience of pitch, that is, its tendency to capture attention regardless of task. If pitch were more salient for musicians, then they would not only perform better on the attention-to-pitch condition but worse on the attention-to-loudness condition. Instead, these results suggest that musicians benefit from a dimension-specific enhancement of control over attention: they are better able to attend to pitch, but only do so when it is task-relevant. Moreover, this enhancement seems to transfer from music to speech, since the dimension-selective attention task used speech stimuli.

The differences in dimension-selective attention-to-pitch performance between musicians and nonmusicians were greatest near the corners of the stimulus space (see Supplementary Materials) where pitch differences between the first and second word were largest (level 1 = 8.71 semitones, level 4 = −5.58 semitones) regardless of the differences in loudness (level 1 = 5.23 dB, level 4 = −7.29 dB). Thus, it is unlikely that nonmusicians were simply unable to hear the differences in pitch between the first and second words. Instead, the group difference in the attend pitch task likely reflects an enhanced ability on the part of the musicians to maintain their focus on pitch information, despite the distracting variation in loudness information from trial to trial. This finding is consistent with the hypothesis that experience repeatedly directing attention to pitch in music may enhance pitch salience across domains. However, this finding is also consistent with a competing explanation, in which individuals with greater domain-general pitch salience are more likely to seek out and to continue musical training. Future longitudinal intervention research is needed to disentangle these two explanations.

In contrast to the attend pitch task, musicians and nonmusicians did not differ in dimension-selective attention to loudness. Like pitch, loudness varies throughout the course of a musical piece and is an important component of musical affect (Ilie & Thompson, 2006). Therefore, it is not immediately clear why musicians do not show enhanced selective attention to loudness as well. According to the OPERA hypothesis (Patel, 2011, 2014), in order for the effects of musical training to transfer to speech, music must place *greater* demands on the processing of a given acoustic

Figure 1

Comparison of Dimension-Selective Attention Performance Between Musicians (n = 24) and Nonmusicians (n = 34)



Note. Error bars represent *SE* of the mean. Thick lines represent the average across participants while thin lines represent the average percentage of "second sound higher" or "second sound louder" responses for each individual. (A) Predicted proportion correct from the model on the attend pitch and loudness tasks for musicians (red) and nonmusicians (black), with error bars representing the 95% confidence intervals. Individual data points displaying the mean proportion correct for each individual are displayed as lightpoints. (B) Percentage of "second sound higher" responses in the attend pitch task as a function of pitch level (left) and loudness level (right) for musicians (red) and nonmusicians (black). Error bars represent *SE* of the mean. Thin lines represent the average percentage of "second sound higher" or "second sound louder" responses for each individual. (C) Percentage of "second sound louder" responses as a function of pitch level (left) and loudness level (right) for musicians (red) and nonmusicians (black). See the online article for the color version of this figure.

dimension compared to speech. At least for nontonal language speakers, music appears to place greater demands on pitch processing compared to speech (see General Introduction). However, there is currently no clear evidence whether or not this is the case for loudness.

In summary, results from Experiment 1 suggest that musical experience is associated with enhanced dimension-selective attention to pitch but not loudness. This suggests that musical training is linked to an enhanced ability to selectively attend to certain acoustic dimensions in speech. In Experiment 2, we investigated whether this enhanced dimension-selective-attention ability might lead musicians to adopt different strategies for integrating across dimensions during speech categorization, relative to nonmusicians.

Experiment 2

Introduction

The aim of Experiment 2 was to examine, in musicians and nonmusicians, the relative weighting of pitch and duration cues during the categorization of two suprasegmental speech features: word emphasis and phrase boundaries. Based on previous research, we predicted that, on average, listeners would weight pitch more highly than duration in the linguistic focus task (Jasmin, 2020) and duration more highly than pitch in the phrase boundary task (Jasmin et al., 2021). In addition, we originally predicted that musicians would place greater weight on pitch relative to duration information across both tasks, due to musical experience being linked to increased pitch salience. However, Experiment 1 showed that while musicians were better able to attend to pitch when it was task-relevant, they had no difficulty ignoring pitch when it was irrelevant. This suggests that musicians may benefit from an enhanced ability to selectively attend to any acoustic dimension with which they have extensive experience due to practice performing and perceiving music, but do not have difficulty ignoring these dimensions. As mentioned above, pitch may be particularly important in music because of its role in conveying melodic structure. However, duration is also important for conveying certain forms of musical structure, including musical phrasing (A. T. Tierney et al., 2011) and beats (Ellis & Jones, 2009). An alternate possibility, therefore, is that, relative to nonmusicians, musicians would place greater weight on the most useful dimension for a given categorization task, weighting pitch more highly for linguistic focus and duration more highly for phrase boundary.

Methods

Participants

The same participants who completed Experiment 1 took part in Experiment 2. To minimize spurious data points, only participants for whom there was a significant relationship ($p < .001$) between at least one of the stimulus dimensions and categorization responses in both the linguistic focus and phrase boundary tasks were included ($n = 82$). Analysis including all participants are included in the Supplementary Materials.

As in Experiment 1, participants were divided into two groups based on their reported years of musical experience, with musicians ($n = 24$) defined as those having at least 6 years of musical experience (J. D. Zhang et al., 2020) and nonmusicians ($n = 34$) defined as those with 0 years of musical experience.

As in Experiment 1, an observed power analysis was conducted to determine whether this experiment had sufficient power to detect the interactions between group and dimension. For focus and phrase tasks, separate mixed-effects logistic regression models were constructed with group (musicians, nonmusicians), pitch level (1–5), and duration level (1–5), and their interaction as fixed effects and response on each trial as the dependent variable. Based on 100 simulations, a sample size of 58 participants (24 musicians, 34 nonmusicians) provided over 80% power to detect a significant interaction between group and duration level in the phrase task (95% confidence interval [96.38%–100%]) and between group and pitch level [96.38%–100%] in the focus task.

The Ethics Committee in the Department of Psychological Sciences at Birkbeck, University of London approved all experimental procedures. Informed consent was obtained from all participants. Participants were compensated for their participation in the form of payment at a standard rate.

Stimuli

The stimuli used for the focus and phrase tasks were obtained from the Multidimensional Battery of Prosody Perception (Jasmin, 2020).

Linguistic Focus

The focus stimuli were derived from recordings made by a Southern British English-speaking voice actor reading aloud two different sentences (capitalization indicating contrastive focus): “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” (Jasmin, Dick, Holt, & Tierney, 2020; Jasmin, 2020). The original stimuli were recorded at a sampling rate of 44.1 kHz, and all morphed stimuli were subsequently presented at this same rate. The first five words from each recording were extracted to obtain two versions of the same phrase (“Dave likes to study music”) that differed in the location of linguistic focus (“study” vs. “music”). The voice morphing software STRAIGHT (Kawahara & Irino, 2005) was used to create stimuli that varied in the extent to which changes in pitch (F0) or duration cued the focused word. The morphing procedure was identical to Experiment 1 with the exception that pitch and duration were morphed while other dimensions were held constant. Both pitch and duration dimensions varied along five levels that reflect the relative contribution of each original recording to the morphed stimulus: 1 (100% contribution of “STUDY music,” 0% contribution of “study MUSIC”), 2 (75% contribution of “STUDY music,” 25% contribution of “study MUSIC”), 3 (50% contribution of “STUDY music,” 50% contribution of “study MUSIC”), 4 (25% contribution of “STUDY music”, 75% contribution of “study MUSIC”), 5 (0% contribution of “STUDY music”, 100% contribution of “study MUSIC”). The use of five levels meant that the stimulus space included tokens in which each dimension could be perceptually ambiguous (level 3). The five levels of pitch were crossed with the five levels of duration, resulting in 25 focus stimuli, one for each combination of pitch and duration levels.

To ensure that the differences along each dimension were large enough to be detected among the general population, we measured the mean F0 and maximum F0 of the words “study” and music’ and the duration of the stressed syllables using Praat (version 6.1.08, Boersma & Weenink, 2019). The differences in mean F0 between study and music at each of the five pitch levels were +7.72, +5.22, +2.76, −0.40, and −2.97 semitones, with negative values reflecting higher mean F0 for “music” compared to “study”. The differences in maximum F0 between study and music at each of the five pitch levels were +7.96, +4.16, +0.16, −2.82, and −5.41 semitones, with negative values reflecting higher maximum F0 for “music” compared to “study”. The differences in duration between the stressed syllables of “study” and “music” at each of the five duration levels were 195.02, 134.24, 84.96, 20.90, and −19.26 ms, with negative values indicating a longer duration of the stressed syllable in “music” compared to “study”.

Phrase Boundary

The phrase stimuli were derived from recordings made by a 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'll be plundered" (Jasmin, 2020; Jasmin et al., 2021). The original stimuli were recorded at a sampling rate of 44.1 kHz, and all morphed stimuli were subsequently presented at this same rate. The first six words from each recording were extracted to obtain two versions of the same phrase ("If Barbara gives up the ship") that differed in the position of the phrase boundary. The recording in which the phrase boundary occurs after the word "up" is referred to as "early closure" and the recording in which the phrase boundary occurs after "ship" is referred to as "late closure." As with the focus stimuli, STRAIGHT software was used to create stimuli that varied in the extent to which pitch (F0) or duration cued the position of the phrase boundary. Both dimensions varied along five levels: 1 (100% contribution of early closure, 0% contribution of late closure), 2 (75% contribution of early closure, 25% contribution of late closure), 3 (50% contribution of early closure, 50% contribution of late closure), 4 (25% contribution of early closure, 75% contribution of late closure), 5 (0% contribution of early closure, 100% contribution of late closure). The inclusion of five levels meant that the stimulus grid included tokens in which each dimension could be perceptually ambiguous (level 3). This resulted in 25 phrase boundary stimuli, one for each combination of pitch and duration levels.

Differences in the mean F0, maximum F0 and duration of the words "up" and "ship" were calculated using Praat (version 6.1.08, Boersma & Weenink, 2019). The differences in mean F0 between "up" and "ship" at each of the five pitch levels were −1.59, −4.02, −4.86, −6.17, and −7.66 semitones, with negative values indicating higher mean F0 for the word "ship" compared to "up". The differences in maximum F0 between "up" and "ship" at each of the five pitch levels were 0.23, −1.96, −4.17, −6.24, and −8.04 semitones, with negative values indicating higher maximum F0 for the word "ship" compared to "up". Differences in duration at each of the five duration levels were 39.36, −2.36, −92.92, −131.40, and −206.54 ms, with negative values indicating longer duration for "ship" compared to "up".

Procedure

The focus and phrase tasks were presented in separate, alternating blocks. Our aim in presenting the tasks in alternating blocks was to keep participants as engaged with the tasks as possible and minimize potential fatigue that might occur during longer and more repetitive online tasks. There were 10 blocks of each task, each consisting of 25 stimuli (250 stimuli per task). In the first block of each task, participants were presented with a set of instructions alongside an example of each recording in which the pitch and duration were unaltered. Participants were asked to play each example 3 times before proceeding to the practice trials. There were two practice trials, consisting of the same stimuli as the examples. During the practice trials, participants listened to a single stimulus and were asked to categorize the stimulus by pressing one of two buttons on the screen. In the focus task, participants were asked to indicate whether the phrase resembled "STUDY music" or "study MUSIC", and in the phrase task, participants were asked to indicate whether the phrase resembled "If Barbara gives up, the ship" or "If Barbara gives up the ship." When the response was incorrect, the word "Nope…" appeared on the screen. The feedback remained on the screen until participants clicked a button to move onto the next trial. The trial structure of the main tasks was identical to the practice except that feedback was no longer provided.

Data Processing and Analysis

Data were then analyzed with a mixed-effects logistic regression model using the lme4 package (Bates et al., 2015) in R (R Core Team, 2022). The dependent variable was the response on each trial. The categorical variable group (musician, nonmusician) was centered (−0.5 and 0.5), and the continuous predictors pitch level (1–5) and duration level (1–5) were standardized by centering and dividing by 2 *SD*s so that the beta coefficients from the model represent the change in log odds given an increase of 1 *SD* of that variable. Participant was included in a random intercept along with random slopes for pitch level and duration level, and their interaction.

Results

Results of the mixed-effects logistic regression model are summarized in Table 3. In the focus task (Figure 2), participants' categorization responses were influenced by both pitch ($\beta = 1.922$, $p < .001$) and duration ($\beta = 2.159$, $p < .001$). An interaction between group × pitch showed that pitch had a greater influence on the categorization responses of musicians compared to nonmusicians ($\beta = 2.474$, $p < .001$). Additionally, there was a small but significant interaction between pitch and duration ($\beta = -0.379$ $p = .048$).

In the phrase task (Figure 3), participants' responses were influenced by both pitch ($\beta = 1.022$, $p < .001$), and duration ($\beta = 3.542$, $p < .001$). An interaction between group × duration ($\beta = 1.077$, $p = .020$) showed that duration had a greater influence

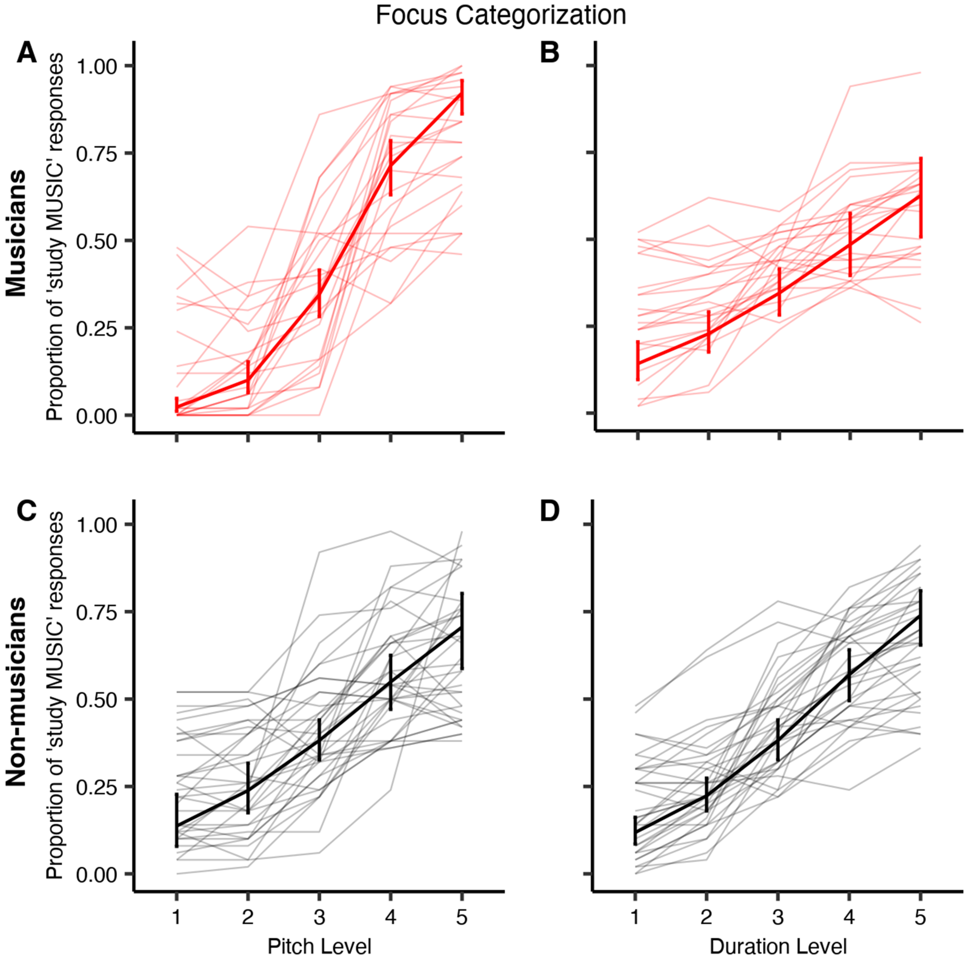
Table 3

Summary of Fixed Effects in a Mixed-Effects Logistic Regression Model for the Categorization Tasks. Reference Level is Provided in Parentheses for Categorical Predictors

Task	Predictor	Estimate	*SE*	*z*	*p*
Focus task	Intercept	−0.483	0.124	−3.892	<.001
	Group (nonmusicians)	−0.156	0.195	−0.801	.423
	Pitch	1.922	0.375	5.127	<.001
	Duration	2.159	0.216	10.003	<.001
	Group × Pitch	2.474	0.593	4.173	<.001
	Group × Duration	−0.522	0.338	−1.546	.122
	Pitch × Duration	0.379	0.192	1.977	.048
	Group × Pitch × Duration	0.382	0.336	1.136	.256
Phrase task	Intercept	0.077	0.133	0.575	.565
	Group (nonmusicians)	−0.141	0.207	−0.681	.496
	Pitch	1.022	0.152	6.744	<.001
	Duration	3.542	0.295	12.006	<.001
	Group × Pitch	0.007	0.235	0.030	.976
	Group × Duration	1.077	0.461	2.334	.020
	Pitch × Duration	0.377	0.195	1.929	.054
	Group × Pitch × Duration	0.215	0.305	0.706	.480

Figure 2

Predicted Proportion of "study MUSIC" Responses as a Function of Pitch Level (A, C) and Duration Level (B, D) in the Focus Task for Musicians (A, B) and Nonmusicians (B, C)



Note. Thick dark lines represent the model predictions with error bars representing the 95% confidence intervals. Thin light lines show the mean proportion of "study MUSIC" responses from each individual participant in the musician and nonmusician groups. See the online article for the color version of this figure.

on the categorization responses of musicians compared to nonmusicians.

Discussion

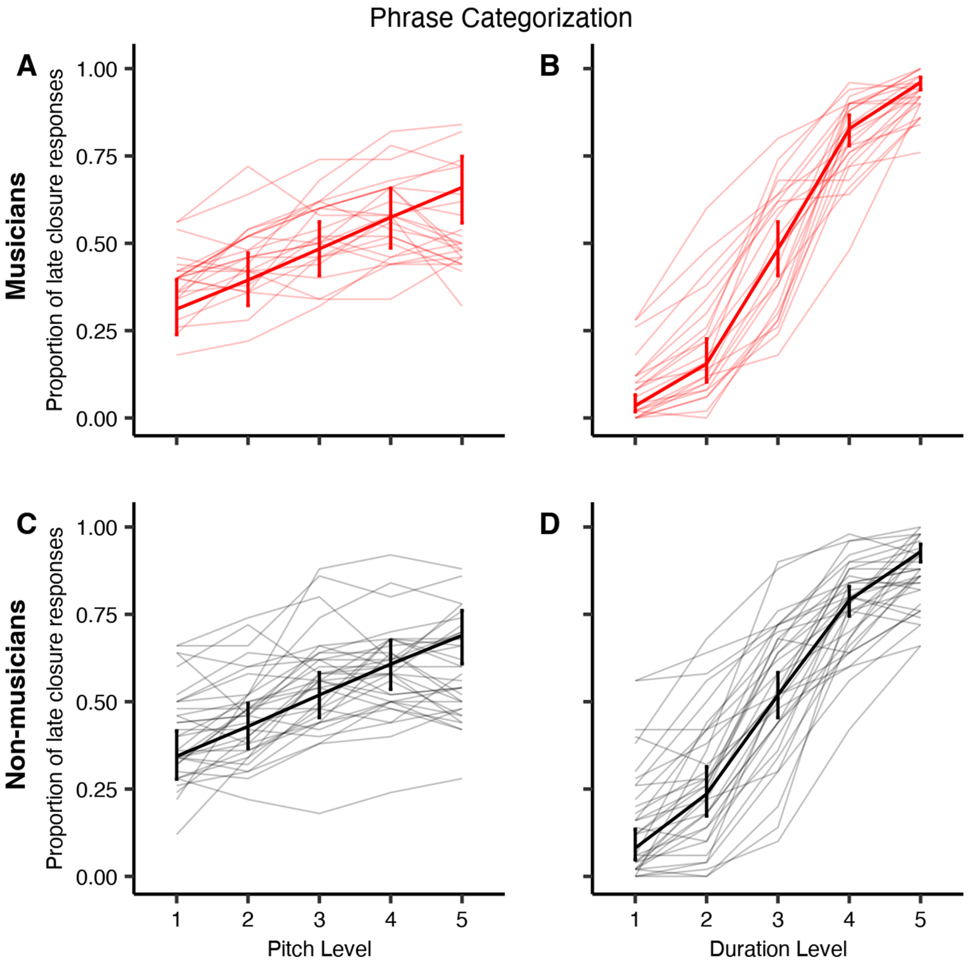
Results from two prosodic categorization tasks showed that the degree of musical experience was linked to listeners' prosodic cue weighting strategies. However, contrary to our predictions, musicians did not simply upweight pitch cues irrespective of task. Instead, musicians weighted pitch more highly in the linguistic focus task and duration more highly in the phrase boundary task.

What could explain musicians' tendency to upweight primary cues during suprasegmental categorization? Our findings are consistent with attentional theories of cue weighting (Francis et al., 2000; Francis & Nusbaum, 2002; Gordon et al., 1993; Heald & Nusbaum, 2014; Holt et al., 2018), which suggest that attention is directed toward acoustic dimensions that are informative. Results from both Experiments 1 and 2 suggest that prior experience directing selective attention to single acoustic dimensions in music to perceive melody and rhythm may lead musicians to adopt a similar perceptual strategy during speech perception. However, we cannot rule out the possibility that individuals with better dimension-selective attention ability may be more likely to seek out musical training. Therefore, future longitudinal work comparing changes dimension-selective attention and cue weighting strategies throughout the course of musical training could provide further insight into the relationship between musical training, dimension-selective attention, and speech perception.

While pitch has perhaps been most extensively studied in experiments of cross-domain transfer of music and speech perception, timing also conveys crucial structural information in music (Patel, 2014). While timing serves an important function in speech perception, the redundancy present in the speech signal may lessen the demands on temporal processing compared to music. Consistent with this idea, musicians show better discrimination ability for both verbal and nonverbal sounds (Sares et al., 2018),

Figure 3

Predicted Proportion of Late Closure ("If Barbara gives up the ship,") Responses as a Function of Pitch Level (A, C) and Duration Level (B, D) in the Phrase Task for Musicians (A, B) and Nonmusicians (B, C)



Note. Thick dark lines represent predicted response proportion with error bars representing the 95% confidence intervals. Thin light lines show the mean proportion of late closure responses from each individual participant in the musician and nonmusician groups. See the online article for the color version of this figure.

as well as larger mismatch negativity (MMN) responses to temporal deviants in speech (Chobert et al., 2011, 2014). Thus, just as with pitch, experience directing attention to temporal cues in music may enhance musicians' ability to attend to temporal information in speech.

One lingering question to be addressed in future research is whether the increased primary weighting and decreased secondary weighting demonstrated by musicians in this study generalize to other acoustic dimensions that are less relevant to music perception. The results of Experiment 1 suggest that musicians' enhanced dimension-selective attention ability does not extend to all acoustic dimensions, because musicians and nonmusicians performed equivalently on the attention-to-loudness task. As a result, if musicians' increased primary weighting is driven by enhanced dimension-selective attention, it may not extend to dimensions such as formant frequency that are important in the perception of both segmental (vowels) and suprasegmental (lexical stress) speech features but are not relevant to music.

General Discussion

In two experiments, we show that musical experience is linked to an enhanced ability to focus on the acoustic dimension most relevant to a given speech perception task, while ignoring less relevant acoustic information. Compared to nonmusicians, musicians showed enhanced dimension-selective attention to pitch but similar attention to loudness (Experiment 1) in speech. Musicians also showed greater reliance on pitch cues relative to duration cues during linguistic focus perception, but greater reliance on duration cues relative to pitch cues during linguistic phrase boundary perception (Experiment 2). These findings suggest that musicians' enhanced ability to selectively attend to certain dimensions leads them to adopt perceptual strategies in which the most reliable cue to a categorization task is upweighted. This interpretation is consistent with attentional theories of cue weighting (Francis & Nusbaum, 2002; Gordon et al., 1993; Holt et al., 2018), which suggest that salient acoustic dimensions receive greater perceptual weight during categorization.

More broadly, our findings support the possibility that experience outside the domain of language may shape listeners' speech perception strategies. As suggested by the OPERA hypothesis, cross-domain transfer can occur when experience in one domain places greater demands on neural resources that are shared across the two domains (Patel, 2011, 2014). Our results provide tentative support hypothesis, suggesting that musical experience can lead to domain-general enhancements in dimension-selective attention abilities that support both speech and music perception. However, in this study, we only examined one type of nonlinguistic experience (music) and cannot differentiate between the effects of musical aptitude versus experience. One promising avenue for future research would be to compare cross-domain transfer effects in populations with different types of auditory expertise. Recent work comparing musicians and audio engineers has found a dissociation in the types of auditory skills influenced by these different types of auditory expertise (Caprini et al., 2021). While musicians showed better auditory selective attention compared to controls, sound engineers showed better memory and recall for auditory scenes. Different types of musical experience may also lead to different perceptual effects, such as enhanced pitch discrimination for musicians trained on variable pitch (string and wind) compared to fixed-pitch (keyboard) instruments (Micheyl et al., 2006) and percussion (Zaltz et al., 2017). Together, these findings suggest that different types of auditory experience can shape different auditory skills. Future work investigating whether and how different types of auditory experience transfer to speech may provide insight into the precise mechanisms that support speech categorization and help inform the development of auditory training protocols to boost speech perception abilities (e.g., Whitton et al., 2014, 2017).

Effect of Musical Experience on Dimension-Selective Attention

Our initial hypothesis was that pitch salience would be greater for musicians due to their history of directing attention to pitch to detect musical features such as melody and harmony. Increased pitch salience would have resulted in enhanced performance on the attention-to-pitch task but impaired performance on the attention-to-loudness task, due to an inability to ignore task-irrelevant pitch information. Instead, we found that while musicians outperformed nonmusicians when attending to pitch and ignoring loudness, there were no differences between groups when attending to loudness and ignoring pitch. This suggests that musicians' enhanced ability to attend to pitch does not reflect bottom-up salience. Instead, it may reflect an enhanced ability to direct attention to specific acoustic dimensions. This broadly aligns with previous work demonstrating enhanced executive function in musicians, including enhanced inhibitory control (Bialystok & DePape, 2009; Moreno & Farzan, 2015; Moussard et al., 2016; Schroeder et al., 2016; Slater et al., 2017; Travis et al., 2011; but see D'Souza et al., 2018; Slevc et al., 2016; A. Tierney et al., 2020), working memory (Clayton et al., 2016; D'Souza et al., 2018; Okada & Slevc, 2018; Slevc et al., 2016; A. T. Tierney et al., 2008; Zuk et al., 2014), and selective attention (Amer et al., 2013; Rodrigues et al., 2013). This prior research has found that musician enhancements in executive function extend broadly across domains (speech vs. music) and modalities (auditory vs. visual). By contrast, our results speak to a more modality and dimension-specific specific process. Consistent with the OPERA hypothesis (Patel, 2011, 2014), our results suggest that musical experience produces cross-domain enhancements to the ability to selectively attend to acoustic dimensions that are directly relevant to music perception and production.

These findings contrast with research in Mandarin speakers, who are better able to attend to pitch and ignore changes in loudness compared to nontonal language speakers but have difficulty ignoring pitch when attending to loudness (Jasmin et al., 2021). In other words, unlike musicians, Mandarin speakers seem to experience increased pitch salience, with pitch tending to capture attention even when it is not task-relevant. One possible explanation for this discrepancy is that while the Mandarin speakers had lifelong language experience, starting at birth, our musicians had comparatively fewer years of experience (6+) and began learning music later in life. Another possible explanation is that language learning in childhood relies on implicit mechanisms, while musical training is primarily explicit. If explicit training later in life leads to changes in dimension-selective attention rather than dimensional salience, then similar effects may be found for second language learning in adulthood. For example, native speakers of a nontonal language learning a tonal language in adulthood may not experience increased pitch salience and this could lead to a pattern like the musicians in the current study (being better able to attend to pitch when it is task-relevant but still able to ignore it when it is not).

Effect of Musical Experience on Prosodic Categorization

Compared to nonmusicians, musicians weighted pitch more highly when categorizing linguistic focus and duration more highly when categorizing phrase boundary location. Prior work has shown that, on average, pitch is the more informative cue for linguistic focus (Breen et al., 2010) while duration is the more informative cue for phrase boundary (de Pijper & Sanderman, 1994; Streeter, 1978). Thus, our results showed that musicians placed greater weight on each dimension only when that dimension was sufficiently informative for categorization. Prosodic cue weighting strategies have been shown to be influenced by both auditory perceptual ability (Jasmin, Dick, Holt, & Tierney, 2020) and language experience (Jasmin et al., 2021). Here we show that musical experience may be another predictor of individual differences in cue weighting strategies. The greater pitch weighting in the linguistic focus task by musicians is consistent with findings from a recent study showing that short-term musical training can influence cue weighting during lexical tone categorization (Wiener & Bradley, 2020). However, our results suggest that the link between musical experience and dimensional weighting may not necessarily be pitch-specific. Instead, although musical experience may be linked to an improved ability to attend to pitch or duration, musicians may only make use of this ability when strongly weighting each dimension would be a successful strategy, due to it being a particularly informative cue to categorization. This possibility can be explored in future longitudinal studies examining shifts in cue weighting across different contrasts and domains (e.g., music vs. speech) throughout the course of musical training.

Our finding that musical experience influences the ability to selectively attend to certain acoustic dimensions during speech perception may have implications for second language learning. One common source of difficulty in second language learning is that individuals need to learn to attend to acoustic cues that are useful for categorization in their new language but were less important in their native language. For instance, native English speakers learning Mandarin have difficulty attending to pitch contour, the most diagnostic cue for tone categorization. However, musical training may confer an advantage when learning lexical tones (e.g., Lee & Hung, 2008; Marie et al., 2011; Zhao & Kuhl, 2015), and has been associated with more robust neural encoding of pitch information (Wong & Perrachione, 2007). Moreover, a recent study found that short-term musical training led to an increase in the weighting of pitch contour during the discrimination of lexical tones (Wiener & Bradley, 2020). Similarly, durational cues can also serve as a vital cue for segmental categorization in many languages (e.g., germinate and singleton consonants in languages such as Italian and Japanese). Although perceiving these durational cues can be a challenge for non-native speakers, there is some evidence to suggest that musical training confers an advantage for discriminating and identifying durational cues in Japanese germinate consonants (Sadakata & Sekiyama, 2011). These findings suggest that musical training may improve the ability to make use of specific acoustic dimensions cues during second language learning. Our results point toward a potential mechanism underpinning this musician advantage. That is, musical training may enhance the ability to selectively attend to acoustic cues (including pitch and duration), facilitating the acquisition of non-native speech contrasts. This link between dimension-selective attention, musical training, and second language learning could be addressed in future longitudinal studies investigating whether changes in the ability to attend to pitch or durational cues throughout the course of musical training correlate with success in learning a second language.

Limitations and Future Directions

Taken together, our findings suggest that musicianship may help explain variation in cue weighting strategies between individuals. It remains unclear, however, whether individual differences in cue weighting strategies remain consistent across domains. The present study employed two tasks with different prosodic contrasts (linguistic focus and phrase boundary); future research could investigate whether perceptual strategies generalize to nonlinguistic tasks (e.g., music perception). Another potential avenue for future research is exploring the extent to which musical training versus aptitude contributes to differences in cue weighting strategies. In the present study, we cannot rule out the possibility that musical aptitude (Mankel & Bidelman, 2018) rather than musical training contributed to the differences between groups. Therefore, future research measuring musical aptitude in trained musicians and nonmusicians may be able to disentangle the extent to which these differences in cue weighting strategies are innate versus experience-dependent. In addition, short-term musical training studies could help disentangle the effects of preexisting musical aptitude versus learning transfer. Nevertheless, our findings suggest musicality as one potential factor underpinning individual differences in prosodic cue weighting strategies.

Our use of an online data collection platform meant that it was not fully possible to control the listening conditions of our participants. Although it is possible that differences in listening conditions could have contributed to the variability in the data, these differences are unlikely to fully explain the differences observed between musicians and nonmusicians. First of all, both groups performed above chance for the unambiguous stimuli, suggesting that participants' listening conditions were sufficient to perform the tasks. Second, differences between musicians and nonmusicians were largely confined to the corners of the stimulus spaces, where acoustic differences between categories were maximal.

Conclusions

In conclusion, here we show an association between musical experience, dimension-selective attention, and prosodic cue weighting strategies. These findings suggest that experience directing attention to acoustic dimensions outside of the domain of language can influence how listeners make use of those dimensions during speech and language processing (Patel, 2011, 2014). These findings also have implications for theoretical models of speech perception. Previous models of speech perception (Toscano & McMurray, 2010) have been based on group-aggregated data that represent the strategies employed by the average listener. Such models implicitly assume a single "optimal" strategy in which the listener relies on the dimension most highly correlated with the statistics of the input. Building upon recent research (Jasmin, Dick, Holt, & Tierney, 2020; Jasmin et al., 2021), our results suggest that individuals differ in their perceptual strategies, which reflect a combination of the statistics of the input, auditory perceptual ability, and auditory experience both within and across domains. Future models accounting for individual differences in the ability to attend to different acoustic dimensions may more closely reflect the variability in listeners' behavior.

References

Amer, T., Kalender, B., Hasher, L., Trehub, S. E., & Wong, Y. (2013). Do older professional musicians have cognitive advantages? *PLOS ONE*, *8*(8), Article e71630. https://doi.org/10.1371/journal.pone.0071630

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00094

Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience*, *25*(3–4), 399–410. PMID: 17943015.

Bialystok, E., & DePape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 565–574. https://doi.org/10.1037/a0012735

Boebinger, D., Evans, S., Scott, S. K., Rosen, S., Lima, C. F., & Manly, T. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America*, *137*(1), 378–387. https://doi.org/10.1121/1.4904537

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* (6.1.08). http://www.praat.org/

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7–9), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Caprini, F., Zhao, S., Chait, M., Agus, T., Pomper, U., Tierney, A., & Dick, F. (2021). *Generalization of auditory expertise in audio engineers and instrumental musicians*. PsyArXiv. https://doi.org/10.31234/osf.io/7fg5h

Carey, D., Rosen, S., Krishnan, S., Pearce, M. T., Shepherd, A., Aydelott, J., & Dick, F. (2015). Generality and specificity in the effects of musical expertise on perception and cognition. *Cognition*, *137*(April 2015), 81–105. https://doi.org/10.1016/j.cognition.2014.12.005

Chobert, J., François, C., Velay, J.-L., & Besson, M. (2014). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex*, *24*(4), 956–967. https://doi.org/10.1093/cercor/bhs377

Chobert, J., Marie, C., François, C., Schön, D., & Besson, M. (2011). Enhanced passive and active processing of syllables in musician children. *Journal of Cognitive Neuroscience*, *23*(12), 3874–3887. https://doi.org/10.1162/jocn_a_00088

Choi, W. (2020). The selectivity of musical advantage. *Music Perception*, *37*(5), 423–434. https://doi.org/10.1525/mp.2020.37.5.423

Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., & Kidd, G. (2016). Executive function, visual attention and the cocktail party problem in musicians and non-musicians. *PloS One*, *11*(7), Article e0157638. https://doi.org/10.1371/journal.pone.0157638

de Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *The Journal of the Acoustical Society of America*, *96*(4), 2037–2047. https://doi.org/10.1121/1.410145

D'Souza, A. A., Moradzadeh, L., & Wiseheart, M. (2018). Musical training, bilingualism, and executive function: Working memory and inhibitory control. *Cognitive Research: Principles and Implications*, *3*(1), Article 11. https://doi.org/10.1186/s41235-018-0095-6

Du, Y., & Zatorre, R. J. (2017). Musical training sharpens and bonds ears and tongue to hear speech better. *Proceedings of the National Academy of Sciences*, *114*(51), 13579–13584. https://doi.org/10.1073/pnas.1712223114

Ellis, R. J., & Jones, M. R. (2009). The role of accent salience and joint accent structure in meter perception. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 264–280. https://doi.org/10.1037/a0013482

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *The Journal of the Acoustical Society of America*, *97*(3), 1893–1904. https://doi.org/10.1121/1.412063

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, *62*(8), 1668–1680. https://doi.org/10.3758/BF03212164

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), 349–366. https://doi.org/10.1037/0096-1523.28.2.349

Gelman, A., Su, Y.-S., Yajima, M., Hill, J., Pittau, M. G., Kerman, J., Zheng, T., & Dorie, V. (2021). *arm: Data analysis using regression and multilevel/hierarchical models* (Version 1.12-2) [Computer software]. https://CRAN.R-project.org/package=arm

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology*, *25*(1), 1–42. https://doi.org/10.1006/cogp.1993.1001

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Heald, S. L. M., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, *8*(2014), 35. https://doi.org/10.3389/fnsys.2014.00035

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*(September 2018), 50–64. https://doi.org/10.1016/j.heares.2018.06.014

Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *The Journal of the Acoustical Society of America*, *132*(6), 3950–3964. https://doi.org/10.1121/1.4765076

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*(4), 319–330. https://doi.org/10.1525/mp.2006.23.4.319

Jasmin, K., Dick, F., Holt, L. L., & Tierney, A. (2020). Tailored perception: Individuals' speech and music perception strategies fit their perceptual abilities. *Journal of Experimental Psychology: General*, *149*(5), 914–934. https://doi.org/10.1037/xge0000688

Jasmin, K., Dick, F., & Tierney, A. T. (2020). The multidimensional battery of prosody perception (MBOPP). *Wellcome Open Research*, *5*(4) https://doi.org/10.12688/wellcomeopenres.15607.1

Jasmin, K., Sun, H., & Tierney, A. T. (2021). Effects of language experience on domain-general perceptual strategies. *Cognition*, *206*(January 2021). 104481. https://doi.org/10.1016/j.cognition.2020.104481

JASP (0.16.2). (2022). https://jasp-stats.org/

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). Kluwer Academic Publishers. https://doi.org/10.1007/0-387-22794-6_11

Kawahara, H., Morise, M., Banno, H., & Skuk, V. G. (2013, October 29–November 1). *Temporally variable multi-aspect N-way morphing based on interference-free speech representations*. 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Kaohsiung, Taiwan (pp. 1–10). https://doi.org/10.1109/APSIPA.2013.6694355

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286–319. https://doi.org/10.2307/413642

Kim, D., Clayards, M., & Goad, H. (2018). A longitudinal study of individual differences in the acquisition of new vowel contrasts. *Journal of Phonetics*, *67*(March 2018), 1–20. https://doi.org/10.1016/j.wocn.2017.11.003

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of Basic and Clinical Physiology and Pharmacology*, *12*(2), 125–144. https://doi.org/10.1515/JBCPP.2001.12.2.125

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*(November 2016), 40–57. https://doi.org/10.1016/j.wocn.2016.08.006

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Lee, C.-Y., & Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *124*(5), 3235–3248. https://doi.org/10.1121/1.2990713

Linnavalli, T., Putkinen, V., Lipsanen, J., Huotilainen, M., & Tervaniemi, M. (2018). Music playschool enhances children's linguistic skills. *Scientific Reports*, *8*(1), Article 8767. https://doi.org/10.1038/s41598-018-27126-5

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, *29*(1), 3–11. https://doi.org/10.1177/002383098602900102

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *WORD*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Madsen, S. M. K., Whiteford, K. L., & Oxenham, A. J. (2017). Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-12937-9

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, *18*(2), 199–211. https://doi.org/10.1162/jocn.2006.18.2.199

Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences*, *115*(51), 13129–13134. https://doi.org/10.1073/pnas.1811793115

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, *23*(10), 2701–2715. https://doi.org/10.1162/jocn.2010.21585

Marques, C., Moreno, S., Luís Castro, S., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–1463. https://doi.org/10.1162/jocn.2007.19.9.1453

Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics*, *62*(2), 253–265. https://doi.org/10.3758/BF03205547

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, *18*(4), 452–463. https://doi.org/10.1016/j.conb.2008.09.005

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*(1–2), 36–47. https://doi.org/10.1016/j.heares.2006.05.004

Moreno, S., & Farzan, F. (2015). Music training and inhibitory control: A multidimensional model. *Annals of the New York Academy of Sciences*, *1337*(1), 147–152. https://doi.org/10.1111/nyas.12674

Morse-Fortier, C., Parrish, M. M., Baran, J. A., & Freyman, R. L. (2017). The effects of musical training on speech detection in the presence of informational and energetic masking. *Trends in Hearing*, *21*(2017), 233121651773942. https://doi.org/10.1177/2331216517739427

Moussard, A., Bermudez, P., Alain, C., Tays, W., & Moreno, S. (2016). Life-long music practice and executive control in older adults: An event-related potential study. *Brain Research*, *1642*(July 2016), 146–153. https://doi.org/10.1016/j.brainres.2016.03.028

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, *104*(40), 15894–15898. https://doi.org/10.1073/pnas.0701498104

Nguyễn, T. A.-T., Ingram, C. L. J., & Pensalfini, J. R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, *36*(1), 158–190. https://doi.org/10.1016/j.wocn.2007.09.001

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians. *The Journal of the Acoustical Society of America*, *125*(1), 328–338. https://doi.org/10.1121/1.3021309

Okada, B. M., & Slevc, L. R. (2018). Individual differences in musical training and executive functions: A latent variable approach. *Memory & Cognition*, *46*(7), 1076–1092. https://doi.org/10.3758/s13421-018-0822-8

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear & Hearing*, *30*(6), 653–661. https://doi.org/10.1097/AUD.0b013e3181b412e9

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, Article 142. https://doi.org/10.3389/fpsyg.2011.00142

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*(February 2014), 98–108. https://doi.org/10.1016/j.heares.2013.08.011

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rodrigues, A. C., Loureiro, M. A., & Caramelli, P. (2013). Long-term musical training may improve different forms of visual attention ability. *Brain and Cognition*, *82*(3), 229–235. https://doi.org/10.1016/j.bandc.2013.04.009

Ruggles, D. R., Freyman, R. L., & Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PLOS ONE*, *9*(1), Article e86980. https://doi.org/10.1371/journal.pone.0086980

Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta Psychologica*, *138*(1), 1–10. https://doi.org/10.1016/j.actpsy.2011.03.007

Sala, G., & Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Directions in Psychological Science*, *26*(6), 515–520. https://doi.org/10.1177/0963721417712760

Sares, A. G., Foster, N. E. V., Allen, K., & Hyde, K. L. (2018). Pitch and time processing in speech and tones: The effects of musical training and attention. *Journal of Speech, Language, and Hearing Research*, *61*(3), 496–509. https://doi.org/10.1044/2017_JSLHR-S-17-0207

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349. https://doi.org/10.1111/1469-8986.00172.x

Schroeder, S. R., Marian, V., Shook, A., & Bartolotti, J. (2016). Bilingualism and musicianship enhance cognitive control. *Neural Plasticity*, *2016*, Article e4058620. https://doi.org/10.1155/2016/4058620

Slater, J., Azem, A., Nicol, T., Swedenborg, B., & Kraus, N. (2017). Variations on the theme of musical expertise: Cognitive and sensory processing in percussionists, vocalists and non-musicians. *European Journal of Neuroscience*, *45*(7), 952–963. https://doi.org/10.1111/ejn.13535

Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition*, *152*(July 2016), 199–211. https://doi.org/10.1016/j.cognition.2016.03.017

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *76*(6), 1690–1695. https://doi.org/10.1121/1.391605

Strait, D., & Kraus, N. (2011). Playing music for a smarter ear: Cognitive, perceptual and neurobiological evidence. *Music Perception*, *29*(2), 133–146. https://doi.org/10.1525/mp.2011.29.2.133

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *The Journal of the Acoustical Society of America*, *64*(6), 1582–1592. https://doi.org/10.1121/1.382142

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, J. G., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Scientific Reports*, *5*(1), Article 1. https://doi.org/10.1038/srep11628

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, *4*(1), 46–64. https://doi.org/10.1037/1528-3542.4.1.46

Tierney, A., Rosen, S., & Dick, F. (2020). Speech-in-speech perception, nonverbal selective attention, and musical training. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(5), 968–979. https://doi.org/10.1037/xlm0000767

Tierney, A. T., Bergeson-Dana, T. R., & Pisoni, D. B. (2008). Effects of early musical experience on auditory sequence memory. *Empirical Musicology Review*, *3*(4), 178–186. https://doi.org/10.18061/1811/35989

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*, *108*(37), 15510–15515. https://doi.org/10.1073/pnas.1103882108

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*(3), 434–464. https://doi.org/10.1111/j.1551-6709.2009.01077.x

Travis, F., Harung, H. S., & Lagrosen, Y. (2011). Moral development, executive functioning, peak experiences and brain patterns in professional and amateur classical musicians: Interpreted in light of a Unified Ttheory of Pperformance. *Consciousness and Cognition*, *20*(4), 1256–1264. https://doi.org/10.1016/j.concog.2011.03.020

Whitton, J. P., Hancock, K. E., & Polley, D. B. (2014). Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise. *Proceedings of the National Academy of Sciences*, *111*(25), E2606–E2615. https://doi.org/10.1073/pnas.1322184111

Whitton, J. P., Hancock, K. E., Shannon, J. M., & Polley, D. B. (2017). Audiomotor perceptual training enhances speech intelligibility in background noise. *Current Biology*, *27*(21), 3237–3247.e6. https://doi.org/10.1016/j.cub.2017.09.014

Wiener, S., & Bradley, E. D. (2020). Harnessing the musician advantage: Short-term musical training affects non-native cue weighting of linguistic pitch. *Language Teaching Research*. Article 136216882097179. https://doi.org/10.1177/1362168820971791

Winter, B. (2014). Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. *BioEssays*, *36*(10), 960–967. https://doi.org/10.1002/bies.201400028

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4), 565–585. https://doi.org/10.1017/S0142716407070312

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422. https://doi.org/10.1038/nn1872

Yu, A. C. L., & Zellou, G. (2019). Individual differences in language processing: Phonology. *Annual Review of Linguistics*, *5*(1), 131–150. https://doi.org/10.1146/annurev-linguistics-011516-033815

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. https://doi.org/10.1007/s10936-009-9142-2

Zaltz, Y., Globerson, E., & Amir, N. (2017). Auditory perceptual abilities are associated with specific auditory experience. *Frontiers in Psychology*, *8*, Article 2080. https://doi.org/10.3389/fpsyg.2017.02080

Zendel, B. R., Tremblay, C.-D., Belleville, S., & Peretz, I. (2015). The impact of musicianship on the cortical mechanisms related to separating speech from background noise. *Journal of Cognitive Neuroscience*, *27*(5), 1044–1059. https://doi.org/10.1162/jocn_a_00758

Zhang, J. D., Susino, M., McPherson, G. E., & Schubert, E. (2020). The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music*, *48*(3), 389–409. https://doi.org/10.1177/0305735618804038

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. https://doi.org/10.1016/j.wocn.2009.11.002

Zhao, T. C., & Kuhl, P. K. (2015). Effect of musical experience on learning lexical tone categories. *The Journal of the Acoustical Society of America*, *137*(3), 1452–1463. https://doi.org/10.1121/1.4913457

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLOS ONE*, *9*(6), Article e99868. https://doi.org/10.1371/journal.pone.0099868

Received December 20, 2021
Revision received November 2, 2022
Accepted November 18, 2022 ■