Musicians have enhanced subcortical auditory and audiovisual processing of speech and music

Gabriella Musacchia*, Mikko Sams†, Erika Skoe*, and Nina Kraus*‡§¶

*Auditory Neuroscience Laboratory, Department of Communication Sciences, ‡Department of Neurobiology and Physiology, and §Department of Otolaryngology, Northwestern University, Evanston, IL 60208; and †Laboratory of Computational Engineering, Helsinki University of Technology, FI-02015 TKK, Helsinki, Finland

Edited by Michael M. Merzenich, University of California School of Medicine, San Francisco, CA, and approved August 3, 2007 (received for review February 16, 2007)

Musical training is known to modify cortical organization. Here, we show that such modifications extend to subcortical sensory structures and generalize to processing of speech. Musicians had earlier and larger brainstem responses than nonmusician controls to both speech and music stimuli presented in auditory and audiovisual conditions, evident as early as 10 ms after acoustic onset. Phase-locking to stimulus periodicity, which likely underlies perception of pitch, was enhanced in musicians and strongly correlated with length of musical practice. In addition, viewing videos of speech (lip-reading) and music (instrument being played) enhanced temporal and frequency encoding in the auditory brainstem, particularly in musicians. These findings demonstrate practice-related changes in the early sensory encoding of auditory and audiovisual information.

brainstem | plasticity | visual | multisensory language

Musicians tune their minds and bodies by using tactile cues to produce notes, auditory cues to monitor intonation, and visuomotor signals to coordinate with the musicians around them. Musicians have been shown to outperform nonmusicians on a variety of tasks, ranging from language (1) to mathematics (2). Over the past decade, an increasing number of scientists have sought to understand what underlies this seemingly ubiquitous benefit of musical training. We now know that the musician's brain has functional adaptations for processing pitch and timbre (3–6) as well as structural specializations in auditory, visual, motor, and cerebellar regions of the brain (7–9). Some studies also suggest that the interplay between modalities is stronger in musicians (10) and, in the case of conductors, that improved audiovisual task performance is related to enhanced activity in multisensory brain areas (11). Because differences between musicians and nonmusicians are seen in so many different brain areas, we reasoned that the musician's basic sensory mechanism for encoding sight and sound may also be specialized. The high fidelity with which subcortical centers encode acoustic characteristics of sound, and recent evidence for visual influence on human brainstem responses (12), allow us to examine in considerable detail whether the representation of auditory and audiovisual elements are shaped by musical experience. Here, we show that musicians, compared with nonmusicians, have more robust auditory and audiovisual brainstem responses to speech and music stimuli.

Speech and music communication are infused with cues from both auditory and visual modalities. Lip and facial movements provide timing or segmentation cues (e.g., of consonant and vowels), as well as more complex information, such as emotional state, that improve the listener's reaction time and recognition of speech (13–17). Similarly, a musician's face and body movements convey cues for time-varying features of music, such as rhythm and phrasing (e.g., the grouping of notes into a division of a composition), the emotional content of the piece (17), and changes to and from consonant and dissonant musical passages (18). Audiovisual perception of speech and music share some commonalities. For example, viewing lip movements or instrumental playing paired with incongruent auditory sounds modifies what people hear (10, 19). Neurophysiological effects of visual influence on auditory processing mirror perceptual effects. Specifically, lip-reading modifies processing in auditory and multimodal cortices (20–22). In addition, multisensory experience has been shown to directly impact both cortical and subcortical brain areas in animals (23–26).

Human subcortical activity can be captured, with exceedingly high fidelity, by recording the evoked brainstem response (27, 28). The neural origins of the brainstem response have been inferred from studies using simultaneous surface and direct recordings during neurosurgery, studies of brainstem pathologies, and data from animals. Contributors to the first five peaks recorded from the scalp (waves I–V) include the auditory nerve, the superior olivary complex, the lateral lemniscus, and the inferior colliculus (27). It is important to note that peaks of the brainstem response generally have more than one anatomical source, and each source can contribute to more than one peak. The latencies of these peaks are consistent with subcortical origins. In addition, brainstem nuclei have high-frequency phase-locking characteristics that are emphasized in recording with high-pass filtering that attenuates (e.g., cortical) low-frequency signal components of electroencephalographic activity (28).

Electrophysiological responses elicited in the human brainstem reflect the frequency and time-varying characteristics of sound and have been studied extensively to click (29), tonal (30), and speech stimuli (31–33). The brainstem response to a speech syllable can be divided into transient and sustained portions (34, 35). The transient response to speech onset is similar to the click-evoked response used as a clinical tool in hearing assessment (28). The sustained portion, called the frequency-following response (FFR), entrains to the periodicity of a sound, with phase-locked interspike intervals occurring at the fundamental frequency (F0) (36, 37). Measurements of the speech-evoked onset response and FFR, such as peak latencies and spectral amplitudes, have been studied extensively. In addition, it has been shown that these two main features of the brainstem response are influenced by viewing phoneme articulations and auditory training (6, 12, 37, 38), thus making these responses suitable tools for the investigation of musicianship effects.

Here, we used the temporal and spectral resolution of the auditory brainstem response to investigate whether, and to what extent, subcortical processing is malleable and shaped by musical

Author contributions: G.M., M.S., and N.K. designed research; G.M. performed research; G.M. and E.S. analyzed data; and G.M., M.S., E.S., and N.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: FFR, frequency-following response; UA, unimodal acoustic; AV, audiovisual; UV, unimodal visual.

¶To whom correspondence should be addressed at: Auditory Neuroscience Laboratory, Northwestern University, 2240 Campus Drive, Evanston, IL 60208. E-mail: nkraus@northwestern.edu.

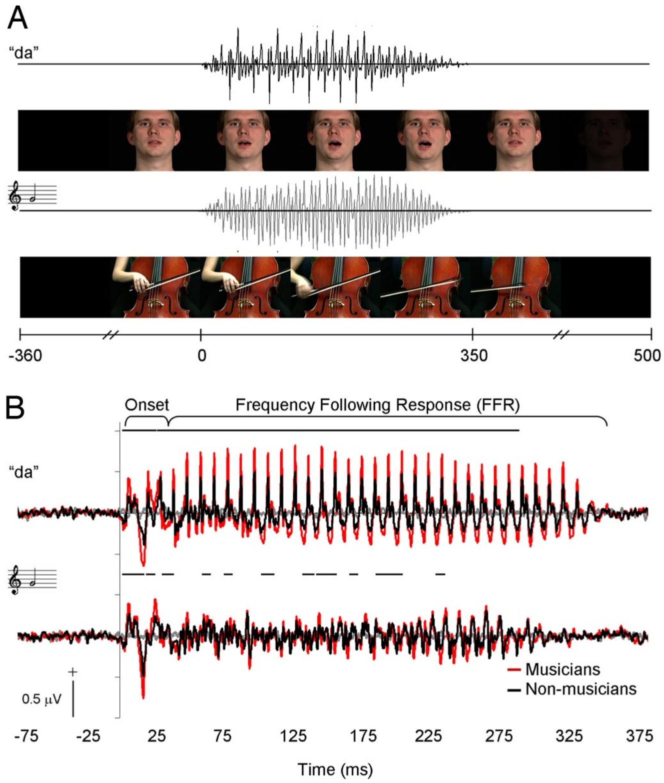

Fig. 1. Stimulus timelines and audiovisual grand averages. (*A*) Auditory and visual components of speech and music stimuli. Visual components were digitized from videos of a speaker uttering ''da'' and a musician bowing a G note on the cello. Acoustic onset for both speech and music occurred 350 ms after the first video frame and simultaneously with the release of consonant closure and onset of string vibration, respectively. Speech and music sounds were 350 ms in duration and similar to each other in envelope and spectral characteristics. (*B*) Grand average brainstem responses to audiovisual speech (*Upper*) and cello (*Lower*) stimuli. Group amplitude differences were assessed by using a sliding-window analysis procedure that resulted in rectified mean amplitude values over 1-ms bins for each subject. Bins with significant differences (t test, $P < 0.05$) are designated by bars over the waveforms for each stimulus type. Amplitude differences in the responses between musicians and controls are evident over the entire response waveforms, especially in the speech condition. UV speech and music stimuli elicited little activity, as indicated by the gray traces.

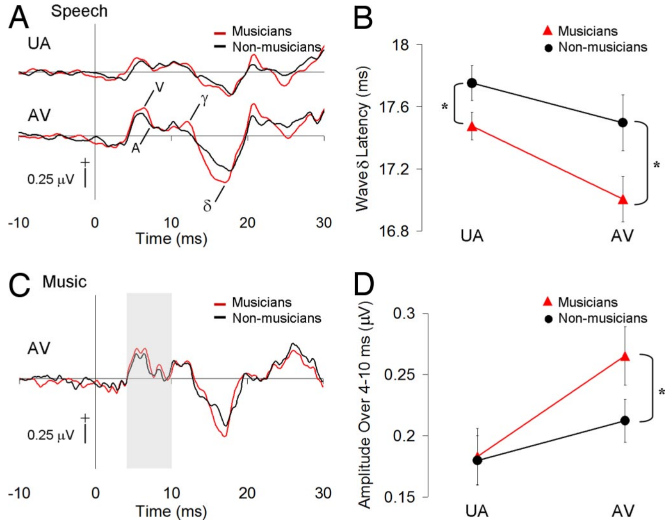

Fig. 2. Musicians have enhanced onset response timing and magnitude. (*A*) Grand average onset responses of the musicians and control subjects to the AV (*Upper*) and UA (*Lower*) speech stimuli. UV speech and music stimuli elicited little activity, as indicated by the gray traces. Prominent peaks of the onset response (V, A, γ, δ) are indicated. Wave δ latencies were earlier in musicians than in controls. (*B*) Mean wave δ latencies for musicians and controls are shown with error bars denoting ± SEM. Musicians had significantly earlier latencies than controls in both the UA and AV conditions. (*C*) Musician and control grand average responses to AV cello stimuli. Mean rectified mean amplitude values were calculated over 4–10 ms (shaded gray) to test whether musicians (red) had larger response magnitude early in the subcortical stream, before cortical excitation. (*D*) Rectified mean amplitudes over 4–10 ms of the AV cello response indicated larger onset responses in musicians than controls to music stimuli.

experience. Although data on musicians and nonmusicians suggest that playing music changes cortical encoding mechanisms, we aimed to test whether musical training engenders plasticity at subcortical levels. We reasoned that auditory and audiovisual stimuli should be used because musical training is multisensory in nature, given its role in developing auditory, audiovisual, and visuomotor skills through extensive practice.

Results

Musicians performed better than controls on the unimodal acoustic (UA) and audiovisual (AV) duration discrimination tasks in the speech condition. ANOVA showed main effects of modality ($F = 23.27$, $P < 0.001$) and group ($F = 7.16$, $P < 0.05$) for error percentage values. Although both groups made fewer errors in the AV condition [$t_{\mathrm{musician\ (mu)}} = 4.86$, $P < 0.01$; $t_{\mathrm{nonmusician\ (nm)}} = 2.79$, $P < 0.05$], musicians performed better than nonmusician controls in both the UA [mean $(\mathrm{M})_{\mathrm{mu}} = 23.4\%$, SD = 14.2; $\mathrm{M}_{\mathrm{nm}} = 35.7\%$, SD = 23.0] and AV conditions ($\mathrm{M}_{\mathrm{mu}} = 8.3\%$, SD = 4.9; $\mathrm{M}_{\mathrm{nm}} = 16.0\%$, SD = 7.8). Musicians did not outperform nonmusicians on the unimodal visual (UV) duration discrimination task, indicating that increased task ability in musicians is limited to tasks involving auditory stimuli in this experiment. Error percentage in the AV speech condition correlated negatively with tonal memory scores from the Musical Achievement Test (MAT) ($r = -0.64$, $P < 0.001$).

Musicians had earlier brainstem responses than nonmusician controls to speech onset in both the UA and AV modalities (Figs. 1*B* and 2). Main effects of group ($F = 6.02$, $P < 0.05$) were observed for wave δ latencies in UA and AV conditions. Speech stimuli elicited earlier wave δ peaks in musicians in the UA ($\mathrm{M}_{\mathrm{mu}} = 17.48$ ms, SD = 0.35; $\mathrm{M}_{\mathrm{nm}} = 17.75$ ms, SD = 0.41) and AV ($\mathrm{M}_{\mathrm{mu}} = 17.01$ ms, SD = 0.58; $\mathrm{M}_{\mathrm{nm}} = 17.50$ ms, SD = 0.65) modalities (Fig. 2*B*). Viewing a speaker's articulation affected the brainstem responses of both groups similarly: there was a main effect of modality ($F = 11.31$, $P < 0.01$), with AV latencies earlier than UA latencies (see means above and Fig. 2*B*). A correlation between wave δ latency and error percentage in the AV speech condition ($r = 0.43$, $P < 0.05$) indicated that the fewer discrimination errors one made, the earlier the wave δ latency.

Musicians also showed an early enhancement of cello sound onset response compared with controls. An analysis of rectified mean amplitude over the onset portions of the cello responses revealed very early group differences in the AV cello condition (Fig. 2*C*). An ANOVA of rectified mean amplitude values taken over 4–10 ms of the AV cello response showed a main effect of subject group ($F = 27.00$, $P < 0.01$). Corrected post hoc t tests revealed that the musicians' AV cello responses were larger than those of controls, even during this early time range ($t = 1.71$, $P < 0.05$).

Striking group differences were observed in the frequency-following portion of the response. Fig. 3 shows the musician and control grand average fast Fourier transform of responses over time for speech and illustrates that musicians have enhanced periodicity encoding (phase-locking), especially relating to the

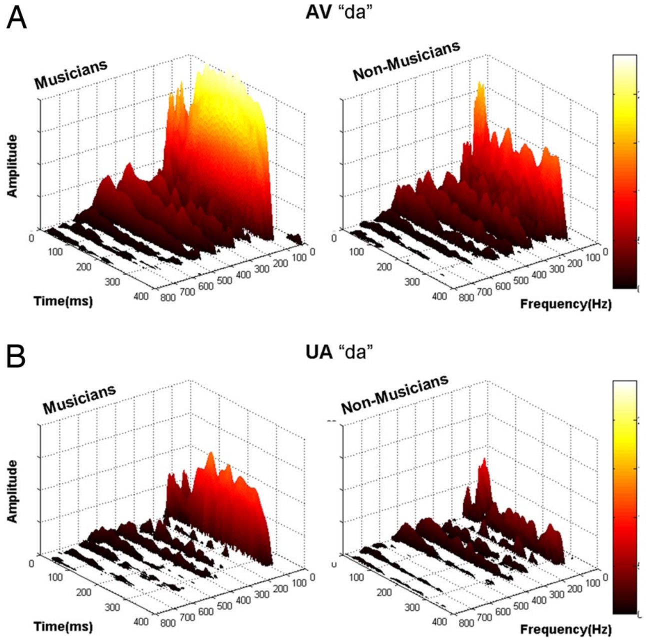

Fig. 3. Musicians have enhanced frequency representation. Narrowband spectrograms were calculated over the entire response to produce time–frequency plots (1-ms resolution) for musician and nonmusician responses to audiovisual (*A*) and unimodal (*B*) speech. Lighter colors indicate greater amplitudes. Musicians have greater spectral energy over the duration of the response than controls, with this difference being most pronounced at F0 (100 Hz). In addition, there was significantly more spectral energy at 100 Hz in the responses to audiovisual in contrast to unimodal auditory stimuli.

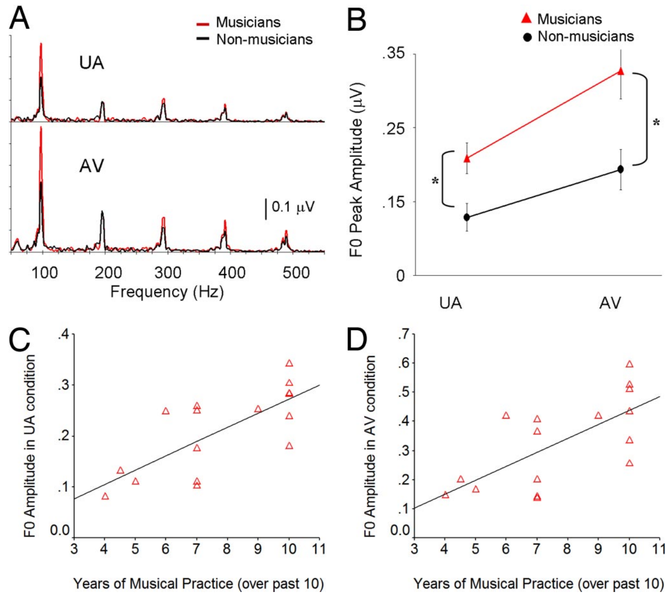

Fig. 4. Enhanced frequency representation in musicians and correlation with musical practice. (*A*) Fast Fourier transform analysis of the entire FFR period (30–350 ms) shows that musicians have more robust F0 peak amplitudes to both unimodal and audiovisual speech stimuli. (*B*) The mean F0 peak amplitudes (±SEMs) were significantly larger in musicians than controls for both unimodal auditory and audiovisual stimuli. (*C* and *D*) Years of consistent musical practice (>3 days/week) over the past 10 years (*x* axis) are plotted against individual peak F0 amplitudes in the UA and AV speech condition (*y* axis). The number of years subjects consistently practiced music correlated highly with the strength of speech pitch encoding (reflected in the peak F0 amplitude) for both UA ($r = 0.73$, $P = 0.001$) and AV ($r = 0.68$, $P < 0.01$) stimuli.

fundamental frequency (F0 = 100 Hz) and throughout the entire FFR period. Statistical analysis performed for F0 and harmonic components showed significant effects only at F0. A pattern similar to that seen for δ wave latency emerged: main effects of modality ($F = 39.96$, $P < 0.001$) and group ($F = 8.13$, $P < 0.01$) were observed for speech. Amplitudes were larger in musicians than in controls for both the UA ($t = 2.81$, $P < 0.0125$; $M_{mu} = 0.21$ μV, SD = 0.08; $M_{nm} = 0.13$ μV, SD = 0.07) and AV conditions ($t = 2.72$, $P < 0.0125$; $M_{mu} = 0.33$ μV, SD = 0.15; $M_{nm} = 0.19$ μV, SD = 0.10) (Fig. 4*B*). In addition, AV responses were larger than the UA ones in both musicians ($t = 5.07$, $P < 0.001$) and controls ($t = 4.54$, $P < 0.001$; see means above). These results suggest that musicians have more robust pitch encoding than controls in both modalities and that viewing phoneme articulations enhances frequency encoding in both groups, particularly in musicians (Fig. 4 *A* and *B*).

Speech-evoked F0 amplitudes correlated positively with how many years musicians had been consistently playing music within the past 10 years (Fig. 4 *C* and *D*). This effect was observed in both the UA ($r = 0.731$, $P = 0.001$) and AV ($r = 0.68$, $P < 0.01$) conditions. In addition, F0 amplitude correlated with how many times per month subjects witnessed musical performances ($r = 0.40$, $P < 0.05$). These data indicate that intensive musical practice and exposure relate to the strength of pitch encoding.

Discussion

This study shows that musicians have more robust brainstem responses to ecologically valid stimuli (speech and music) than controls. The earlier latencies and larger magnitude of onset responses exhibited by musicians suggest that this group has a more synchronous neural response to the onset of sound, which is the hallmark of a high-functioning peripheral auditory system (28). These peaks represent neural activity early in the afferent processing stream, before activation of primary auditory cortex (39). Musicians also exhibited an enhanced representation of the F0, which is widely understood to underlie pitch perception (40).

Our data show a correlation between the amount of practice and strength of F0 representation, suggesting that musicians acquire an enhanced representation of pitch through training. Accurate pitch coding is vital to understanding a speaker's message and identity, as well as the emotional content of a message. Because no correlations were seen with music aptitude or even basic pitch discrimination tasks and F0 encoding, it may be that encoding enhancement is not related to how well one does, but rather to consistency and persistency of practice.

We have established a relationship between musicianship and strength of unisensory and multisensory subcortical encoding. However, our data cannot definitively answer which aspect (or aspects) of musicianship is the fueling force. Musical training involves discrimination of pitch intonation, onset, offset, and duration aspects of sound timing as well as the integration of multisensory cues to perceive and produce notes. Indeed, musicians have been shown to outperform nonmusicians on a variety of tasks, including language (1), visuospatial (41), and mathematical (2) tests. It is also possible that because of their musical training, musicians have learned to pay more attention to the details of the acoustic stimuli than nonmusicians. The robust nature of the differences demonstrated here may open new lines of research that focus on disentangling how these factors contribute to subcortical specialization in musicians.

Given that musicians have more experience with musical stimuli than nonmusicians, it may be initially surprising that the largest observed group differences are in the frequency-following region of the speech condition. The relative paucity of group differences for the musical stimuli may be due to a floor effect given the overall reduced response amplitudes for the cello stimuli for both groups (Figs. 1 and 3). Because cello stimuli elicited smaller FFR responses than speech stimuli, any differences between musicians and nonmusicians may have been harder to detect. The acoustic differences between the sounds may in part account for the differences in the FFR amplitude between speech and music. Although the frequencies of harmonics H2–H5 were the same for speech and music stimuli, the relative amplitude of these components differed. Vocal fold vibrations produce a harmonic spectrum that has large amplitudes of frequencies at the fundamental and the first two formants (in this case, 100, 700–800, and 1,200–1,300 Hz, respectively) with relatively small amplitudes of frequencies between them. This results in an acoustic waveform with a robust fundamental periodicity (Fig. 1). On the other hand, a vibrating string produces a harmonically richer sound with the largest spectral peaks falling at the second through sixth harmonics (200–600 Hz). These harmonics interact to produce an acoustic waveform with a less salient periodicity at the fundamental (Fig. 1). Therefore, our results may reflect a general tuning preference in the auditory system to sounds with robust fundamental frequencies. This suggests that, although speech may elicit brainstem responses with larger signal-to-noise ratios than cello sounds, this enhancement is not exclusive to speech. Further work with other musical stimuli is needed to determine whether or not spectral encoding of music differs between musicians and nonmusicians. Alternatively, but less likely, we can speculate that brainstem structures exhibit a speech-encoding bias, perhaps because of the vastly greater exposure to speech in both groups.

Three mechanisms for brainstem plasticity observed in this study can be suggested. One is that top–down influences, originating from complex, multisensory training, guide plasticity in peripheral areas. This suggestion is derived from the reverse hierarchy theory, which states that learning modifies the neural circuitry that governs performance, beginning with the highest level and gradually refining lower sensory areas (42). Our data corroborate the prediction of this theory that physiological changes correlate with the length of training. A second mechanism is that afferent peripheral structures exhibit Hebbian rules of plasticity (43). Specifically, joint activity of pre- and postsynaptic auditory brainstem neurons stimulated during musical perception and performance leads to a strengthening of the synaptic efficacy of brainstem mechanisms responsible for encoding sound. And finally, a combination of these two mechanisms suggests reciprocal afferent and efferent plasticity that develops and updates concurrently, thus strengthening cortical and subcortical centers simultaneously.

We show auditory brainstem enhancement with the addition of visual stimuli in both groups. Visual influence on auditory brainstem function has been previously shown in humans (12) and is supported by well established lines of research that document how multisensory interactions develop and change with experience in animal brainstem nuclei, such as the superior and inferior colliculi (24, 44, 45). Audiovisual interaction in the colliculi is thought to be accomplished primarily by corticofugal modulation (26). Whether visual stimuli and experience with multisensory stimuli modulate the human auditory brainstem response via feedforward or corticofugal mechanisms is still unknown. The interconnectedness of the auditory afferent pathway (46) as well as efferent anatomical projections from primary and nonprimary cortices to the inferior colliculus (47–50) provide the anatomical bases for either mechanism.

Overall, the results of this study suggest that high-level, complex training, such as learning to play music, impacts encoding mechanisms in peripheral sensory structures. Learning-related increases in cortical activity and neurobiological evidence for increased arborization and neurogenesis in the adult mammalian brain after complex stimulation, as seen in the work by van Praag *et al.* (25), support this interpretation. As in that study, neural specialization through musical training may derive from the richness of musical training. "Critical periods" of musical development (51) as well as the development of pitch, timbre, and melody discrimination skills, which are present as early as 6 months of age (52), may also contribute to the degree of adaptive change. It is likely that the multisensory encoding mechanisms develop and are strengthened by a reciprocal relationship between cortical and subcortical processes, as has been suggested to explain correlations between brainstem and cortical deficits (32, 53, 54). Our data show that musicians have pervasive subcortical specializations that enhance auditory and audiovisual encoding of music and speech sounds, indicating that musical training impacts neural mechanisms beyond those specific to music processing. These findings have practical implications when considering the value of musical training in schools and investigations of auditory training strategies for people with speech-encoding deficits.

Materials and Methods

Twenty-nine adult subjects (mean age, 25.6 ± 4.1 years; 14 females) with normal hearing (<5 dB pure-tone thresholds from 500 to 4,000 Hz), normal or corrected-to-normal vision (Snellen Eye Chart, 2001), and no history of neurological disorders gave their informed consent to participate in this experiment. Subjects completed a musical history form that assessed beginning age and length of musical training, practice frequency and intensity, as well as how often they attended musical performances and listened to music. All subjects were given the Seashore's Test of Musical Talents and self-identified musicians or subjects with any musical experience were given two Musical Achievement Tests (MATs). Subjects who were categorized as musicians (n = 16) were self-identified, began playing an instrument before the age of 5 years, had 10 or more years of musical experience, and practiced more than three times a week for 4 or more hours during the last 10 years. Controls (n = 13) were categorized by the failure to meet the musician criteria, and, as such, a subset of control subjects had some musical experience. Subjects with perfect pitch were excluded from this study.

Six types of stimuli were presented: the UA speech syllable "da" (55), the UA musical sound of a cello being bowed (note G2, recorded from a keyboard synthesizer), the UV video of a male speaker articulating the syllable "da," the UV video of a musician bowing a cello, and the congruent pairings of UA and UV tokens to make AV speech and music tokens (Fig. 1*A*). Both acoustic sounds were 350 ms in length and shared the same (±2 Hz) fundamental frequency (F0 = 100 Hz) and second (H2 = 200 Hz), third (H3 = 300 Hz), fourth (H4 = 400 Hz), and fifth (H5 = 500 Hz) harmonics. Video clips of a speaker's face saying "da" and a cellist bowing G2 were edited to be 850 ms in length (FinalCut Pro 4; Apple, Cupertino, CA). When auditory and visual stimuli were presented together, sound onset was 350 ms after the onset of the first frame. Acoustic onset occurred synchronously with release of consonant closure in the speech condition and onset of string vibration in the music condition.

Speech and music tokens were presented in separate testing sessions, with session order alternated across subjects. In each session, 12 blocks of 600 tokens each were presented with a 5-min break between blocks (Neurobehavioral Systems, Albany, CA; 2001). This yielded 2,400 sweeps per condition (speech and music) for each stimulus type (UA, AV, UV). Acoustic stimuli were presented with alternating polarities. Order of presentation (UA, UV, AV) was randomized across subjects. To control for attention, subjects were asked to silently count the number of target stimuli they saw or heard and then report that number at the end of each block. Target stimuli were slightly longer in duration than the nontargets (auditory target, 380 ms; visual target, 890 ms) and occurred 4.5 ± 0.5% of the time. Performance accuracy was measured by counting how many tokens the subject missed (error percentage).

Continuous electroencephalographic data were recorded from Cz (10–20 International System, earlobe reference, forehead ground), off-line filtered (70–2,000 Hz), epoched, and averaged to result in individual artifact-free averages of at least 2,000 sweeps per stimulus type (music, speech) and condition (UA, UV, AV) (Compumedics, El Paso, TX). Brainstem responses to UV stimuli resulted in neural activity that was indistinguishable from background nonstimulus activity, as has been shown in a previous report of visual influence on brainstem activity (12). Therefore, response measurements in the UV condition were not analyzed.

All analyses were done in parallel for the speech and music conditions. Brainstem onset response peaks (waves V, A, δ, and γ) were picked from each individual's responses (Fig. 2*A*), yielding latency and amplitude information. One rater who was blind to subject group and condition picked the peak voltage fluctuation, and another rater confirmed the first rater's marks. Peak latencies were calculated by subtracting the latency of sound onset (time 0) from the latency of the peak voltage fluctuation for each wave. Strength of pitch encoding was measured by peak amplitudes at F0 (100 Hz), H2 (200 Hz), H3 (300 Hz), H4 (400 Hz), and H5 (500 Hz) of fast Fourier transforms over the FFR period. Magnitude of response was calculated in 1-ms bins over the entire length of the response, and to focus on the onset response, again over just the 4- to 10-ms portion. Two-way repeated-measures ANOVAs and Bonferroni-corrected post hoc *t* tests, when applicable, were used with brainstem and error percentage measures to test whether responses in UA and AV conditions differed between and within groups. Independent *t* tests were applied to the musical aptitude tests. Correlations between behavioral and brainstem measures were also performed.

We thank Dr. Scott Lipscomb, members of the Auditory Neuroscience Laboratory at Northwestern University, and the subjects who participated in this experiment. This work was supported by National Science Foundation Grant 0544846 and National Institutes of Health Grant R01 DC01510.

1. Magne C, Schon D, Besson M (2006) *J Cogn Neurosci* 18:199–211.
2. Schmithorst VJ, Holland SK (2004) *Neurosci Lett* 354:193–196.
3. Zatorre RJ (1998) *Brain* 121:1817–1818.
4. Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M(1998) *Nature* 392:811–814.
5. Peretz I, Zatorre RJ (2005) *Annu Rev Psychol* 56:89–114.
6. Wong PC, Skoe E, Russo NM, Dees T, Kraus N (2007) *Nat Neurosci* 10:420–422.
7. Schlaug G, Jancke L, Huang YX, Steinmetz H (1995) *Science* 267:699–701.
8. Ohnishi T, Matsuda H, Asada T, Aruga M, Hirakata M, Nishikawa M, Katoh A, Imabayashi E(2001) *Cereb Cortex* 11:754–760.
9. Gaser C, Schlaug G (2003) *J Neurosci* 23:9240–9245.
10. Saldana HM, Rosenblum LD (1993) *Percept Psychophys* 54:406–416.
11. Hodges DA, Hairston WD, Burdette JH (2005) *Ann NY Acad Sci* 1060:175–185.
12. Musacchia G, Sams M, Nicol T, Kraus N (2006) *Exp Brain Res* 168:1–10.
13. Massaro DW, Cohen MM (1983) *J Exp Psychol Hum Percept* 9:753–771.
14. Sumby WH, Pollack I (1954) *J Acoust Soc Am* 26:212–215.
15. Summerfield Q (1979) *Phonetica* 36:314–331.
16. Drake C, Palmer C (1993) *Music Percept* 10:343–378.
17. Vines BW, Krumhansl CL, Wanderley MM, Dalca IM, Levitin DJ (2005) *Ann NY Acad Sci* 1060:462–466.
18. Thompson WF, Graham P, Russo F (2005) *Semiotica* 156:203–227.
19. McGurk H, MacDonald J (1976) *Nature* 264:746–748.
20. Sams M, Aulanko R, Hamalainen M, Hari R, Lounasmaa OV, Lu ST, Simola J (1991) *Neurosci Lett* 127:141–145.
21. Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997) *Science* 276:593–596.
22. Calvert GA (2001) *Cereb Cortex* 11:1110–1123.
23. Thompson RF (1986) *Science* 233:941–947.
24. Wallace MT, Meredith MA, Stein BE (1998) *J Neurophysiol* 80:1006–1010.
25. van Praag H, Kempermann G, Gage FH (2000) *Nat Rev Neurosci* 1:191–198.
26. Hyde PS, Knudsen EI (2002) *Nature* 415:73–76.
27. Jacobson JT (1985) *The Auditory Brainstem Response* (Pro Ed, Austin, TX).
28. Hall JW, III (1992) *Handbook of Auditory Evoked Responses* (Allyn and Bacon, Needham Heights, MA).
29. Hood L (1998) *Clinical Applications of the Auditory Brainstem Response* (Singular, San Diego).
30. Galbraith GC, Doan BQ (1995) *Int J Psychophysiol* 19:203–214.
31. King C, Warrier CM, Hayes E, Kraus N (2002) *Neurosci Lett* 319:111–115.
32. Banai K, Nicol T, Zecker SG, Kraus N (2005) *J Neurosci* 25:9850–9857.
33. Johnson KL, Nicol TG, Kraus N (2005) *Ear Hear* 26:424–434.
34. Russo N, Nicol T, Musacchia G, Kraus N (2004) *Clin Neurophysiol* 115:2021–2030.
35. Kraus N, Nicol T (2005) *Trends Neurosci* 28:176–181.
36. Hoormann J, Falkenstein M, Hohnsbein J, Blanke L (1992) *Hear Res* 59:179–188.
37. Krishnan A, Xu YS, Gandour J, Cariani P (2005) *Cogn Brain Res* 25:161–168.
38. Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N (2005) *Behav Brain Res* 156:95–103.
39. Celesia GG (1968) *Arch Neurol* 19:430–437.
40. Moore BCJ (2003) *Introduction to the Psychology of Hearing* (Academic, London), 5th Ed, pp 195–231.
41. Brochard R, Dufour A, Despres O (2004) *Brain Cogn* 54:103–109.
42. Ahissar M, Hochstein S (2004) *Trends Cogn Sci* 8:457–464.
43. Hebb DO (1949) *The Organization of Behavior* (Wiley, New York).
44. Stein BE, Jiang W, Wallace MT, Stanford TR (2001) *Prog Brain Res* 134:143–156.
45. Hyde PS, Knudsen EI (2001) *J Neurosci* 21:8586–8593.
46. Popper AN, Fay RR (1992) *The Mammalian Auditory Pathway: Neurophysiology* (Springer, New York).
47. Saldana E, Feliciano M, Mugnaini E (1996) *J Comp Neurol* 371:15–40.
48. Winer JA, Larue DT, Diehl JJ, Hefti BJ (1998) *J Comp Neurol* 400:147–174.
49. Schofield BR, Coomes DL (2005) *Hear Res* 199:89–102.
50. Lim HH, Anderson DJ (2007) *J Neurophysiol* 97:1413–1427.
51. Trainor LJ (2005) *Dev Psychobiol* 46:262–278.
52. Trehub SE (2003) *Nat Neurosci* 6:669–673.
53. Wible B, Nicol T, Kraus N (2005) *Brain* 128:417–423.
54. Abrams DA, Nicol T, Zecker SG, Kraus N (2006) *J Neurosci* 26:1131–1137.
55. Klatt DH (1977) *J Acoust Soc Am* 67:971–995.