INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

The role of musical experience in the perceptual weighting of acoustic cues for the obstruent coda voicing contrast in American English

Michelle Cohn[1], *Georgia Zellou*[1], *Santiago Barreda*[1]

[1]University of California, Davis

mdcohn@ucdavis.edu, gzellou@ucdavis.edu, sbarreda@ucdavis.edu

Abstract

This study examines the role of musical experience on listeners' phoneme judgments across noise conditions. Individuals with 10+ years of musical training were compared with nonmusicians in their use of three acoustic cues in categorizing post-vocalic obstruent voicing: fundamental frequency, vowel duration, and spectral composition in two listening conditions (silence and multitalker babble, MTB). Results demonstrate that musicians display steeper phonemic categorization for coda /t/ and /d/ on the basis of all three cues of interest. Additionally, musicians and nonmusicians show different cue weighting patterns in MTB than in silence. The findings are discussed with reference to their implications for theories of experience-driven plasticity and individual differences in the perceptual organization of phonemes.

Index Terms: musical training, categorical perception, auditory experience, obstruent voicing

1. Introduction

There is a growing interest in musicianship as a factor in experience-driven neuroplasticity [1]–[3]. Empirical work has shown that relative to nonmusicians, musicians show more precise neural encoding and enhanced sensitivity to acoustic properties shared by speech and music, relative to nonmusicians, including fundamental frequency (f0), timing, and the spectrum [3]–[5]. Yet, the extent to which musical training can affect speech perception is an open question; currently there is a theoretical debate as to the nature of cross-domain plasticity between speech and music, e.g. [6]–[8]. Indeed, past work has demonstrated mixed results in the literature: with some work suggesting that musical training may confer an advantage when listening to speech in noise (e.g., competing talkers, white noise, etc.) [5]–[8], but others showing no difference between musicians and nonmusicians in similar tasks [13]–[17]. We ask whether some of the differences observed for musicians/nonmusicians in prior work exploring speech-in-noise perception may be driven by differences in *acoustic cue weighting* between the groups.

1.1. Acoustic cue weighting

Multiple acoustic properties can serve as cues for a given phoneme category [18]. For example, the perception of final stop voicing can be cued by the duration of the preceding vowel, formant transitions, and other acoustic features. However, listeners have been shown to vary in the relative importance, or weighting, they give to each cue when making a phoneme judgment, known as 'cue weighting' or 'perceptual weighting'. Some variation in cue-weighting can be attributed to language-specific patterns [19], [20]. For example, in identifying English /r-l/, native Japanese speakers weigh differences in the second formant frequency ($F2$) more heavily while native English speakers rely more on $F3$ [20].

Listeners' patterns of cue weighting can also reflect their attentional strategies in speech perception, particularly in difficult listening situations. In some cases, this can lead to a *re-ranking* of cues. For example, listeners weigh formant-frequency transitions more heavily than vowel duration while making a coda obstruent voicing judgment in silence, yet these preferences are reversed when the same stimuli are presented in multitalker babble (overlayed audio of multiple, competing speakers; MTB) [21], [22]. Still, other work has reported individual variation above and beyond what can be attributed to language experience and listening conditions [23], [24].

1.2. Musicianship and cue weighting

We ask whether nonlinguistic experience, in particular musical training, might shape cue weighting strategies across listening conditions. There are several theoretical predictions as to how skills developed through long-term musical training could transfer to speech perception. On the one hand, *bottom-up accounts* posit that musical training sharpens subcortical and cortical encoding for acoustic features shared by speech and music [25]–[27]. These accounts might predict that musicians would give more perceptual weight to cues sharpened by their experience in both domains, relative to nonmusicians, such as f0, timing, and the spectrum [3]–[5]. Other bottom-up accounts propose that music-to-speech transfer only occurs for features thought to be more heavily 'targeted' by music than in speech, such as f0, e.g., OPERA Hypothesis: [28], as music requires greater precision for pitch (perceived f0) than other features; this varies from speech, which uses more precise spectral and durational cues than for f0, cf. [29]. Thus, one prediction would be that musicians only show stronger weighting for f0, relative to nonmusicians, but no differences for cues that are used in more fine-grained ways in speech, such as timing (e.g., voice onset time) or the spectrum (e.g., formant frequency transitions).

Top-down accounts, on the other hand, posit that transfer from musical training to speech perception primarily impacts higher order cognitive mechanisms, such as attention [30], [31]; one prediction is that musicians and nonmusicians may show different cue weighting strategies across noise conditions based on the attentional demands of more difficult listening situations. In line with this account, in a speaker-gender identification task, [32] found that musicians and nonmusicians showed similar cue weighting strategies for f0 and vocal tract length in silence, but that musicians relied more heavily on f0 when the signal was degraded. Whether musicians/nonmusicians might also vary in their cue weighting strategies for *phonemic contrasts*, however, has not, to our knowledge, been previously explored.

http://dx.doi.org/10.21437/Interspeech.2019-3103

1.3. Current Study

We test whether musicians and nonmusicians differ on the basis of their use of three acoustic cues signaling the post-vocalic obstruent voicing contrast (e.g., /bi**t**/ 'beat' vs. /bi**d**/ 'bead'). We manipulated phonetic properties shown to reliably distinguish [±voice] for obstruent codas in American English: vowel duration [33], vowel f0 [34], and the spectrum, which includes multiple acoustic features, such as $F1$ transition time [35], release burst amplitude [36], and voicing during closure. One prediction is that musicians will show differences in cue weighting strategies based on their increased perceptual acuity for these three cues [3]–[5] and based on their enhanced attention to particular acoustic properties. Another possibility is that both musicians and nonmusicians show similar perceptual weighting of f0, timing, and duration given their similar and extensive experience with speech perception.

2. Methods

2.1. Subjects

Forty-four adult subjects (27 female, 17 male) were recruited from the pool of UC Davis undergraduate students for the present study. Participants were native English speakers, reported no prior experience with a tonal language (e.g., Mandarin Chinese, Thai, etc.) and did not report having any auditory disorders. All participants passed a hearing screening (≤ 25dB from 250-8000 Hz) [37] and fulfilled the inclusion criteria for either the musician or nonmusician group. Musicians (n=22) reported at least 10 years of instrumental practice (mean=12.83y, sd=2.34) and current weekly practice (mean=8.35h, sd=3.93). Nonmusicians (n=22) had minimal musical experience (mean=0.73y, sd=1.08) that occurred at least 7 years ago and did not play any musical instruments at the time of the study. Subjects did not significantly differ in terms of age [$t(40.145)=-1.004$, $p=0.322$]. Subjects were compensated for their time with a $10 Amazon gift card. Subjects also participated in a related speech production study. In total, the study lasted around 40 minutes.

2.2. Stimuli

The target stimuli, "beat" and "bead", were recorded by a native California English speaker (male, age 25) in a carrier sentence ("The word is [beat/bead]") in a sound attenuated booth at a sampling rate of 44.1 kHz with a Shure WH20 XLR head-mounted microphone. The spectral manipulation consisted of the original vowel formants and final consonant (/t/, /d/) from the "beat" and "bead" tokens, excluding the final stop release (following [22]) since stops in word-final position are frequently unreleased.

The duration of the vowel /i/ in both target words was annotated in Praat [38], excluding the release of the initial /b/, starting where $F2/F3$ were clearly visible and ending where the vowel became aperiodic. The vowel duration of /i/ in each of the bVt and bVd tokens was then adjusted along a continuum spanning from 180 to 300 ms in thirteen 10 ms steps. The minimum and median f0 levels were measured from the carrier phrase "The word is bead" recording (min f0=119.96 Hz and median f0=139.06 Hz, respectively) to establish plausible low and high fundamental frequency conditions for the final target word (low f_0=120 Hz, high f_0=140 Hz), and each syllable was monotonized at each f0 level in Praat. Pilot testing confirmed that listeners heard the values as low and high f0s in the context of the sentence frame. The same beginning frame ("The word is b…") from a single recording was used for all tokens. The intensity contours of each vowel were matched using Praat VocalToolkit [39] at a mean intensity of 60 dB. In total, there were 52 tokens (2 words x 2 f0 levels x 13 duration steps).

For the noise condition, we selected a recording of 10+ multitalker babble (MTB) from an hour-long binaural recording from YouTube.com of a crowded bar with no background music (normalized to 60 dB).

2.3. Procedure

Stimuli were presented binaurally over headphones (Sennheiser HD 280 Pro) in a sound attenuated booth, randomized across all stimulus dimensions. On each trial, subjects were asked to identify the word they heard (either "beat" or "bead") using a labeled E-Prime response box (label-button correspondence was counterbalanced across subjects; intertrial interval = 1000ms). Stimuli were presented at 60 dB in two noise condition blocks: Silence or MTB (counterbalanced across subjects). In the silent condition, subjects were not presented any interfering background noise. In the MTB condition, subjects heard the MTB continuously presented throughout the experiment (i.e., during stimulus presentation, ISIs, etc.), following protocol from prior MTB studies [40] (SNR: 0dB). Note that all subjects heard the identical MTB recording in the noise block. Within each block, presentation order of the stimuli was randomized. Subjects responded to two repetitions of each of the 52 stimuli, for a total of 208 trials across the two conditions.

2.4. Analysis

Responses were coded with a 1 for a "bead" response and 0 for a "beat" response. The probability of "bead" responses was modeled with a mixed effects logistic regression with the *lme4* R package [41]. Fixed effects included Group (two levels: musician, nonmusician), Condition (two levels: silence, MTB), Vowel duration (standardized), F0 (two levels: 120 or 140 Hz), and Spectral composition (preserved cues from the vowel-coda of "bead" or "beat"). We additionally tested the three-way interactions between Group, Condition, and the three acoustic cues manipulated. Random effects included by-Subject random intercepts and by-Subject random slopes for Duration. Lme4 syntax is: Group*Condition(Word + Dur + f0) + (Condition | Subject). Effects were treatment coded (ref=Group:Nonmusicians, Condition: Silence, F0: 140, Spectrum: "beat").

3. Results

The logistic regression model revealed significant main effects for Group, where musicians showed fewer "bead" responses than nonmusicians (p<0.001), and for two of the acoustic cues of interest: preserved spectral information signaling "bead" and increased vowel duration were significant (p<0.001) predictors of [+voice] responses. There was a trend toward significance for a lower f0 (f_0 = 120 Hz) in increasing voiced judgment (p=0.068). No main effect of Condition was observed.

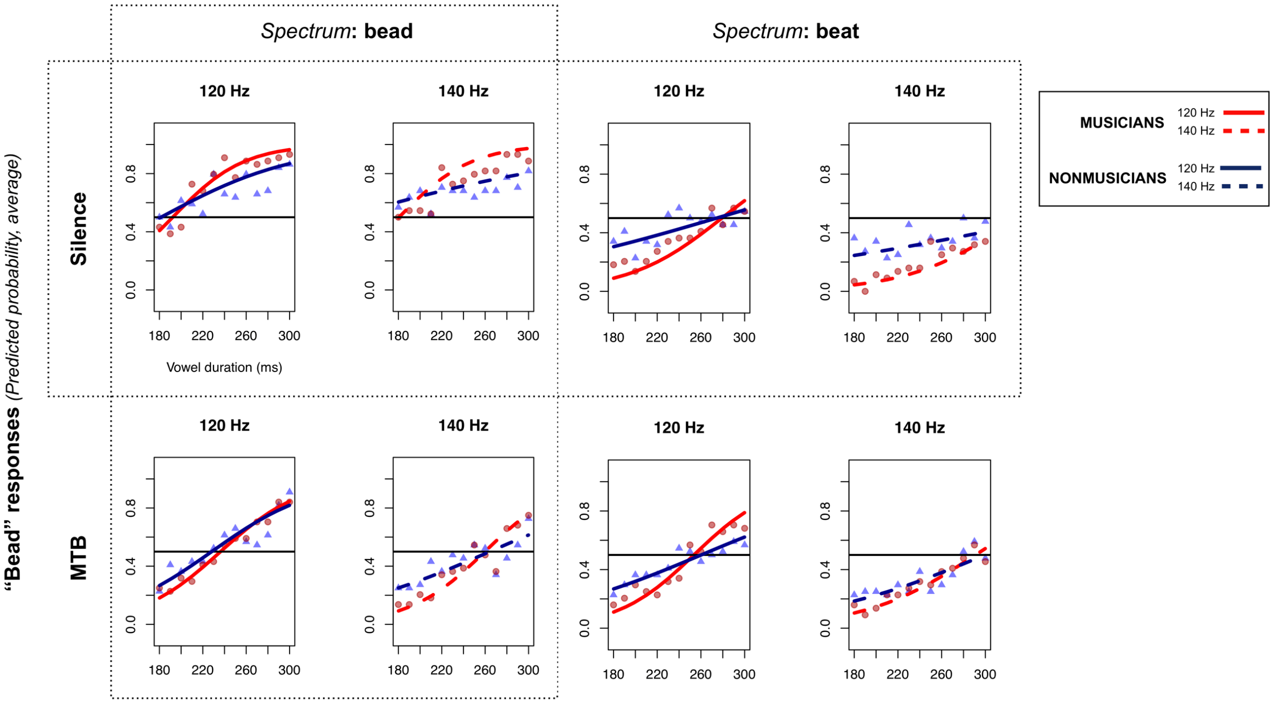

Figure 1: *Mean raw data for "bead" judgements (0 = "beat", 1 = "bead") at each duration step for musicians (circle) and nonmusicians (triangle). Logistic functions showing the predicted probability from the model are overlaid for silence (top row) and multitalker babble (MTB, bottom row).*

The model also showed two-way interactions between Condition and the three acoustic cues: in MTB, listeners made more "bead" judgments on the basis of increasing vowel duration (p<0.001) and a lower f0 (p<0.05), compared with no background noise (i.e., in "silence"). Additionally, listeners made significantly *fewer* voiced judgments on the basis of preserved spectral cues (p<0.001) in MTB than in silence (Table 1).

Table 1: *Model outputs for "bead" responses*

	Coef	*SE*	*z*	*p*	
(Intercept)	-0.53	0.11	-4.83	<0.001	***
$Condition_{MTB}$	-0.20	0.13	-1.59	0.113	
$Group_{MUS}$	-0.84	0.16	-5.14	<0.001	***
Dur	0.30	0.05	6.70	<0.001	***
$F0_{120}$	0.16	0.09	1.83	0.068	
$Spect_{BEAD}$	1.18	0.09	13.17	<0.001	***
$Condition_{MTB}$ *$Group_{MUS}$	0.63	0.19	3.32	<0.001	***
$Condition_{MTB}$*Dur	0.21	0.06	3.25	<0.01	**
$Condition_{MTB}$*F0	0.32	0.13	2.53	0.011	*
$Condition_{MTB}$* $Spect_{BEAD}$	-0.73	0.13	-5.81	<0.001	***
$Group_{MUS}$*Dur	0.48	0.07	6.82	<0.001	***
$Group_{MUS}$*$F0_{120}$	0.30	0.14	2.22	0.025	*
$Group_{MUS}$* $Spect_{BEAD}$	1.19	0.14	8.40	<0.001	***
$Condition_{MTB}$*$Group_{MUS}$*Dur	-0.14	0.10	-1.38	0.169	
$Condition_{MTB}$ *$Group_{MUS}$*F0	-0.18	0.18	-0.97	0.33	
$Condition_{MTB}$ *$Group_{MUS}$* $Spect_{BEAD}$	-1.19	0.18	-6.21	<0.001	***

(*** p<0.001, ** p<0.01, * p<0.05), Observations = 9152

Furthermore, we observed two-way interactions between Group and all three acoustic cues of interest: musicians displayed significantly more "bead" responses based on the preserved spectral cues in the vowel (p<0.001), for increasing duration (p<0.001), and for a lower f0 (p<0.05), relative to nonmusicians (see Figure 1). Additionally, we observed a three-way interaction between Condition, Group, and Spectral cues (p<0.001): while musicians show more "bead" responses on the basis of spectral cues, they do so less in MTB than in silence, relative to nonmusicians. No other three-way interactions were significant in the model.

3.1. Cue weighting

The relative weight of each cue can be observed in the beta coefficients from the logistic regression model for the main effects (e.g., f0, spectrum, duration) adjusted by the interaction coefficients of the cues by Group and Condition (e.g., F0 + $Group_{MUS}$*F0 + $Cond_{MTB}$*F0 + $Cond_{MTB}$*$Group_{MUS}$*F0; recall that the contrasts were treatment coded). These coefficients are plotted in Figure 2.

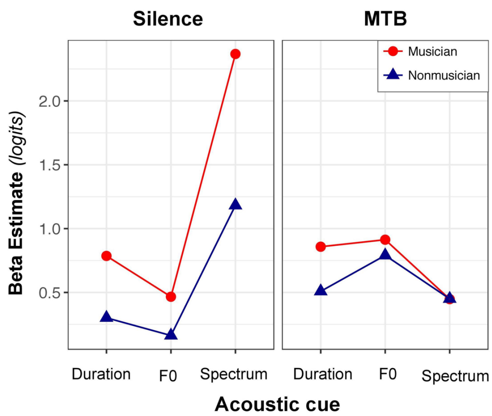

Figure 2: *Beta estimates from the logistic regression models for each Group in silence and multitalker babble.*

We find that the groups show some similarities and some differences in their cue weighting strategies within and across noise conditions. In Silence, musicians weigh all three cues more heavily (spectrum, duration, f0); though the cue weighting pattern is similar for nonmusicians (spectrum > duration > f0). In MTB, we see that both group's cues are re-ranked, wherein duration and f0 are stronger cues than the spectrum.

4. Discussion

In this study, we tested whether musical training influences listeners' perceptual cue weighting and categorical perception when making a phoneme judgment in their native language. We examined musicians' and nonmusicians' cue weighting strategies for duration, vowel f0, and spectral cues for the post-vocalic /t-d/ distinction (e.g., "beat" vs. "bead") in silence and in multitalker babble (MTB). For both musicians and nonmusicians, we found that longer vowel duration and preserved spectral cues (e.g., formant transitions, voicing duration closure) led to a [+voice] response for both groups, consistent with past work, e.g. [33]–[35], with a weak effect of lower vowel f0.

Listeners' sensitivity to each of these cues, however, systematically differed based on their musical background. Musicians displayed reliably *steeper* categorization functions for vowel duration, f0, and spectral information, relative to nonmusicians (see Figure 1). This is in line with studies that have reported musicians' enhanced perceptual acuity for each of these cues, relative to nonmusicians [3]–[5]. One interpretation is that musicians may be better able to use fine-grained differences along an acoustic dimension to classify stimuli, an ability that may be especially honed via musical experience, in line with bottom-up accounts, e.g. [25]–[27]. However, that musicians show greater sensitivity to spectral and duration cues counters predictions made by the OPERA theoretical framework that acoustic properties that require greater "precision" for music (cf. [29]) will show cross-domain transfer to speech perception [6], [28], which would only predict differences for f0.

We also found differences in cue weighting across noise conditions for both groups. Consistent with other studies [21], [22], we observed a shift from stronger weighting of the spectrum for [+voice] judgments in silence toward greater weighting of duration and f0 in multitalker babble (see Table 1 and Figure 2). This re-ranking is even more pronounced for musicians, who demonstrated the largest difference in spectrum weighting across the noise conditions. That musicians cue into duration and f0 more strongly in MTB suggests that they might more strongly encode the *relevant* or more *reliable* acoustic cues available, particularly as MTB has been shown to more heavily mask the spectrum [22]. One interpretation for these patterns is that musical training may sharpen attentional mechanisms, which allows musicians to better hone in on phonemically contrastive cues [30], supporting top-down accounts of cross-domain plasticity from music to speech perception [30], [31]. Still, the lack of a different cue weighting schema between the groups suggests that overall cue weighting strategies are largely driven by listeners' language, rather than musical, experience.

In general, our findings support theories that musical training transfers to speech perception. Musicians displayed sharper categorization functions for duration, f0, and the spectrum while making phoneme judgments, suggesting they may have more precise mappings of acoustic phonetic properties to speech sound categories, in line with [3]–[5]. Furthermore, in difficult listening conditions, musicians show more pronounced re-ranking, relative to nonmusicians.

Still, there are remaining questions and avenues for future research. While we hypothesize that these categorization differences will generalize to other obstruent final contrasts, additional study is needed to test other word pairs, including items that do not vary as greatly for the subject groups; for example, the target word "beat" may be a more frequent term for musicians than nonmusicians. Furthermore, it is important to note that the paradigm used in this study was an identification paradigm, which forces a binary judgment; one question is whether musicians' and nonmusicians' responses would vary if given a continuous, sliding scale to respond. This could speak to whether the groups differ in the gradience of their sound categories at a more fine-grained level. Additionally, as our task did not provide feedback, listeners did not have a target acoustic cue to monitor for. Future studies could use tasks with feedback test the groups' flexibility in cue reweighting over the course of an experiment; along with paradigms that test discrimination abilities, these types of experiments could tease apart the contribution of bottom-up sensory encoding and top-down attentional processing differences for those with and without musical training.

5. Conclusions

Overall, this work suggests that musical training may impact listeners' cue weighting strategies for a phonemic contrast in their native language, a novel finding. Our findings shed light on the nature of music-to-speech transfer, with support for both bottom-up theories of acoustic encoding as well as for top-down theories of attention, but countering targeted-feature accounts that limit transfer to particular properties (e.g., f0). Finally, that we see systematic variation among listeners with the same language background further points to both individual differences and plasticity in the perceptual organization of phonemes.

6. Acknowledgments

Thank you to Antoine Shahin for advice and support on the experiment.

7. References

[1] T. F. Münte, E. Altenmüller, and L. Jäncke, "The musician's brain as a model of neuroplasticity," *Nat. Rev. Neurosci.*, vol. 3, no. 6, p. 473, 2002.

[2] A. Parbery-Clark, E. Skoe, C. Lam, and N. Kraus, "Musician Enhancement for Speech-In-Noise," *Ear Hear.*, vol. 30, no. 6, p. 653, Dec. 2009.

[3] D. Schön, C. Magne, and M. Besson, "The music of speech: Music training facilitates pitch processing in both music and language," *Psychophysiology*, vol. 41, no. 3, pp. 341–349, May 2004.

[4] J. Chobert, C. François, J.-L. Velay, and M. Besson, "Twelve months of active musical training in 8-to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time," *Cereb. Cortex*, vol. 24, no. 4, pp. 956–967, 2012.

[5] G. M. Bidelman and C. Alain, "Musical Training Orchestrates Coordinated Neuroplasticity in Auditory Brainstem and Cortex to Counteract Age-Related Declines in Categorical Vowel Perception," *J. Neurosci.*, vol. 35, no. 3, pp. 1240–1249, Jan. 2015.

[6] A. D. Patel, “Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis,” *Hear. Res.*, vol. 308, pp. 98–108, 2014.
[7] I. Peretz and M. Coltheart, “Modularity of music processing,” *Nat. Neurosci.*, vol. 6, no. 7, p. 688, 2003.
[8] E. Fedorenko, M. K. Behr, and N. Kanwisher, “Functional specificity for high-level linguistic processing in the human brain,” *Proc. Natl. Acad. Sci.*, vol. 108, no. 39, pp. 16428–16433, Sep. 2011.
[9] A. Parbery-Clark, D. L. Strait, S. Anderson, E. Hittner, and N. Kraus, “Musical Experience and the Aging Auditory System: Implications for Cognitive Abilities and Hearing Speech in Noise,” *PLOS ONE*, vol. 6, no. 5, p. e18082, May 2011.
[10] D. L. Strait, A. Parbery-Clark, S. O’Connell, and N. Kraus, “Biological impact of preschool music classes on processing speech in noise,” *Dev. Cogn. Neurosci.*, vol. 6, pp. 51–60, Oct. 2013.
[11] K. Meha-Bettison, M. Sharma, R. K. Ibrahim, and P. R. M. Vasuki, “Enhanced speech perception in noise and cortical auditory evoked potentials in professional musicians,” *Int. J. Audiol.*, vol. 57, no. 1, pp. 40–52, Jan. 2018.
[12] B. R. Zendel, C.-D. Tremblay, S. Belleville, and I. Peretz, “The impact of musicianship on the cortical mechanisms related to separating speech from background noise,” *J. Cogn. Neurosci.*, vol. 27, no. 5, pp. 1044–1059, 2015.
[13] D. R. Ruggles, R. L. Freyman, and A. J. Oxenham, “Influence of Musical Training on Understanding Voiced and Whispered Speech in Noise,” *PLOS ONE*, vol. 9, no. 1, p. e86980, Jan. 2014.
[14] D. Boebinger, S. Evans, S. Rosen, C. F. Lima, T. Manly, and S. K. Scott, “Musicians and non-musicians are equally adept at perceiving masked speech,” *J. Acoust. Soc. Am.*, vol. 137, no. 1, pp. 378–387, Jan. 2015.
[15] S. M. K. Madsen, K. L. Whiteford, and A. J. Oxenham, “Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds,” *Sci. Rep.*, vol. 7, no. 1, p. 12624, Oct. 2017.
[16] E. M. Anaya, D. B. Pisoni, and W. G. Kronenberger, “Long-term musical experience and auditory and visual perceptual abilities under adverse conditions,” *J. Acoust. Soc. Am.*, vol. 140, no. 3, pp. 2074–2081, 2016.
[17] I. Yeend, E. F. Beach, M. Sharma, and H. Dillon, “The effects of noise exposure and musical training on suprathreshold auditory processing and speech perception in noise,” *Hear. Res.*, vol. 353, pp. 224–236, Sep. 2017.
[18] L. Lisker, “‘Voicing’ in English: A catalogue of acoustic features signaling/b/versus/p/in trochees,” *Lang. Speech*, vol. 29, no. 1, pp. 3–11, 1986.
[19] P. Escudero, T. Benders, and S. C. Lipski, “Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners,” *J. Phon.*, vol. 37, no. 4, pp. 452–465, Oct. 2009.
[20] P. Iverson *et al.*, “A perceptual interference account of acquisition difficulties for non-native phonemes,” *Cognition*, vol. 87, no. 1, pp. B47–B57, 2003.
[21] C. Wardrip–Fruin, “On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants,” *J. Acoust. Soc. Am.*, vol. 71, no. 1, pp. 187–195, 1982.
[22] C. Wardrip-Fruin, “The effect of signal degradation on the status of cues to voicing in utterance-final stop consonants,” *J. Acoust. Soc. Am.*, vol. 77, no. 5, pp. 1907–1912, 1985.
[23] V. Hazan and S. Rosen, “Individual variability in the perception of cues to place contrasts in initial stops,” *Percept. Psychophys.*, vol. 49, no. 2, pp. 187–200, Mar. 1991.
[24] J. Schertz, T. Cho, A. Lotto, and N. Warner, “Individual differences in phonetic cue use in production and perception of a non-native sound contrast,” *J. Phon.*, vol. 52, pp. 183–204, 2015.
[25] A. Shahin, L. E. Roberts, C. Pantev, L. J. Trainor, and B. Ross, “Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds,” *Neuroreport*, vol. 16, no. 16, pp. 1781–1785, 2005.
[26] G. M. Bidelman, M. W. Weiss, S. Moreno, and C. Alain, “Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians,” *Eur. J. Neurosci.*, vol. 40, no. 4, pp. 2662–2673, 2014.
[27] A. J. Shahin, “Neurophysiological Influence of Musical Training on Speech Perception,” *Front. Psychol.*, vol. 2, 2011.
[28] A. D. Patel, “Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis,” *Front. Psychol.*, vol. 2, 2011.
[29] R. J. Zatorre, P. Belin, and V. B. Penhune, “Structure and function of auditory cortex: music and speech,” *Trends Cogn. Sci.*, vol. 6, no. 1, pp. 37–46, 2002.
[30] D. L. Strait and N. Kraus, “Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise,” *Front. Psychol.*, vol. 2, p. 113, 2011.
[31] N. Kraus and T. White-Schwoch, “Unraveling the biology of auditory learning: a cognitive–sensorimotor–reward framework,” *Trends Cogn. Sci.*, vol. 19, no. 11, pp. 642–654, 2015.
[32] C. D. Fuller, J. J. Galvin, R. H. Free, and D. Başkent, “Musician effect in cochlear implant simulated gender categorization,” *J. Acoust. Soc. Am.*, vol. 135, no. 3, pp. EL159–EL165, Feb. 2014.
[33] M. Chen, “Vowel length variation as a function of the voicing of the consonant environment,” *Phonetica*, vol. 22, no. 3, pp. 129–159, 1970.
[34] W. A. Castleman and R. L. Diehl, “Effects of fundamental frequency on medial and final [voice] judgments,” *J. Phon.*, vol. 24, no. 4, pp. 383–398, 1996.
[35] C. S. Crowther and V. Mann, “Native language factors affecting use of vocalic cues to final consonant voicing in English,” *J. Acoust. Soc. Am.*, vol. 92, no. 2, pp. 711–722, 1992.
[36] B. H. Repp, “Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants,” *Lang. Speech*, vol. 22, no. 2, pp. 173–189, 1979.
[37] J. Reilly, V. Troiani, M. Grossman, and R. Wingfield, “An introduction to hearing loss and screening procedures for behavioral research,” *Behav. Res. Methods*, vol. 39, no. 3, pp. 667–672, Aug. 2007.
[38] P. Boersma and D. Weenink, *Praat: doing phonetics by computer*. 2018.
[39] R. Corretge, “Praat vocal toolkit: A praat plugin with automated scripts for voice processing,” Retrieved 2016-09-05, from http://www. praatvocaltoolkit. com, 2012.
[40] S. Messaoud-Galusi, V. Hazan, and S. Rosen, “Clarifying the speech perception deficit in dyslexic children,” in *Proceedings of the 16th International Congress of Phonetic Sciences*, 2007, pp. 697–700.
[41] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting Linear Mixed-Effects Models Using lme4,” *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, Oct. 2015.