Auditory categorical processing for speech is modulated by inherent musical listening skills

Kelsey Mankel[a,b], Jacob Barber[a] and Gavin M. Bidelman[a,b,c]

During successful auditory perception, the human brain classifies diverse acoustic information into meaningful groupings, a process known as categorical perception (CP). Intense auditory experiences (e.g., musical training and language expertise) shape categorical representations necessary for speech identification and novel sound-to-meaning learning, but little is known concerning the role of innate auditory function in CP. Here, we tested whether listeners vary in their intrinsic abilities to categorize complex sounds and individual differences in the underlying auditory brain mechanisms. To this end, we recorded EEGs in individuals without formal music training but who differed in their inherent auditory perceptual abilities (i.e., musicality) as they rapidly categorized sounds along a speech vowel continuum. Behaviorally, individuals with naturally more adept listening skills ("musical sleepers") showed enhanced speech categorization in the form of faster identification. At the neural level, inverse modeling parsed EEG data into different sources to evaluate the contribution of region-specific activity [i.e., auditory cortex (AC)] to categorical neural coding. We found stronger categorical processing in musical sleepers around the timeframe of P2 (~180 ms) in the right AC compared to those with poorer musical listening abilities. Our data show that listeners with naturally more adept auditory skills map sound to meaning more efficiently than their peers, which may aid novel sound learning related to language and music acquisition. *NeuroReport* 31: 162–166

NeuroReport 2020, **31**:162–166

Keywords: auditory event-related potentials, categorical perception, electroencephalography, individual differences, Profile of Music Perception Skills

[a]School of Communication Sciences and Disorders; [b]Institute for Intelligent Systems, University of Memphis and [c]University of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis, Tennessee, USA

Correspondence to Gavin Bidelman, School of Communication Sciences and Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA
Tel: +1 901 678 5826; fax: +1 901 678 2472; e-mail: gmbdlman@memphis.edu

Received 3 October 2019 **Accepted** 11 October 2019

Introduction

Music training is associated with enhanced auditory processing, including the categorical perception (CP) of speech [1–4] and music stimuli [5]. Musicianship has been attributed to sharper speech identification [1,4], faster labeling speeds [1,2], and higher perceptual sensitivities for within- vs. between-category contrasts [3]. Whether these speech advantages truly reflect experience (training) or other predispositions remains unclear. Musicians' improved speech processing may instead reflect increased motivation [6], enhanced auditory attention and cognitive advantages [7], or even innately superior auditory abilities (e.g., musicality) that exist without formal music training [8,9].

While musicality is a multidimensional trait, we focus on abilities pertaining to complex listening, following a long tradition of describing musical aptitude via receptive skills [8,10,11]. Recently, we demonstrated even nonmusicians vary in innate music abilities and people with exceptionally high levels of musicality show stronger neural processing of speech [8], a finding typically attributed to trained musicians [1,12]. Positive associations between phoneme discrimination and musicality (but not music training itself) have been observed even after controlling for cognitive and socioeconomic status (SES) [9]. Irrespective of formal training, music aptitude also predicts stronger CP for musical chords [5]. Such findings highlight the need to distinguish innate from experience-dependent characteristics for defining neuroplasticity. Presumably, individual differences in auditory processing may at least partially drive sound categorization skills, which would temper assumptions that experience, per se, drives musicians' enhancements in speech processing [8,9].

Neural evidence suggests CP emerges in the brain by 100–200 ms, in the timeframe of the N1-P2 event-related potentials (ERPs) [4,13,14]. Enhanced N1-P2 amplitudes, for example, have been linked with stronger identification [15] and faster labeling speeds during speech categorization [1,2]. Categorical effects have also been reported for the later N2 and P3 waves, whose amplitudes differentiate stimuli at the categorical boundary and between native vs. nonnative phonetic contrasts [14]. Neuroimaging studies have further implicated several brain regions in auditory categorical processing, including the superior temporal gyrus/sulcus, middle temporal gyrus, premotor cortex, supramarginal gyrus, inferior parietal cortex, planum temporale, and inferior frontal gyrus [4,16]. These neural signatures of CP are also associated with formal music training [1,2,4,14,17]. While long-term auditory

0959-4965 DOI: 10.1097/WNR.0000000000001369

experience might tune categorical speech processing, it is unclear (1) how much CP is susceptible to individual differences and (2) whether musicians' enhancements in this process might be partially due to innate auditory sensitivities irrespective of formal experience.

Here, we tested whether individual differences in musicality – in the absence of formal music training – affect speech categorization. To this end, we measured EEGs in nonmusicians who varied in their musical listening abilities as they classified vowels along an acoustic-phonetic continuum. If individual differences in auditory perceptual skills drive CP, we predicted that individuals with higher levels of innate musicality ("musical sleepers") [8,11] would show stronger behavioral CP and neural responses to speech. Our data confirm that musical sleepers categorize speech more efficiently, as evidenced by stronger categorical coding compared to those with poorer music aptitude.

Methods

Participants

The sample included $N = 14$ young adults (seven females; age: $\mu \pm \sigma = 24.9 \pm 1.7$ years). This sample size is comparable to studies assessing musicians and nonmusicians [1,4] and allowed us to test the premise that such variations in CP might result from differences in inherent musicality rather than formal training. Participants were right-handed, had normal hearing (thresholds ≤ 25 dB HL, 250–8000 Hz), no tone language experience or neurological disorders, and a collegiate level of education (17.3 ± 3.00 years). Average parental education, a common measure of SES [highest parental education: 1 (high school without diploma or GED) to 6 (doctoral degree)], was 4.14 ± 0.57 indicating a bachelor's level education [18]. To isolate effects of inherent music listening abilities on auditory processing [8], participants were required to have <3 years total of formal music training (0.57 ± 0.76 years) and no music experience within the past 5 years. All gave written consent according to a protocol approved by the UofM Institutional Review Board.

Behavioral test of musicality

The brief Profile of Music Perception Skills (PROMS) assessed aptitude related to receptive musical skills [11]. This test comprises same-different tasks of melody, tuning, accent, and tempo discrimination. Each subtest contains 18 trials, and higher scores indicate stronger musical listening abilities (max score = 72). The PROMS can differentiate listeners based on their musical experience (professional vs. amateur musicians vs. nonmusicians), and is sensitive to detect untapped musical potential among nonmusicians ('musical sleepers') [8,11]. For details see [11].

Stimuli

We used a synthetic five-step vowel continuum to assess CP for speech (see Fig. 1 in ref. [1]). Tokens were 100 ms. Fundamental (F0), second (F2), and third formant (F3) frequencies were identical across tokens (F0: 100 Hz; F2: 1090 Hz; and F3: 2350 Hz). First formant (F1) was varied across five equidistant steps (430–730 Hz), yielding a perceptual continuum from /u/ to /a/.

Behavioral data

The task was identical to our recent reports [1,13]. Stimuli were delivered binaurally through ER-2 earphones at 82 dB SPL. Listeners heard 200 trials of each speech token (random order) and were instructed to label them as 'u' or 'a' via the keyboard as fast and accurately as possible. The interstimulus interval was 400–600 ms (20 ms steps; jittered).

Individual identification curves were fit with a sigmoid: $P = 1/[1 + e^{-\beta 1(x-\beta 0)}]$, where P is the proportion of trials labeled as /a/, x is the token number, and β0 and β1 are the location and slope parameters of the psychometric fit. Large β1 values indicate steeper slopes and stronger CP. Response times (RTs) were computed as the average labeling speed per token. RTs outside 250–2500 ms were excluded as fast guesses or lapses in attention [13,15].

EEG recordings

EEGs were recorded from 64 electrodes at 10-10 scalp locations (500 Hz sample rate) (Synamps RT amplifiers, Neuroscan). Impedances were <10 kΩ. Preprocessing was performed in BESA Research (v7) (BESA, GmbH). Blinks were nullified in the continuous data via spatial filtering [10]. Trials with voltages ≥120 μV were discarded from averaging. Recordings were epoched (-200 to 800 ms), baseline corrected, filtered from 1 to 30 Hz, and averaged across trials to compute ERPs for each speech token per listener.

Source analysis

We transformed raw ERPs to source space using BESA's AEP virtual source montage [19]. We then extracted source waveforms (units nAm) from the radially-oriented dipoles in left and right auditory cortices (AC) to assess categorical speech coding (Talairach coordinates of [-37, -18, 17] and [37, -18, 17], respectively) [e.g., 4,20]. Neural correlates of CP emerge around the N1 and P2 waves [2,13,20]. Thus, N1 was measured as the peak negativity between 100 and 160 ms; P2 as the peak positivity between 160 and 220 ms. Categorical coding in the ERPs was evaluated as the difference between P2 responses to prototypical tokens (endpoints) and the ambiguous midpoint [i.e., mean (Tk1, Tk5) – Tk3; ΔP2] [4,15]; more negative ΔP2 latencies mirror categorical processing whereby /u/ and /a/ exemplars elicit faster latencies than stimuli at the ambiguous midpoint of the continuum (i.e., Tk1/5 < Tk3).

Results

Behavioral data

Listeners were highly separable based on a median split of their PROMS scores, revealing some nonmusicians have inherently better music listening skills (Fig. 1a) [8]. Low and high musicality groups did not differ in age, handedness, years of musical training, years of education, or SES (all *P*s > 0.05). The high PROMS group demonstrated superior performance not only on the PROMS total score (Wilcoxon; normal approximation, one sided: Z = 1.79, P = 0.001) but also the tempo (Z = 1.79, P = 0.0363) and tuning subtests (Z = 2.78, P = 0.0027).

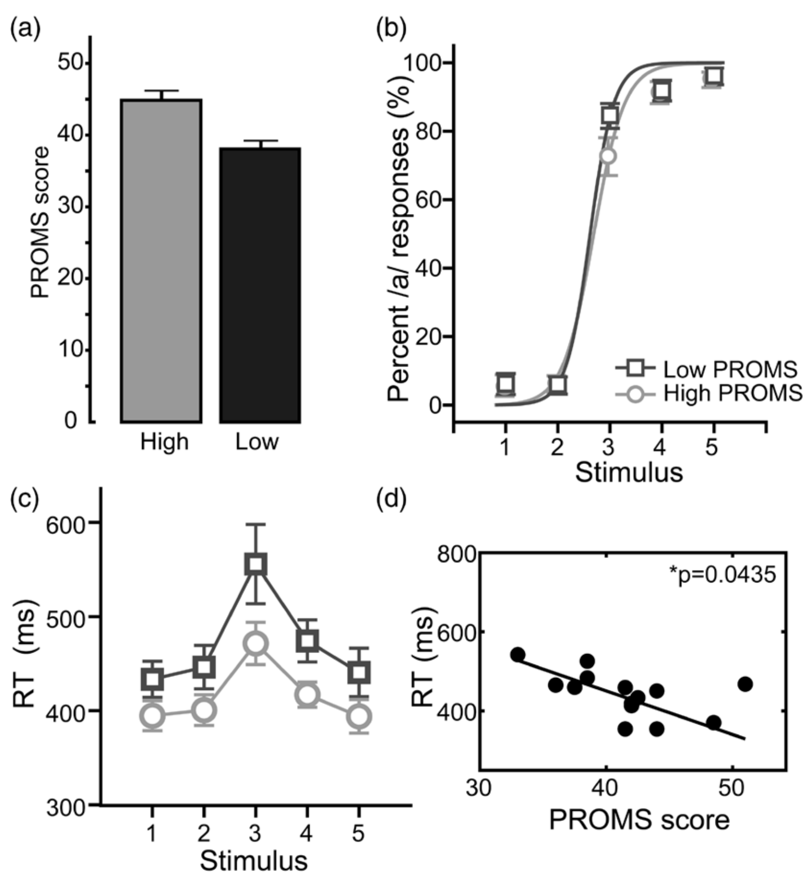
Groups did not differ in their slopes of speech identification (Z = −0.9594, P = 0.3374; Fig. 1b) but the high PROMS group showed marginally faster RTs across the continuum ($F_{1,52}$ = 3.87, P = 0.0544, Cohen's d = 1.136; Fig. 1c). Higher PROMS scores were also associated with faster RTs (r = −0.55, P = 0.0435; Fig. 1d). Thus, while speech CP was not stronger in one group over the other, per se, faster RTs suggest individuals with better music listening skills were more efficient at making speech categorization judgments.

Source event-related potentials

Figure 2 shows scalp topographies and source waveforms from left/right AC. Two-way, mixed model ANOVAs evaluated group × hemisphere effects on ERP measures (random factor = subjects nested within group). Pooled across tokens, we found sole hemispheric difference in N1 latencies (LH > RH; $F_{1,128}$ = 7.15, P = 0.0085, d = 1.54). Subsequent analysis focused on changes in P2 (i.e., ΔP2) between phonetic (Tk1/5) and nonphonetic (Tk3) tokens as a marker of speech identification and categorical processing [1,2,4,13,15].

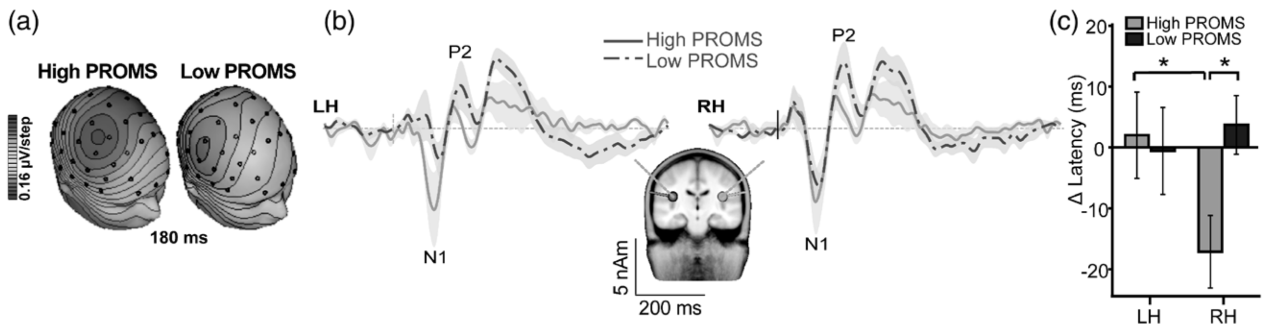
ΔP2 latencies showed a group × hemisphere interaction ($F_{1,12}$ = 6.79, P = 0.023, d = 1.50; Fig. 2c). Tukey-Kramer contrasts revealed the high PROMS group had more negative (i.e., more categorical) ΔP2 latencies in the RH than the low PROMS group (t_{12} = −2.33, P = 0.038). Stronger speech differentiation, as indexed by ΔP2 latencies, was also observed in RH vs. LH specifically in the high PROMS group (RH < LH; t_{12} = 3.01, P = 0.0108).

Fig. 1



Behavioral data. (a) PROMS scores reveal some individuals have naturally more adept auditory perceptual skills in the absence of any formal music training (median split of into high- and low-musicality groups). High and low PROMS groups did not differ in identification (b), but highly musical listeners were more efficient (faster) at identifying categorical speech stimuli (c). (d) Better musical perceptual skills correlated with faster average RTs. *$P < 0.05$, error bars ± 1 SEM. PROMS, Profile of Music Perception Skills.

Fig. 2



Neural data. ERPs reflect categorical coding around the timeframe of P2. (a) Scalp topographies (180 ms) and (b) grand average auditory cortical source waveforms by PROMS groups. (c) Phonetic (Tk1/5) vs. nonphonetic (Tk3) contrast of P2 (i.e., ΔP2) assessed categorical processing, which was stronger for individuals with better musical listening abilities. ΔP2 latencies were more categorical (i.e., negative) in the right hemisphere for the high PROMS group. $^*P < 0.05$, error bars/shading ± 1 SEM. ERPs, event-related potentials; PROMS, Profile of Music Perception Skills.

No other contrasts including amplitude measures were significant.

Discussion

Our findings show 'musical sleepers' who exhibit naturally higher levels of musicality (but are nevertheless nonmusicians) are more efficient at categorizing speech sounds (i.e., faster RTs). Neural responses mimicked behavioral benefits in that high PROMS listeners showed more categorical ERPs within right AC. Our results provide novel evidence that listeners' efficiency of categorical processing at both behavioral and neural levels varies with their inherent auditory perceptual skills.

Critically, our data reveal natural auditory sensitivities, in the absence of music training, are associated with improvements in auditory categorical processing. While both groups demonstrated categorical speech processing – sharp transitions at the perceptual boundary and delayed RTs for ambiguous relative to prototypical tokens [21,22] – those with higher inherent musicality exhibited faster RTs. Interestingly, the faster RTs observed in our high music aptitude listeners mirrors results in formally trained musicians, who also show faster speech categorization [1,2,14]. While faster RTs in the high PROMS group may reflect faster initiation/execution of motor responses, group differences in neural activity occurred as early as P2 (i.e., well before RTs). This implies precognitive, premotor brain activity (i.e., sensory coding) contributes to improved categorization efficiency in high aptitude nonmusicians.

That RTs but not identification differed between groups suggests inherent auditory skills improve the decision processes and/or speed of access to internalized categories rather than a sharpening of those phonetic representations, per se [21,22]. A lack of difference in identification may have been expected given that our listeners were native speakers, highly familiar with English vowel contrasts [2,14,23]. In comparison, categorical identification (in addition to RTs) is more acute in actual trained musicians [cf. 1,4]. Collectively, this and other studies lead us to infer that long-term music experience improves sound classification accuracy and speeded access to speech representations above and beyond natural or inherent abilities observed here.

At the neural level, we found stronger categorical neural encoding of speech in people with better music perceptual skills. More categorical responses were observed in the right AC of the high compared to the low PROMS group. Our data here (all nonmusicians) parallel recent findings in highly experienced musicians, which report plasticity in AC activity is associated with more categorical coding of sound [4]. Previously, we asserted musicality-related differences in auditory CP require (1) strong experiential plasticity rather than subtle innate function, or (2) tasks requiring top-down processing and/or attention [8]. Results of this study extend previous findings by demonstrating both behavioral and ERP enhancements in musical sleepers during active speech identification. Results here show that certain individuals exhibit naturally superior auditory system function, specifically in the right AC and timeframe of P2 (~160–200 ms), which contributes to enhanced speech perception.

Given the known LH dominance of language processing, it is surprising that categorical neural responses to speech were more apparent in right AC. One possibility is that highly musical individuals recruit additional resources in RH for speech processing, which could contribute to faster ERP latencies and RTs seen here. Indeed, greater training-related plasticity is observed in right compared to left AC [24]. However, previous cross-sectional studies do not isolate innate differences of brain structure and function before training. Although the role of 'native abilities' has been acknowledged in studies on musicianship and auditory perceptual abilities [23], they have never been fully tested to the extent here [cf. 8].

While we do not refute the existence of experience-dependent plasticity of musicianship [8,25], the "musical brain" is likely an interplay between predispositions, environmental factors, and training [8]. Our data reveal inherent differences in brain function can contribute to more efficient and robust categorical speech processing. Thus, inherent auditory skills may at least partially contribute to experience-dependent neuroplastic effects reported in studies on musical training and speech-language function [1,4]. Other studies have similarly shown associations between speech discrimination and musicality (i.e., rhythm perception) but not music training itself [9]. While music-based interventions for communication disorders are promising [25], care should be taken to understand individuals' experience and inherent auditory function to maximize learning or therapeutic benefits. We add to the growing body of evidence that some individuals have naturally superior auditory abilities, which enables better neural processing and perception of speech [8,9].

Acknowledgements

The authors thank G. Lewis for her comments on earlier versions of this article. Requests for materials should be directed to G.M.B. (gmbdlman@memphis.edu).

This study was supported by NIH/NIDCD grant (R01DC016267).

Conflicts of interest

There are no conflicts of interest.

References

1 Bidelman GM, Weiss MW, Moreno S, Alain C. Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur J Neurosci* 2014; **40**:2662–2673.
2 Bidelman GM, Alain C. Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J Neurosci* 2015; **35**:1240–1249.
3 Wu H, Ma X, Zhang L, Liu Y, Zhang Y, Shu H. Musical experience modulates categorical perception of lexical tones in native Chinese speakers. *Front Psychol* 2015; **6**:436.
4 Bidelman GM, Walker B. Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network. *Neuroimage* 2019; **201**:116022.
5 Howard D, Rosen S, Broad V. Major/Minor triad identification and discrimination by musically trained and untrained listeners. *Music Percept* 1992; **10**:205–220.
6 McAuley JD, Henry MJ, Wedd A, Pleskac TJ, Cesario J. Effects of musicality and motivational orientation on auditory category learning: a test of a regulatory-fit hypothesis. *Mem Cognit* 2012; **40**:231–251.
7 Yoo J, Bidelman GM. Linguistic, perceptual, and cognitive factors underlying musicians' benefits in noise-degraded speech perception. *Hear Res* 2019; **377**:189–195.
8 Mankel K, Bidelman GM. Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proc Natl Acad Sci U S A* 2018; **115**:13129–13134.
9 Swaminathan S, Schellenberg EG. Musical competence and phoneme perception in a foreign language. *Psychon Bull Rev* 2017; **24**:1929–1934.
10 Gordon EE. *Musical aptitude profile manual*. Boston, MA: Houghton Mifflin; 1965.
11 Law LN, Zentner M. Assessing musical abilities objectively: construction and validation of the profile of music perception skills. *Plos One* 2012; **7**:e52508.
12 Parbery-Clark A, Skoe E, Kraus N. Musical experience limits the degradative effects of background noise on the neural processing of sound. *J Neurosci* 2009; **29**:14100–14107.
13 Bidelman GM, Moreno S, Alain C. Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage* 2013; **79**:201–212.
14 Dittinger E, D'Imperio M, Besson M. Enhanced neural and behavioural processing of a nonnative phonemic contrast in professional musicians. *Eur J Neurosci* 2018; **47**:1504–1516.
15 Bidelman GM, Walker BS. Attentional modulation and domain-specificity underlying the neural organization of auditory categorical perception. *Eur J Neurosci* 2017; **45**:690–699.
16 Guenther FH, Nieto-Castanon A, Ghosh SS, Tourville JA. Representation of sound categories in auditory cortical maps. *J Speech Lang Hear Res* 2004; **47**:46–57.
17 Elmer S, Meyer M, Jäncke L. Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cereb Cortex* 2012; **22**:650–658.
18 Entwislea DR, Astone NM. Some practical guidelines for measuring youth's race/ethnicity and socioeconomic status. *Child Dev* 1994; **65**:1541–1545.
19 Scherg M, Von Cramon D. Evoked dipole source potentials of the human auditory cortex. *Electroencephalogr Clin Neurophysiol* 1986; **65**:344–360.
20 Bidelman GM, Lee CC. Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *Neuroimage* 2015; **120**:191–200.
21 Pisoni DB, Tash J. Reaction times to comparisons within and across phonetic categories. *Percept Psychophys* 1974; **15**:285–290.
22 Bidelman GM, Sigley L, Lewis GA. Acoustic noise and vision differentially warp the auditory categorization of speech. *J Acoust Soc Am* 2019; **146**:60.
23 Foster NE, Zatorre RJ. A role for the intraparietal sulcus in transforming musical pitch information. *Cereb Cortex* 2010; **20**:1350–1359.
24 Bermudez P, Lerch JP, Evans AC, Zatorre RJ. Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb Cortex* 2009; **19**:1583–1596.
25 Habib M, Lardy C, Desiles T, Commeiras C, Chobert J, Besson M. Music and dyslexia: A new musical training method to improve reading and related disorders. *Front Psychol* 2016; **7**:26.