NORTHWESTERN UNIVERSITY

Musical training enhances brainstem and cortical representation of speech and music

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Communication Sciences and Disorders

By

Gabriella Andrea-Eugenia Musacchia

EVANSTON, ILLINOIS

June, 2008

Abstract

Musical training enhances brainstem and cortical representation of speech and music

Gabriella Musacchia

Learning to play an instrument requires the progressive tuning almost all of a student's senses. He trains his ears to hear consonance and dissonance and eyes to read sheet music or read other musicians' body language, for example. The protracted training that is required to become a musician has been shown to bestow perceptual advantages and shape the anatomy and function of the cortex. What is not known is whether musical training shapes subcortical responses and the extent to which putative encoding specializations extend to non-musical or multimodal stimuli. To this end, we explored visual influence on human brainstem responses, how the musician's brainstem differed from non-musicians, and how these differences related to cortical encoding mechanisms. We found that musicians have specialized brainstem mechanisms for encoding pitch periodicity of speech sounds and that the relationship between low- and high-level function is more strongly related in this group, compared to non-musicians. These effects were seen when subjects were listening to music or speech sounds alone and when they viewed concomitant video tokens simultaneously. These data indicate that representation of sight and sound in the human brainstem, which is the neural gateway to higher level function, can be shaped my musical training and that speech and music encoding at this level share neural resources.

Acknowledgements

These studies were supported by grants from National Institute of Health R01 DC01510 and National Science Foundation 0544846. I would like to thank Nina Kraus, Ph.D. and Mikko Sams, Ph.D., for leading this work with integrity, insight and dedication as well as committee members Steve Zecker, Ph.D., Anne Bradlow, Ph.D. and Scott Lipscomb, Ph.D. for their commitment to this effort. The author would also like to acknowledge past and present members of the Auditory Neuroscience Laboratory, Erin Hayes, Ph.D., Trent Nicol, Krista Lynne Johnson, Nicole Russo and Dan Abrams for being the harshest critics and most devoted supporters of this work.

List of Abbreviations

FFR: frequency-following response

UA: unimodal acoustic

AV: audiovisual

UV: unimodal visual

AA: auditory alone

RMA: rectified mean amplitude

CANS: central auditory nervous system

Mus: musician

Nmus: non-musicians

Table of Contents

List of Tables and Figures

Introduction

One of the fundamental issues in the field of communication science is how our central auditory nervous system (CANS) is shaped by auditory training and adapts to environmental changes. Auditory *plasticity*, or neural reorganization as a result of experience, can be divided into developmental, compensatory and learning-related changes (Musiek & Berge 1998). Developmental plasticity occurs in every person born with an intact hearing system as they are exposed to, and eventually learn the sounds around them and the language(s) of their caregivers (Doupe 1997;Hall 1992;Jing & Benasich 2006). When innate hearing is degraded, when hearing is lost or when brain trauma results in an auditory cortex lesion, the neural system strengthens non-auditory activity and connectivity in order to compensate for the deficit. In some cases, compensatory mechanisms can augment other senses beyond normal-hearing capabilities (Bavelier et al. 2000;Roder et al. 2002). While both of the former types of plasticity co-occur with learning, maturational and critical periods of stimulus sensitivity define them. Learning-related plasticity on the other hand can be induced at any stage of development and is evident throughout adulthood (Ahissar 2001;Bergan et al. 2005).

A large portion of auditory neuroscience is devoted to understanding what kinds of auditory training promote learning-related changes and the neural mechanisms that underlie them. Investigations in this area focus on how training shapes different areas of the CANS, and the extent to which changes either develop from peripheral to central structures ("bottom-up" plasticity) or proceed from more cognitive areas to primary and subcortical areas ("top-down"). For example, while training on a frequency discrimination task leads to tonotopic reorganization of the auditory cortex (Recanzone, Schreiner, & Merzenich 1993), this type of reorganization also governs a sharpening of frequency tuning curves in the inferior colliuclus, a subcortical nucleus (IC) (Suga et al. 2002). The extent to which learning promotes changes in human subcortical (brainstem) and cortical structures can be effectively measured with evoked potentials. Language-learning impaired children, for example, develop more robust cortical and brainstem responses to speech after extensive training on frequency and temporal discrimination tasks (Hayes et al. 2003).

Professional musicians represent an ideal model to investigate auditory plasticity because they have undergone protracted auditory training with a set of stimuli that non-musicians are only occasionally exposed to (Munte et al. 2003). Musical expertise is attained through rigorous training, practice and performance schedules that tune the senses. For example, Chicago Symphony Orchestra players spend an approximate average of 12 hours per week from September to March in performances alone (http://www.cso.org/). While non-musician music lovers may receive the benefits of this training in the form of an auditory experience, a gulf of auditory training exists between the performer and the perceiver. In this way, studies that compare perceptual and neurophysiological differences between musicians and non-musicians reveal differences between a normal and a highly-trained auditory system.

Intensive musical practice bestows behavioral advantages in multimodal and linguistic domains. Musicians display less right-handedness preference (Jancke, Schlaug, & Steinmetz 1997) and have finer control of motor timing (Kincaid, Duncan, & Scott 2002). Conductors, who are perhaps masters among experts, can discriminate the pitch, timing and location of auditory and audiovisual targets more accurately than non-musicians (Hodges, Hairston, & Burdette 2005). Musicians also have finer discrimination abilities of linguistic cues that are important for important for speech perception. They can detect weaker manipulations of speech and music pitch than non-musicians (Schon, Magne, & Besson 2004) and smaller differences in non-native languages (Marques et al. 2007). They also have an enhanced ability to extract prosodic cues, which transmit the emotional content of a message, from spoken sentences (Thompson, Schellenberg, & Husain 2003).

Not surprisingly, the complex, multisensory nature of musical training shapes neural anatomy and function in many areas of the brain. Musicians have more grey matter volume, which is dense with neural cell bodies, in motor, auditory and visual cortical areas of the brain (Gaser & Schlaug 2003c). The white matter fissure of axonal projections that connects the right and left hemispheres is also larger in musicians (Schlaug et al. 1995), indicating more efficient communication between the lobes. Musicians have enhanced encoding of pitch, timing and timbre features sound (Pantev et al. 2001). This is especially true when they are listening to the instruments that they themselves play (Munte, Nager, Beiss, Schroeder, & Altenmuller 2003;Pantev et al. 2003). In many of these and other studies, the degree of anatomical or functional specialization was related to the duration, or intensity, of musical training.

Given what we know about plasticity in the CANS and how musical training promotes learning-related neural changes, we hypothesized that musicians' subcortical systems may also be "tuned", relative to non-musicians. We formulated four experimental aims to test this hypothesis. Our first aim was to test whether musicians had specialized subcortical systems. Second, because musicians show behavioral advantages in multiple modalities and in the linguistic domain, we asked whether specialization was limited to the auditory modality or to musical sounds. In order to test this aim, we needed to establish the method for obtaining brainstem responses to seen and heard stimuli. This question allowed us to consider the extent to which music and language share neural resources. Third, we asked whether musical training would enhance particular aspects of sound encoding (e.g. pitch or timbre), or whether musical experience influenced neural encoding in a more pervasive manner. Finally, we aimed to understand the relationship between subcortical and cortical activity, and whether musicians had stronger or weaker correlations between low- and high-level processing.

Our experimental paradigm was to record brainstem and cortical responses from musicians and non-musicians when they listened to and viewed speech and music tokens. Our method was built on well-established neurophysiological methods that Nina Kraus and her colleagues have developed for recording human brainstem responses, as thoroughly described in the Methods sections of Chapters I-IV. Analyses was performed on established measures of brainstem and cortical electrophysiology and compared to behavioral measures of musicianship, psychophysical scores and self-reported musical history answers.

The following chapters describe the methods and results that answered our four experimental aims. Overall, we found that the musician's auditory system is highly tuned to encode the sight and sound of musical and linguistic cues, both at the brainstem and cortical level. At the level of the brainstem, the representation of pitch is particularly enhanced in musicians. We also found faster cortical response timing in musicians and a relationship between cortical measures, periodicity encoding at the brainstem level, and the extent of musical training. The data compiled here represent an advance of knowledge on four fronts 1) that musical training induces plasticity at subcortical levels of encoding, 2) that these learning-related changes are not limited to the auditory and musical domains and 3) that brainstem encoding fidelity is related to cortical representation and 4) that the brainstem-cortical relationship is strengthened by musical training. Of particular importance is the generalization of musical specialization to speech encoding. The implications of this are that musical training changes the brain to encode not only music, but language, more efficiently.

Chapter 1

Our experimental aims necessitated developing a method to record brainstem responses to seen and heard stimuli. The following manuscript describes the methods and results of our work that accomplished that.

Manuscript - Seeing speech affects acoustic information processing in the human brainstem

Exp Brain Res. 2006 Jan;168(1-2):1-10. Epub 2005 Oct 11.

Gabriella A.E. Musacchia[1], Mikko Sams[4], Trent G. Nicol[1], Nina Kraus[1, 2, 3]

Auditory Neuroscience Laboratory, Departments of [1]Communication Sciences, [2]Neurobiology and Physiology and [3]Otolaryngology, Northwestern University, Evanston, IL, USA; [4]Laboratory of Computational Engineering, Helsinki University of Technology, Finland

Keywords: auditory, visual, brainstem, multisensory, speech

Abstract

Afferent auditory processing in the human brainstem is often assumed to be determined by acoustic stimulus features alone and immune to stimulation by other senses or cognitive factors. In contrast, we show that lipreading during speech perception influences early acoustic processing. Event-related brainstem potentials were recorded from 10 healthy adults to concordant (acoustic-visual match), conflicting (acoustic-visual mismatch) and unimodal stimuli. Audiovisual interactions occurred as early as ~11 ms post-acoustic stimulation and persisted for the first 30 ms of the response. Furthermore, the magnitude of interaction depended on

audiovisual pairings. These findings indicate considerable plasticity in early auditory processing.

Introduction

Natural perceptions are rich with sensations from the auditory and visual modalities (Marks 1982). As a friend says hello, we are cheered by their friendly tone and the sight of their smile. At a concert, we are amazed at the sight and sound of a trumpet player's technique. One of the most ubiquitous and well-studied examples of audiovisual integration in humans is seeing and hearing speech. Although acoustic and visual information are seamlessly combined without conscious control (Marks 2004), seeing articulation greatly aids speech acquisition (Kent 1984) and perception (Grant & Seitz 2000;Green 1987), especially in noisy environments (MacLeod & Summerfield 1987;Sumby & Pollack 1954). In addition, seeing articulation that does not match acoustic speech can drastically change what people "hear" (MacDonald & McGurk 1978;Sekiyama et al. 2003). A prevalent model of how audiovisual integration is accomplished posits that information from different modalities is processed along unisensory streams, which converge in cortical structures (MacLeod & Summerfield 1987;Massaro 1998). The combined representation is then processed in a feed-forward fashion that does not affect early, subcortical processing. While this hypothesis has proven to account for copious multisensory phenomena, evidence of audiovisual interaction in subcortical structures encourages modification of the model. These observations prompted the current study, which investigates the timing of seen and heard speech interactions in the human brainstem.

Neuroimaging data have consistently identified cortical sites that show audiovisual effects to speech, and evoked potential data show that effects in these areas happen as early as ~100 milliseconds (ms) post-acoustic onset. Speech processing areas, such as primary auditory cortex, posterior superior temporal cortex (Binder et al. 2000), Broca's area (Burton, Small, & Blumstein 2000) and pre-motor cortex (Scott & Johnsrude 2003;Watkins & Paus 2004) have also shown activity during observation of visual articulatory movements (Calvert et al. 1997;Calvert et al. 1999;Campbell et al. 2001;Nishitani & Hari 2002). In these studies, audiovisual stimuli elicited response enhancement, relative to the sum of the unimodal responses, in multisensory cortices. Sensory-specific cortices, on the other hand, demonstrate response decrements due to audiovisual interaction (Bushara et al. 2003;Klucharev & Sams 2004;Saito et al. 2005). Activity in sensory-specific and superior temporal cortices was affected by visual articulatory information as early as ~100 ms post-acoustic onset in electroencephalogram (EEG) and magnetoencephalogram (MEG) studies (Linkenkaer-Hansen et al. 1998;Lu et al. 1991;Mottonen et al. 2002;Sams et al. 1991). Nonspeech stimuli have been shown to elicit audiovisual interactions at earlier latencies (~90 ms) over primary auditory areas (Giard & Peronnet 1999).

At the subcortical level, neurons of the superior colliculus (SC) have been shown to receive convergent auditory and visual inputs, as well as exhibit audiovisual response properties (Stein, Huneycutt, & Meredith 1998;Wallace, Meredith, & Stein 1993;Wallace, Meredith, & Stein 1998). This compelling line of research has revealed a predominance of supra-additive responses to convergent audiovisual stimuli (from the same time or location) with sub-addition, or suppression, observed less often. Orientation accuracy and audiovisual response properties of neurons in the SC neurons are severely degraded when the ecto-sylvian cortex is deactivated (Jiang, Jiang, & Stein 2006;Jiang & Stein 2003;Perrault, Jr. et al. 2003;Stein et al. 2002). These data suggest that cortical activity is necessary for audiovisual responses to occur in subcortical structures. However, lesions of the SC also disrupt orientation to audiovisual stimuli (Burnett et al. 2004) and there are some audiovisual areas of the SC that do not receive descending projections from the cortex (Wallace et al. 2004b). Because the time course of afferent and efferent audiovisual response properties is not known, we cannot tell when interaction first occurs or the time course of corticofugal modulation.

The principal aim of this investigation was to test whether viewing articulatory gestures influenced the subcortical response to acoustic speech. Our approach was to record event-related responses to seen and heard speech using well-established methodology for recording the unimodal auditory speech-evoked brainstem response (Cunningham et al. 2001;King et al. 2002;Kraus & Nicol 2005;Russo et al. 2004;Russo et al. 2005;Wible, Nicol, & Kraus 2004;Wible, Nicol, & Kraus 2005). The speech-evoked response has been shown to be similar in precision to the click-evoked brainstem response, whose reliability and replicability have enabled its widespread clinical use. Peak-latency differences to click stimuli as small as a few milliseconds can be diagnostically significant in individuals with audiological or neurological abnormalities (Jacobson 1985;Moller 1999). Similarly, small delays in brainstem timing can distinguish normal and language-learning impaired groups using speech (Cunningham, Nicol, Zecker, Bradlow, & Kraus 2001;Hayes, Warrier, Nicol, Zecker, & Kraus 2003;King, Warrier, Hayes, & Kraus 2002;Russo, Nicol, Zecker, Hayes, & Kraus 2005;Wible, Nicol, & Kraus 2004;Wible, Nicol, & Kraus 2005).

The hypothesis for the current study was that acoustic and visual speech generates audiovisual (AV) interactions in human subcortical structures. The time course of the interaction, recorded by evoked potentials, could help inform the extent to which AV mechanisms operate early or late in the processing stream. To investigate this, an acoustic speech syllable was paired with either concordant or conflicting visual articulatory gestures

(Klucharev, Mottonen, & Sams 2003). Brainstem responses were recorded when unimodal stimuli were presented separately and together. This presentation paradigm enabled two complementary data analysis strategies. Modulation effects, or, how the unimodal acoustic response is changed by the addition of visual stimuli, could be identified by differences between the AV response and responses to the unimodal acoustic (UA) stimulus. In addition, AV response features that deviated from the mathematical combination of the UA and unimodal visual (UV) responses could be considered evidence of true, nonlinear, audiovisual interaction mechanisms.

Materials and Methods

Subjects. Ten adults (five females and five males; ages 18-35, mean age 25) participated in this experiment after giving informed, written consent. This experiment was carried out in accordance with the ethical principles laid down in the 1964 Declaration of Helsinki. All subjects performed visual and auditory tests to confirm normal or corrected 20/20 vision (Logarithmic Visual Acuity Chart “2000”, Precision Vision) and hearing thresholds at or below 20 dB HL for octaves from 500 to 4000 Hz. The testing session was conducted in a sound-attenuated booth with a background sound level of 34 dB SPL. Subjects were seated in a comfortable chair, facing a 15.2 cm x 19.2 cm projection screen, 2.3 m away.

Stimuli and presentation sequence. The acoustic stimulus consisted of a five-formant synthetic 100 ms speech syllable, /da/, created with a DH Klatt synthesizer. Following a 10 ms consonant burst, a 30 ms formant transition was followed by a 60 ms steady-state vowel with a fundamental frequency of 100 Hz. Additional stimulus details can be found in previous studies (Bradlow et al. 1999). The consonant burst was amplified by an additional 3 dB (CoolEdit Pro 2000, Syntrillium), in order to elicit robust responses to acoustic onset. Visual stimuli were created from a digital recording of a male speaker articulating /da/, /du/ and /fu/ utterances. All three articulations were edited to 19 frames that began and ended with the same neutral resting position (FinalCut Pro 4, Apple Software and MorphMan 4.0, Stoik Imaging). Each frame was presented for 33.15 ms (sd=1.2), which brought the total visual stimulus duration to 630 ms. The release of the consonant was edited to occur at frame 11 for all three visual tokens. When presented together, acoustic speech onset occurred synchronously with presentation of the 11th frame (Figure 1).

Stimulus sequences were delivered with Presentation software (Neurobehavioral Systems, Inc., 2001) and presented in separate blocks of UA, AV and UV stimuli. The rate of presentation for all three stimulus conditions was 1.59/s. In the UA stimulus sequence, short blocks of 200 acoustic stimuli were presented at 84 dB SPL binaurally through ear inserts (ER-3, Etymotic research). Both stimulus polarities (condensation and rarefaction) were presented equally to ensure that the cochlear microphonic did not affect the brainstem response. To control for attention, subjects were asked to count how many sets of 50 /da/ tokens they heard. In the AV stimulus sequence, the synthesized speech syllable was paired with randomly presented /da/ ($AV_{Concordant}$, 40%), /fu/ ($AV_{Conflicting}$, 40%) and /du/ (target, 20%) visual utterances. The UV stimulus sequence consisted of randomly presented visual tokens (/da/ 40%, /fu/ 40%, /du/ 20%). To control for attention in the AV and UV condition, subjects were asked to watch the video and count the number of /du/ tokens presented in each block.

Recording parameters. Continuous EEG was acquired with Neuroscan 4.3 (Compumedics, El Paso, TX) from Cz (impedance < 5 kΩ), referenced to the nose, band pass filtered from 0.05 to 3000 Hz and digitized at 20,000 Hz. Simultaneously, online averaged evoked potentials were collected with an artifact criterion of > +/- 65μV to ensure that at least 1000 good repetitions per condition were collected. These averages were not used for data analysis. Instead, the continuous EEG was processed offline to create the epoched averages for each condition. The continuous file was band pass filtered from 75 to 2000 Hz to select the brainstem response frequencies (Hall, 1992). The EEG was then divided into epochs (20 ms pre- to 120 ms post-acoustic onset). An artifact criterion of > +/- 65μV was applied to the epochs created from the continuous files in order to reject epochs that contained myogenic and eye blink artifacts. The remaining epochs were then separately averaged, according to stimulus type, and contained between 1000 and 1100 sweeps per non-target type. In order to correct for DC drift, the mean amplitude of the 20 ms epoch immediately preceding acoustic onset was subtracted from the response.

Response measurements. Signal-to-noise ratios were calculated by comparing the pre- (-20 to 0 ms) and post- (0 to 100 ms) stimulus periods. The timing of the brainstem response was quantified by peak-latency and cross-correlation measures. The peaks of Waves V, γ, ε, and κ (Figure 2A) were chosen by visual inspection for all subjects, in all conditions, by two investigators. Cross-correlation measures (Pearson's r) were performed over a latency range that included Wave γ and the completion of its negative trough (8 to 20 ms). This analysis technique shifts one waveform in time to obtain a maximal correlation value. The lag at which this maximum correlation is attained is an indication of a response timing difference. Peak latency measures and cross-correlations provide information about when the response culminates in time and the degree of neural synchrony.

To assess the effects of visual speech on the size of the acoustic response, rectified mean amplitude (RMA) of the periodic and onset portions of the response was calculated. Individuals' latencies for Waves V, ε, and κ were used to delineate the per-subject time ranges for RMA calculations. Onset RMAs were calculated between V and ε; RMAs, spectral analysis and cross –correlations of the frequency following response (FFR) were calculated between ε and κ.

Data analysis. Modulation effects, or changes in the acoustic response due to the addition of visual stimuli, were investigated using a repeated-measure ANOVA with three levels as within-subject factors (UA, $AV_{Concordant}$ and $AV_{Conflicting}$). Interaction effects, or the difference between the AV responses and the summed unimodal responses, were explored using a repeated-measure ANOVA with four levels as within-subjects factor ($AV_{Concordant}$, $AV_{Conflicting}$ and their summed unimodal counterparts). Greenhouse-Geisser corrections were applied if applicable. Protected paired t-tests were performed subsequent to significant ANOVAs. Correlation values and lags were subjected to single-sample t-tests to determine if they differed from zero.

Results

Results I: Description of responses

The grand average responses of all subjects to the three unimodal stimuli (UA /da/, UV /da/ and UV /fu/) are illustrated in Figure 2A. The onset of the acoustic stimulus elicited a series of transient, biphasic peaks. Figure 2B shows that the vowel portion of the stimulus evoked an FFR, which reflects phase-locking to the waveform of the stimulus (Galbraith et al. 1995;Marsh, Brown, & Smith 1975).

In all subjects, and evident in the average, the first prominent peak, Wave V (UA mean latency 6.16 ms, sd= 0.34), was followed by a negative trough, previously reported as Wave A (Russo, Nicol, Musacchia, & Kraus 2004). Wave V mean latency and standard deviation was similar to the normative values reported in previous studies. A positive peak that was not observed in previous studies followed Waves V and A. Some differences in response morphology were expected due to differences between the current and previous stimuli. To avoid confusion between the present and previously reported peaks, the Greek alphabet was used to describe peaks following Wave A. The positive peak following Wave A was referred to as Wave γ. The periodic portion of the response (FFR) began with a positive peak, Wave ε, and ended at a negative peak, Wave κ. Neither the /da/ or /fu/ UV responses elicited replicable peaks across subjects and exhibited low SNRs (0.94 and 1.32 respectively), indicating that the visual stimulus alone elicited little evoked activity with the recording parameters and electrode placement reported here.

Results could not be explained by differences across conditions in signal-to-noise ratios (SNR) or overall electrical activity, as measured by the RMA over pre-stimulus periods. SNR values demonstrated that the signal measurements were distinguishable from noise in the UA and AV conditions (SNR_{UA}=5.23 sd=1.02, $SNR_{Concordant}$=5.68 sd=1.20, $SNR_{Conflicting}$=4.55 sd=1.47). SNR values were not significantly different across conditions (F(2,18)=0.96; p=0.44, ε=0.96). The overall electrical activity generated by electrical noise and non-stimulus related EEG activity, measured by the RMA over –20 to 0 ms, was not significantly different across UA, AV and UV conditions (F(2,18)=0.495; p= 0.63).

Results II: Lipreading delays the brainstem response to speech onset

The presentation of either visual stimulus modulated the timing of the brainstem response to speech at Wave γ (Figure 3, Table 1). There were no differences in Wave V, ε, and κ latencies across conditions. Latency differences at Wave γ were evident across conditions ($F(2,18)=6.77$; $p<0.05$, $\varepsilon=0.51$) and prolonged in both $AV_{Concordant}$ and $AV_{Conflicting}$ responses, relative to the UA response ($p_{Concordant}<0.01$, $t=3.26$; $p_{Conflicting}<0.01$, $t=3.11$). Wave γ latencies in the concordant condition were prolonged in nine out of ten subjects and in seven out of ten in the conflicting condition. Wave γ latencies did not differ significantly between the two AV conditions.

Inter-peak intervals between Wave V and Wave γ ($\gamma_{latency}-V_{latency}$) were computed to confirm that the modulation delay occurred subsequent to Wave V. The inter-peak interval difference was evident across UA and AV conditions ($F(2,18)=4.88$; $p<0.05$, $\varepsilon=0.56$) and was prolonged in both the $AV_{Concordant}$ ($p=0.02$, $t=2.53$) and $AV_{Conflicting}$ ($p=0.01$, $t=2.85$) conditions when compared to the UA condition. A prolonged inter-peak interval was evident in nine out of the ten individuals in the $AV_{Concordant}$ condition and in seven subjects in the $AV_{Conflicting}$. This finding, combined with the null result for Wave V latencies across conditions ($F(2,18)=.87$; $p=0.44$, $\varepsilon=0.84$), indicated that modulation of the unimodal response did not begin previous to Wave γ.

A maximal correlation between UA and $AV_{Concordant}$ onset responses occurred with a lag of 0.69 ms ($p<0.05$, $t=2.66$). The maximum correlation between UA and $AV_{Conflicting}$ responses (0.36 ms lag) was not significantly different from zero.

The difference between the two AV conditions and their computed UA+UV counterparts revealed a true nonlinear audiovisual interaction at Wave γ. Wave γ latencies were different across conditions ($F(3,27)=6.21$; $p<0.05$, $\varepsilon=0.38$) with delays evident in both the $AV_{Concordant}$ ($p<0.01$, $t= 2.91$) and $AV_{Conflicting}$ ($p<0.01$, $t= 3.17$) responses when compared to their respective unimodal sums. Nine out of ten individuals exhibited this latency interaction in the $AV_{Concordant}$ and eight out of ten in the $AV_{Conflicting}$ condition.

Inter-peak intervals between Wave V and γ also demonstrated an interaction ($F(3,27)=4.46$; $p=0.011$, $\varepsilon=0.39$). Prolonged intervals were evident in both $AV_{Concordant}$ ($p<0.05$, $t=2.06$) and $AV_{Conflicting}$ ($p=0.011$, $t=2.71$) conditions compared to their respective unimodal sums. Again, no differences in interaction effects were observed between concordant and conflicting conditions. It is important to note that our data reflect some variance in Wave γ delay across individuals. The perceptual or subject characteristics that may have contributed to this variance were not pursued in this study, but are an intriguing direction of future research.

Results III: Two types of visual stimuli modulate the size of the acoustic brainstem response to speech differently

The two types of visual stimuli modulated the size of the acoustic brainstem response differently. RMA values, as measured between waves V and ε, were different across UA (Mean RMA 0.26 μV, $sd=0.11$), $AV_{Concordant}$ (Mean RMA 0.19 μV, $sd=0.05$) and $AV_{Conflicting}$ (Mean RMA 0.21 μV, $sd=0.06$) conditions ($F(2,18)=5.82$; $p<0.01$, $\varepsilon=0.59$) and were diminished in both the $AV_{Concordant}$ ($p<0.01$, $t=3.31$) and $AV_{Conflicting}$ ($p<0.05$, $t=2.37$) responses compared to the UA. In contrast to the onset timing finding, in which both $AV_{Concordant}$ and $AV_{Conflicting}$ Wave γ

latencies were delayed to the same degree, Table 2 and Figure 4 shows a greater suppression in the $AV_{Concordant}$ response than the $AV_{Conflicting}$ response ($p<0.05$, $t=2.47$).

The size of the AV onset responses compared to their summed unimodal counterparts revealed an audiovisual interaction effect. The onset RMA values in both AV conditions were smaller than those in the summed unimodal responses ($F(3,27)=11.26$; $p<0.01$, $\varepsilon=0.40$; $p_{Concordant}<0.01$, $t=4.97$; $p_{Conflicting}<0.01$, $t=3.01$). The extent of the AV suppression over the onset response was not correlated with the length of the Wave γ delay for either concordant or conflicting stimuli. No statistical evidence of modulation or AV interaction was observed over the FFR region of the responses, using the three methods described in *Response Measurements*.

Discussion

The results of the current study demonstrate that seeing facial movements (lipreading) delays and suppresses the amplitude of the human brainstem response to acoustic speech. The effect of audiovisual delay, on average 1.3 ms, was evident in both $AV_{Concordant}$ and $AV_{Conflicting}$ conditions and occurred as early as 11 ms post-acoustic stimulation. Although both the $AV_{Concordant}$ and $AV_{Conflicting}$ RMAs were smaller compared to the UA condition, the extent of diminution depended on the type of facial movement. The $AV_{Concordant}$ response was more suppressed than those to the $AV_{Conflicting}$ response. The observed effects in the audiovisual conditions could not be attributed to activity elicited by the visual stimuli alone, because measures of the summed unimodal responses (UA+UV) did not differ from UA responses.

These results suggest that early auditory processing is susceptible to visual influence. The observed differences between the latency of Wave γ elicited by UA and AV stimuli are, to our knowledge, the earliest reported audiovisual speech interaction. The time frame of the delay,

~11 ms post-acoustic stimulus, precludes the possibility of audiovisual interaction from simultaneous visual information at acoustic onset, because visual information takes longer to propagate to brainstem structures than acoustic information (Wallace, Meredith, & Stein 1998). Therefore, the interaction must be due to the processing of visual information that precedes acoustic stimulation. The authors suggest two hypotheses as to how this may be accomplished.

One hypothesis is that visual information that precedes acoustic stimulation engages cortical gating or attention mechanisms that directly modulate subcortical acoustic processing. Although early components of the acoustic-evoked response (latency range 2-40 ms) have not generally shown replicable effects of attention (for review, see (Picton & Hillyard 1974), some effects have been observed. In audiovisual conditions and in cases of very difficult acoustic target detection, effects of attention have been observed between 20 and 50 ms post-acoustic onset (Hillyard et al. 1987;Hoormann, Falkenstein, & Hohnsbein 1994;Teder-Salejarvi et al. 2002;Woldorff, Hansen, & Hillyard 1987). The results of these studies suggest that early auditory processing could be selectively tuned by mechanisms recorded as slow ‘anticipatory’ evoked responses to stimulus cues. The audiovisual effects described in these studies produced considerably smaller delays than those observed here. Although hypotheses regarding speech vs. nonspeech stimuli cannot be derived directly from this study, it is possible that lipreading may produce larger differences between unimodal and bimodal stimuli than those observed to nonspeech stimuli. The complexity of speech stimuli, relative to flashes and tones for example, or the extensive experience humans have with lipreading may contribute to the difference in effect size.

Converging evidence from animal and human studies also suggests that the corticofugal system has a role in attentional modulation of subcortical auditory nuclei (for review, see (Suga & Ma 2003) as low as the cochlear nucleus (Oatman & Anderson 1977). In these studies, activity in the auditory nuclei was reduced when subjects attended to visual stimuli, which parallels the amplitude suppression observed in the current study. Recent investigations have shown that the synthesis of acoustic and visual cues in the cat SC is greatly compromised when areas of the auditory cortex are deactivated (Jiang & Stein 2003), indicating that the cortex plays a functional role in mediating audiovisual integration in the superior colliculus. The cortical gating/attentional hypothesis could also explain the range of audiovisual delay across individuals. Target identification scores were used only to ensure 80% correct identification, and statistical analysis of the responses was not performed. Therefore, it is possible that the extent of delay is related to greater attentional focus and higher hit rates.

The alternative hypothesis is that ongoing activity in visual brainstem nuclei, combined with afferent acoustic processing, increases the degree of neural asynchrony, relative to unimodal processing, recorded as total electrical activity from the scalp. A fundamental property of event-related potentials is that a decrease in synchrony of firing, for example due to aggregate neural populations firing at slightly different times, results in longer peak latencies (Hall 1992). Visual or audiovisual nuclei in the brainstem that do not fire in concert with those involved in UA processing could produce the observed delay. Excitation of different brainstem nuclei with opposite dipoles could also produce the observed cancellation, or suppression, of total electrical activity recorded from the surface of the scalp. Although AV fMRI data from the human SC have been limited to nonspeech stimuli (Calvert 2001), acoustic and visual cues that coincide in time and space have been shown to produce enhancement, rather than the suppression seen here.

It is possible that acoustic stimuli (presented with ear inserts) were encoded as spatially
disparate from the visual tokens (projected in front of the subject). However, the observed difference between the RMA of the $AV_{Concordant}$ and $AV_{Conflicting}$ responses would be unexpected, given that the spatial disparity would be equal across the two conditions. Response suppression, like that observed in the current study, has previously been shown in the acoustic and visual spatial maps of the barn owl brainstem to spatially concordant cues (Hyde & Knudsen 2001), prompting the theory that concordant stimuli are 'easier' to process. It is conceivable that the audiovisual response to our primary means of communication, speech, engages a similar interaction mechanism.

Although single-channel ERP recording precludes localization, the timing of the AV effects observed in this study is consistent with activation of nuclei before thalamus and cortex. The latency differences between UA and AV responses take place before initial excitation of the human primary auditory cortex, detected in direct intracranial recordings at 12-15 ms post-acoustic stimulation (Celesia 1968). It is important to note that Celesia and colleagues used rapid-onset click stimuli, which elicit earlier latencies than tone or speech stimuli (Hall 1992). Tone stimuli have been shown to elicit a peak of activity at 13.5 ms post-acoustic onset in the human thalamus and at 17 ms in the auditory cortex (Yvert et al. 2002). Because the audiovisual delay observed in the current study occurred at about 11 ms post-acoustic stimulation, i.e. before reported activation of auditory cortex and thalamus, it is reasonable to suggest that the interaction is taking place in the afferent brainstem pathway.

Although Wave V latency was not prolonged in the audiovisual conditions, this does not preclude the contribution of Wave V generators to later peaks. Studies designed to determine the sources of scalp-recorded auditory brainstem response indicate that the inferior colliculus and lateral lemniscus are the primary generators of Wave V (Gardi, Merzenich, & McKean 1979). However, these studies also consistently demonstrated that the onset discharge of single units in multiple generator sites corresponds in time to the latency of several different (II-V) waves. Deduction of where the AV interactions are taking place is furthered by evidence of converging acoustic and visual inputs on neurons in the superior colliculus (Meredith & Stein 1986b). Despite the localization constraints of ERPs, nuclei of the midbrain emerge as the most likely generators of interaction in the current study.

The results of this study cannot clearly differentiate between speech and nonspeech effects because there were no nonspeech controls. However, because the stimuli were in fact speech tokens, we can discuss the implications of our findings in terms of both speech-specific and more generalized audiovisual interaction hypotheses.

One implication is that speech is processed via a specialized module in which the articulatory gestures could influence afferent speech processing in a way that is unique from nonspeech tokens. A long-debated question is whether speech is processed differently than nonspeech sounds (Chomsky 1985;Hauser, Chomsky, & Fitch 2002). Separate brain mechanisms have been shown to be active for acoustic speech and nonspeech processing (e.g., (Binder, Frost, Hammeke, Bellgowan, Springer, Kaufman, & Possing 2000;Tervaniemi & Hugdahl 2003) and a strong relationship between phoneme perception and motor imitation has been found (Gallese et al. 1996). A related implication is that extensive experience with audiovisual speech results in plasticity of the system such that visual articulatory gestures have unique access to the auditory brainstem. This would suggest that speech is processed in a qualitatively different way from nonspeech, and that precursors of phonetic discrimination operate at the level of the brainstem to discern the degree of audiovisual concordance for later processing.

Alternatively, any visual cue that facilitates attention to acoustic stimulus onset, regardless of linguistic content, may modulate early auditory brainstem activity. Subtle differences in the pre-acoustic visual quality (such as that between /da/ and /fu/ visual facial movements) independent of their concordance, or lack thereof, to the accompanying sound, may be responsible for the effect.

These findings challenge the prevailing view about the human brainstem as a passive receiver/transmitter of modality-specific information. Future investigations on the nature of early audiovisual interactions, and the experimental conditions that contribute to the extent of these effects, will most likely have a great impact on our understanding of sensory processing. The results of the current study are reflections of a new zeitgeist in science today: that our neural system is an active information seeker that incorporates multisensory information at the earliest possible stage in order to discern meaningful objects from the world around it.

Acknowledgments

NIH R01 DC01510 supported this work. The authors wish to thank their collaborators in the Auditory Neuroscience Laboratory at Northwestern University and in the Laboratory of Computational Engineering at the University of Helsinki.

Chapter 2

With the knowledge gained from the previous experiment, we designed our main experiment with musicians and non-musicians. The following paper describes how the musician's brainstem response to seen and heard speech and music differs from non-musicians.

Manuscript - Musicians have enhanced subcortical auditory and audiovisual processing of speech and music.

Proc Natl Acad Sci U S A. 2007 Sep 26.

Gabriella Musacchia[1], Mikko Sams[2], Erika Skoe[1], and Nina Kraus[1,3]

Auditory Neuroscience Laboratory, Departments of [1]Communication Sciences, [2]Neurobiology and Physiology and [3]Otolaryngology, Northwestern University, Evanston, IL, USA; [4]Laboratory of Computational Engineering, Helsinki University of Technology, Finland

Abstract

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 non-musician controls to both speech and music stimuli presented in auditory and audiovisual conditions, evident as early as 10 ms post-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 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.

Text

Musicians tune their minds and bodies, using tactile cues to produce notes, auditory cues to monitor intonation and visuo-motor signals to coordinate with the musicians around them. Musicians have been shown to outperform non-musicians on a variety of tasks, ranging from language (Magne, Schon, & Besson 2006) to mathematics (Schmithorst & Holland 2004). 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 (Pantev et al. 1998;Peretz & Zatorre 2005;Wong et al. 2007a;Zatorre 1998a) as well as structural specializations in auditory, visual, motor and cerebellar regions of the brain (Gaser & Schlaug 2003c;Ohnishi et al. 2001;Schlaug, Jancke, Huang, Staiger, & Steinmetz 1995). Some studies also suggest that the interplay between modalities is stronger in musicians (Saldana & Rosenblum 1993), and in the case of conductors, that improved audiovisual task performance is related to enhanced activity in multisensory brain areas (Hodges, Hairston, & Burdette 2005). Because differences between musicians and non-musicians 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 brainstem responses (Musacchia et al. 2006a), allow us to examine in considerable detail whether the representation of auditory and audiovisual elements are shaped by to musical experience.   Here we show that musicians, compared to non-musicians, 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 (Drake & Palmer 1993;Massaro & Cohen 1983;Sumby & Pollack 1954;Summerfield 1979;Vines et al. 2005). 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 (Vines, Krumhansl, Wanderley, & Levitin 2005), and changes to and from consonant and dissonant musical passages (Thompson, Graham, & Russo 2005).   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 (McGurk & MacDonald 1976;Saldana & Rosenblum 1993). Neurophysiological effects of visual influence on auditory processing mirror perceptual effects. Specifically, lip-reading modifies processing in the auditory and multimodal cortices (Calvert, Bullmore, Brammer, Campbell, Williams, McGuire, Woodruff, Iversen, & David 1997;Calvert 2001;Sams, Aulanko, Hamalainen, Hari, Lounasmaa, Lu, & Simola 1991). In addition, multisensory experience has been shown to directly impact both cortical and subcortical brain areas in animals (Hyde & Knudsen 2002;Thompson 1986;Van, Kempermann, & Gage 2000;Wallace, Meredith, & Stein 1998).

Human subcortical activity can be captured, with exceedingly high fidelity, by recording the evoked brainstem response (Hall 1992;Jacobson 1985). 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 cochlear nucleus, the superior olivary complex, the lateral lemniscus, and the inferior colliculus (Jacobson 1985). 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 which are emphasized in the recording with high-pass filtering that attenuates low frequency signal components of electroencephalographic activity (Hall 1992). Electrophysiological responses elicited in the human brainstem reflect the frequency and time-varying characteristics of sound, and have been studied extensively to click (Hood 1998), tonal (Galbraith & Doan 1995), and speech stimuli (Banai et al. 2005;Johnson, Nicol, & Kraus 2005;King, Warrier, Hayes, & Kraus 2002). The brainstem response to a speech syllable can be divided into transient and sustained portions (Kraus & Nicol 2005;Russo, Nicol, Musacchia, & Kraus 2004). The transient response to speech onset is similar to the click-evoked response used as a clinical tool in hearing assessment (Hall 1992). 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) (Hoormann et al. 1992;Krishnan et al. 2005). Measurements of the speech-evoked onset response and FFR, such as peak latencies and spectral amplitudes, have been extensively studied. In addition, it has been shown that these two main features of the brainstem response are influenced by viewing phoneme articulations and auditory training (Krishnan, Xu, Gandour, & Cariani 2005;Musacchia, Sams, Nicol, & Kraus 2006a;Russo, Nicol, Zecker, Hayes, & Kraus 2005;Wong, Skoe, Russo, Dees, & Kraus 2007a); thus making these responses suitable tools for the investigation of musicianship effects.

Here, we utilized 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 experience. Although data on musicians and non-musicians 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 visuo-motor 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. Analysis of variance showed main effects of modality ($F=23.27$, $p<0.001$) and group ($F=7.16$, $p<0.05$) for Error% values. Although both groups made fewer errors in the AV condition ($t_{mu}=4.86$, $p<0.01$; $t_{nm}=2.79$, $p<0.05$), musicians performed better than non-musician controls in both the UA ($M_{mu}=23.4\%$ $SD=14.2$, $M_{nm}=35.7\%$ $SD=23.0$) and AV conditions ($M_{mu}=8.3\%$ $SD=4.9$, $M_{nm}=16.0\%$ $SD=7.8$). Musicians did not outperform non-musicians 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% in the AV speech condition correlated negatively with tonal memory scores from the Musical Achievement Test (MAT) ($r=-.64$, $p<0.001$).

Musicians had earlier brainstem responses than non-musician controls to speech onset in both the UA and AV modalities (Figs. 5B & 6). 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 (*Mmu*=17.48 ms *SD*= 0.35, *Mnm*=17.75 ms *SD*= 0.41) and AV (*Mmu*=17.01 ms *SD*=0.58, *Mnm*=17.50 ms *SD*=0.65) modalities (Fig. 2B). 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. 6B*). A correlation between Wave δ latency and Error% 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 to controls. An analysis of Rectified Mean Amplitude (RMA) over the onset portions of the cello responses revealed very early group differences in the AV cello condition (Fig. 6C). Analysis of RMA 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. Figure 7 shows the musician and control grand average Fast Fourier Transform (FFT) of responses over time for speech and illustrates that musicians have enhanced periodicity encoding (phase-locking), especially relating to fundamental frequency (F0=100

Hz) 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 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, $p<0.0125$; *Mmu* 0.21 uV, *SD* 0.08; *Mnm* 0.13 uV, *SD* 0.07) and AV conditions (t=2.72, $p<0.0125$; *Mmu* 0.33 uV, *SD* 0.15; *Mnm* 0.19 uV, *SD* 0.10) (Fig. 8B). 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 (Figure 8B).

Speech-evoked F0 amplitudes correlated positively with how many years musicians had been consistently playing music within the past 10 years (Fig. 4C & D). This effect was observed in both the UA (r=.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 (Hall 1992). These peaks represent neural activity early in the afferent processing stream, prior to activation of primary auditory cortex (Celesia 1968). Musicians also exhibited an enhanced representation of the F0, which is widely understood to underlie pitch perception (Moore 2003).

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 non-musicians on a variety of tasks, including language (Magne, Schon, & Besson 2006), visuospatial (Brochard, Dufour, & Despres 2004) and mathematical (Schmithorst & Holland 2004) 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 non-musicians, it may be initially surprising that the largest observed group differences are in the frequency following region of 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 (Figures 1 and 3). Because cello stimuli elicited smaller FFR responses than speech stimuli, any differences between musicians and non-musicians may have been harder to detect. The acoustic differences of the sounds may in part account for the differences in the FFR amplitude between speech and music. Although the frequency components of H1-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 1200-1300 Hz, respectively) with relatively small amplitudes of frequencies between them. This results in an acoustic waveform with a robust fundamental periodicity (Figure 1). On the other hand, a vibrating string produces a harmonically richer sound with the largest spectral peaks falling at the first through fifth harmonics (200-600 Hz). These harmonics interact to produce an acoustic waveform with a less salient periodicity at the fundamental (Figure 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 non-musicians. Alternatively, but less likely, we can speculate that brainstem structures exhibit a speech-encoding bias, perhaps due to 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 (Ahissar & Hochstein 2004). Our data corroborate the prediction of this theory that physiological changes correlate with the length of training. An alternative to the top-down hypothesis is that afferent peripheral structures exhibit Hebbian rules of plasticity (Hebb 1949). 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 (Musacchia, Sams, Nicol, & Kraus 2006a) 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 (Hyde & Knudsen 2001;Stein et al. 2001;Wallace, Meredith, & Stein 1998). Audiovisual interaction in the colliculi is thought to be accomplished primarily by corticofugal modulation (Hyde & Knudsen 2002). Whether visual stimuli and experience with multisensory stimuli modulate the human auditory brainstem response via feed-forward or corticofugal mechanisms is still unknown. The interconnectedness of afferent pathway (Popper 1992) as well as efferent anatomical projections from primary and nonprimary cortices to the inferior colliculus (Lim & Anderson 2007;Saldana, Feliciano, & Mugnaini 1996a;Schofield & Coomes 2005;Winer et al. 1998) provide the anatomical bases for either a corticofugal or feed-forward mechanism, respectively.

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 following complex stimulation, as seen in van Praag et al.'s work (2005), 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 (Trainor 2005) as well as the development of pitch, timbre and melody discrimination skills, which are present as early as 6 months of age (Trehub 2003), 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 (Abrams et al. 2006;Banai, Nicol, Zecker, & Kraus 2005;Wible, Nicol, & Kraus 2005). 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

29 adult subjects (mean age 25.6 ± 4.1 years, 14 females) with normal hearing (<15 dB pure-tone thresholds from 500 to 4000 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 (MAT). Subjects who were categorized as musicians (N=16) were self-identified, began playing an instrument before the age of five, had 10 or more years of musical experience, and practiced more than three times a week for four 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" (Klatt 1980), 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. 1A). Both acoustic sounds were 350 ms in length and shared the same (±

2 Hz) fundamental frequency (F0=100 Hz), first (H1=200 Hz), second (H2=300 Hz), third (H3=400 Hz), and fourth (H4=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 Software). 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-minute break between blocks (Neurobehavioral Systems Inc. 2001). This yielded 2400 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%).

Continuous electroencephalographic (EEG) data were recorded from Cz (10-20 International System, earlobe reference, forehead ground), offline filtered (70-2000 Hz), epoched and averaged to result in individual artifact-free averages of at least 2000 sweeps per stimulus type (music, speech) and condition (UA, UV, AV) (Compumedics, El Paso, TX, USA). Brainstem responses to UV stimuli resulted in neural activity that was indistinguishable from background non-stimulus activity, as has been shown in a previous report of visual influence on brainstem activity (Musacchia, Sams, Nicol, & Kraus 2006a). 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 2A), 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 in 40-ms overlapping epochs from 30 to 350 ms, using a sliding-window technique. 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 4-10 ms portion. Two-way repeated measures ANOVAs and Bonferroni corrected *post hoc* t-tests, when applicable, were employed with brainstem and Error% 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.

Acknowledgements: NSF 0544846 and NIH R01 DC01510 supported this work. The authors wish to thank Scott Lipscomb, members of the Auditory Neuroscience Laboratory at Northwestern University, and the subjects who participated in this experiment.

Chapter 3

In order to test our third aim to understand the relationship between subcortical and cortical activity, we analyzed the simultaneously recorded cortical data we had recorded from our musicians and non-musicians.

Manuscript - Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians

Gabriella Musacchia[1], Dana Strait[1], and Nina Kraus[1,2]

Auditory Neuroscience Laboratory, Departments of [1]Communication Sciences, [2]Neurobiology and Physiology and [3]Otolaryngology, Northwestern University, Evanston, IL, USA; [4]Laboratory of Computational Engineering, Helsinki University of Technology, Finland

Abstract

Musicians have a variety of perceptual and cortical specializations compared to non-musicians. Recent studies have shown that potentials evoked from primarily brainstem structures are enhanced in musicians, compared to non-musicians. Specifically, musicians have more robust representations of pitch periodicity and faster neural timing to sound onset when listening to sounds or both listening to and viewing a speaker. However, it is not known whether musician-related enhancements at the subcortical level are correlated with specializations in the cortex. Does musical training shape the auditory system in a coordinated manner or in disparate ways at cortical and subcortical levels? To answer this question, we recorded simultaneous brainstem and cortical evoked responses in musician and non-musician subjects. Brainstem response periodicity was related to early cortical response timing across all subjects, and this relationship was stronger in musicians. Peaks of the brainstem response evoked by sound onset and timbre cues were also related to cortical timing. Neurophysiological measures at both levels correlated with musical skill scores across all subjects. In addition, brainstem and cortical measures correlated with the age musicians began their training and the years of musical practice. Taken together, these data imply that neural representations of pitch, timing and timbre cues and cortical response timing are shaped in a coordinated manner, and indicate corticofugal modulation of subcortical afferent circuitry.

Introduction

Playing music is a cognitively complex task that requires, at minimum, sensations from the sound he or she is playing, the sight of sheet music and the touch of the instrument to be utilized and integrated. Proficiency at doing so accumulates over years of consistent training, even in cases of high innate talent. Not surprisingly, instrumental musicians exhibit behavioral and perceptual advantages over non-musicians in music-related areas such as pitch discrimination (Tervaniemi et al. 2005) and fine motor control skills (Kincaid, Duncan,

& Scott 2002). Musicians have also shown perceptual improvements over nonmusicians in both native and foreign linguistic domains (Magne, Schon, & Besson 2006;Marques, Moreno, Luis, & Besson 2007). It is thought that neural plasticity related to musical training underlies many of these differences (Hannon & Trainor 2007).

Highly-trained musicians exhibit anatomical, functional and event-related specializations compared to non-musicians. From an anatomical perspective, musicians have more neural cell bodies (grey matter volume) in auditory, motor and visual cortical areas of the brain (Gaser & Schlaug 2003b). Not surprisingly, professional instrumentalists, compared to amateurs or untrained controls, have more activation in auditory areas such as Heschel's gyrus (Schneider et al. 2002) and the planum temporale (Ohnishi, Matsuda, Asada, Aruga, Hirakata, Nishikawa, Katoh, & Imabayashi 2001) to sound. Musical training also promotes plasticity in somatosensory regions; with string players demonstrating larger areas of finger representation than untrained controls (Elbert et al. 1995). With regard to evoked potentials (EPs) thought to arise primarily from cortical structures, musicians show enhancements of the P1-N1-P2 complex to pitch, timing and timbre features of music, relative to non-musicians (Pantev, Roberts, Schulz, Engelien, & Ross 2001). Trained musicians show particularly large enhancements when listening to the instruments that they themselves play (Munte, Nager, Beiss, Schroeder, & Altenmuller 2003;Pantev, Ross, Fujioka, Trainor, Schulte, & Schulz 2003). Musicians' cortical EP measures are also more apt to register fine-grained changes in complex auditory patterns and are more sensitive to pitch and interval changes in a melodic contour than non-musicians ((Fujioka et al. 2004;Pantev, Ross, Fujioka, Trainor, Schulte, & Schulz 2003). Moreover, musician-related plasticity is implicated in these and other studies because enhanced cortical EP measures have been correlated to the length of musical training or musical skill.

Recent studies from our laboratory have suggested that playing a musical instrument also "tunes" neural activity peripheral to cortical structures (Musacchia et al. 2007a;Wong, Skoe, Russo, Dees, & Kraus 2007a). These studies showed that evoked responses thought to arise predominantly from brainstem structures were more robust in musicians than in nonmusician controls. The observed musician-related enhancements corresponded to stimulus features that may be particularly important for processing music. One such example is observed with the frequency following response (FFR), which is thought to be generated primarily in the inferior colliculus and consists of phase-locked inter-spike intervals occurring at the fundamental frequency (F0) of a sound (Hoormann, Falkenstein, Hohnsbein, & Blanke 1992;Krishnan, Xu, Gandour, & Cariani 2005). Because F0 is understood to underlie the percept of pitch, this response is hypothesized to be related to the ability to accurately encode acoustic cues for pitch. Enhanced encoding of this aspect of the stimulus would clearly be beneficial to pitch perception of music. Accordingly, our previous studies demonstrated larger peak amplitudes at F0 and better pitch tracking in musicians relative to nonmusicians. Another example was observed with Wave δ (~8 ms post-acoustic onset) of the brainstem response to sound onset, which has been hypothesized to be important for encoding stimulus onset (Musacchia, Sams, Nicol, & Kraus 2006a;Musacchia, Sams, Skoe, & Kraus 2007a). Stimulus onset is an attribute of music important for denoting instrument attack and rhythm, and therefore it is perhaps unsurprising that we observed earlier Wave δ responses in musicians than non-musicians. More importantly, FFR and Wave δ enhancement in musicians was observed with both music and speech stimuli and was largest when subjects engaged multiple senses by simultaneously lip-reading or watching a musician play. This suggests that while these enhancements may be motivated by music-related tasks, they are pervasive and apply to other stimuli that possess those stimulus characteristics.

A key point to be noted regarding prior EP studies showing musician-related enhancements is that none have attempted to relate enhancements in measures thought to arise from brainstem structures (e.g., the FFR) with measures thought to arise largely from cortical regions (e.g., P1, N1 and P2 potentials). One crucial piece of information that could be gleaned from this approach would be that we may be able to determine which stimulus features are relevant to cortical EP enhancements in musicians. Such determinations could be made because musician-related enhancements in brainstem responses correspond to representations of specific stimulus features (e.g. pitch, timing and timbre).

The implications of these data could be strengthened considerably if the EP data were also correlated with performance on music-related behavioral tasks. Previous work has suggested that short and long-term experience with complex auditory tasks (e.g. language, music, auditory training) may shape subcortical circuitry likely through corticofugal modulation of sensory function (Banai, Abrams, & Kraus 2007;Krishnan, Xu, Gandour, & Cariani 2005;Russo, Nicol, Zecker, Hayes, & Kraus 2005;Song et al. 2008). Correlations between measures of brainstem and cortical EPs that coincide with improved performance on a musical task would provide support for the notion that specific neural elements are recruited to perform a given task, and that such selections are mediated in a top-down manner through experience (e.g., Reverse Hierarchy Theory; Ahissar & Hochstein, 2004), presumably via reciprocal cortical-subcortical interactions. Although Reverse Hierarchy Theory (RHT) has been used to consider visual cortical function, it is our view that this mechanism applies to subcortical sensory processing and that the application of its principles can explain the malleability of early sensory levels.

The idea of a cognitive-sensory interplay between subcortical and cortical plasticity is not new, and theories of learning increasingly posit a co-operation between “bottom-up” ad “top-down” plasticity [for review, see (Kral & Eggermont 2007), and when reaction times to auditory stimuli are shorter (Galbraith et al. 2000). The FFR is also selectively activated when verbal stimuli are consciously perceived as speech Galbraith, 1997 480 /id} and is larger to a speech syllable than to a time-reversed version of itself (Galbraith et al. 2004). In addition, several lines of research suggest that subcortical activity is enhanced in people who have had protracted linguistic (Krishnan, Xu, Gandour, & Cariani 2005;Xu, Krishnan, & Gandour 2006)or musical training(Musacchia, Sams, Skoe, & Kraus 2007a;Wong, Skoe, Russo, Dees, & Kraus 2007a) and degraded in people with certain communication disorders (Banai, Nicol, Zecker, & Kraus 2005;Russo, Larson, & Kraus 2008). Malleability of the human brainstem response is not restricted to lifelong training, however, as short-term auditory training has also been shown to enhance the FFR in children and adults (Russo, Nicol, Zecker, Hayes, & Kraus 2005;Song, Skoe, Wong, & Kraus 2008). Physiological work in animals demonstrates that improved signal processing in subcortical structures is mediated by the corticofugal system during passive and active auditory exposure (Yan & Suga 1998;Zhou & Jen 2007). Prior anatomical findings suggest several potential routes that propagate action potentials from the auditory cortex to subcortical centers such as the medial geniculate body and inferior colliculus (IC) (Huffman & Henson, Jr. 1990;Kelly & Wong 1981;Saldana, Feliciano, & Mugnaini 1996b). Consistent with this notion of reciprocal cortical-subcortical interaction, the current work investigates the relationship between experience and the representation of stimulus features at the sensory and cortical level.

In order to examine the relationship between evoked-potentials and experience, we recorded simultaneous brainstem and cortical EPs in musicians and non-musician controls. Because previous data showed that musician-related effects extend to speech and multisensory stimuli, the speech syllable “da” was presented in three conditions: when subjects listened to the auditory sound alone, when the subjects simultaneously watched a video of a male speaker saying “da”, and when they viewed the video alone. Our analysis focused on comparing measures of the speech-evoked brainstem response that have been previously reported as enhanced in musicians with well-established measurements of cortical activity (e.g., the P1-N1-P2 complex). Thus, we were particularly interested in the representation of the timing of sound onset, pitch and timbre in the brainstem response. By correlating these neurophysiological measures and comparing them to behavioral scores on tests of musical skill and auditory perception, we were able to establish links between brainstem measures, cortical measures and behavioral performance and to show which relationships were strengthened by musical training.

Materials and Methods

Subjects

Participants in this study consisted of 26 adults (mean age 25.6 ± 4.1 years, 14 females) with normal hearing (<15 dB HL pure-tone thresholds from 500 to 4000 Hz). We assume that all listeners had similar audiometric profiles because we are unaware of any data suggesting that normal-hearing musicians have a different audiometric profile than normal-hearing non-musicians. Participants were selected to have normal or corrected-to-normal vision (Snellen Eye Chart, 2001) and no history of neurological disorders. All participants gave their informed consent before participating in this study in accordance with the Northwestern

University Institutional Review Board regulations. Subjects categorized as musicians (N=14) were self-identified, began playing an instrument before the age of five, had 10 or more years of musical experience, and practiced more than three times weekly for four or more hours consistently over the last 10 years. Controls (N=12) were categorized by the failure to meet the musician criteria.

Musical aptitude measures

We administered two in-house measures of auditory and musical skill: Seashore's Test of Musical Talents (Seashore 1919) and Colwell's Musical Achievement Test (MAT-3) (Colwell 1970). Seashore's test consists of six subtests: Pitch, Rhythm, Loudness, Time, Timbre and Tonal Memory. Each subtest is a two–alternative forced choice auditory discrimination task that asks listeners to judge whether the second sound (or sequence) is different from the first. Because of its use of pure and complex sine waves, and the method of evaluation, the Seashore battery of listening tests is widely-understood to measure basic psychoacoustic skills rather than actual musical aptitude. The MAT-3 consists of 5 subtests and was designed as an entrance exam for post-secondary instrumental students. Accordingly, some MAT-3 tests were too advanced for the nonmusicians.

We administered MAT-3 subtests of Tonal Memory and Solo Instrument Recognition (I) to all subjects. Musicians were also given MAT-3 tests of Melody Recognition, Polyphonic Chord Recognition and Ensemble Instrument Recognition (II). Introductory verbal instruction was provided at the start of each test and subtest, with musical examples for each question provided via a portable stereo system. Bivariate correlation tests among tests of musical skill and neurophysiological measures were conducted and independent t-tests between groups were conducted to determine the extent of musician-related differences.

Stimuli and recording procedure

Stimuli were presented binaurally via insert earphones (ER-3; Etymotic Research, Elk Grove Village, IL) while the subject sat in a comfortable chair centered 2.3 m from a 15.2 cm x 19.2 cm projection screen. The speech syllable “da” was presented in three conditions: 1) when subjects heard the sound alone and simultaneously watched a captioned video (A); 2) when, instead of a captioned movie, subjects viewed a video token of a male speaker saying “da” simultaneously (AV); and 3) when subjects viewed the video of the speaker without sound (V). The synthesized speech syllable (Klatt 1980) was 350 ms in duration with a fundamental frequency of 100 Hz. F1 and F2 of the steady state were 720 Hz and 1240 Hz, respectively. Video clips of a speaker’s face saying “da” were edited to 850 ms durations (FinalCut Pro 4, Apple Software). When auditory and visual stimuli were presented together, the sound onset occurred 460ms after the onset of the first video frame. The acoustic onset occurred synchronously with the visual release of consonant closure.

Stimuli were presented in 12 blocks of 600 stimulus repetitions with a 5-minute break between blocks (Neurobehavioral Systems Inc., 2001). Each block consisted of either A, V or AV stimuli, with modality of presentation order randomized across all subjects. Auditory stimuli were presented at 84 dB SPL in alternating polarities. This presentation level insured that the signal was clearly audible and well above threshold to all subjects. To control for attention, subjects were asked to silently count the number of target stimuli they saw or heard and to report that number at the end of each block. Target stimuli were slightly longer in duration than the standards (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 targets the subject missed (Error%).

General neurophysiology recording procedure

Electroencephalographic (EEG) data were recorded from Ag-AgCl scalp electrode Cz (10-20 International System, earlobe reference, forehead ground) with a filter passband of 0.5 to 2000 Hz and a sampling rate of 20 kHz (Compumedics, El Paso, TX, USA). Following acquisition, the EEG data were highpass and lowpass filtered offline to emphasize brainstem or cortical activity, respectively (see below).

Although there is ample evidence that generators in brainstem structures figure prominently in what we refer to as the “brainstem” and “cortical” responses, it is worthn noting that these far-field evoked potentials do not reflect the activity of brainstem or cortical structures exclusively. Because far-field responses record the sum of all neuroelectric activity, higher-level activity (e.g. thalamic, cortical) may be concomitantly captured to some degree in both the onset and FFR measures and vice-versa. Neural generators that contribute to the human brainstem response have been identified primarily through simultaneous surface and intracranial recordings of responses to clicks during neurosurgery (Hall 1992;Jacobson 1985). The cochlear nucleus, the superior olivary complex, the lateral lemniscus, and the inferior colliculi have been shown to predominantly contribute to the first five transient peaks (Waves I-V, ~1-6 ms post–acoustic onset) recorded from the scalp. Pure tones and complex sounds evoke the FFR which is thought to primarily reflect phase-locked activity from the inferior colliculus (Hoormann, Falkenstein, Hohnsbein, & Blanke 1992;Krishnan, Xu, Gandour, & Cariani 2005;Smith, Marsh, & Brown 1975b;Smith, Marsh, & Brown 1975a). Moreover, the FFR can emerge at latency of ~6 ms, which precedes the initial excitation of primary auditory cortex (~12 ms) (Celesia 1968;Moushegian, Rupert, & Stillman 1973). Finally, and perhaps most convincingly, cryogenic cooling of the IC greatly decreases or eliminates the FFR (Smith, Marsh, & Brown 1975b). Despite this evidence, it is possible that evoked FFR activity may reflect concomitant cortical activity after cortical regions have been activated (e.g. ~12 ms). At longer latencies, the FFR most likely reflects a mix of afferent brainstem activity, cortically modulated efferent effects, and synchronous cortical activity. According to these data and for the sake of parsimony and accord with previous studies, we utilize the terms “brainstem” and “cortical” in this study to denote high-and low-pass filtered EP responses, respectively.

Brainstem response analysis

After acquisition, a highpass filter of 70 Hz was applied to the EEG data. Typically, this type of passband is employed to emphasize the relatively fast and high-frequency neural activity of putative brainstem structures. After filtering, the data were epoched from -100 to 450 ms, relative to acoustic onset. A rejection criterion of ± 35 μV was applied to the epoched file so that responses containing high myogenic or extraneous activity above or below the criterion were excluded. The first 2000 epochs that were not artifact-rejected from each condition (A, V, AV) were then averaged for each individual. We then assessed measures of the brainstem response that reflect stimulus features that have been shown to differ between musicians and non-musicians. Brainstem onset response peak, Wave γ, was picked from each individual’s responses, yielding latency and amplitude information. The FFR portion of the brainstem response was submitted to a fast Fourier transform (FFT). Strength of pitch encoding was measured by peak amplitudes at F0 (100 Hz) and timbre representation by peak amplitudes at harmonics H2 (200 Hz), H3 (300 Hz), H4 (400 Hz), and H5 (500 Hz) as picked by an automatic peak-detection program. Because we assessed measures that have previously been shown to differ between musicians and non-musicians, we used one-tailed independent t-tests to assess group differences in brainstem response measures.

Cortical response analysis

EEG data were lowpass filtered offline at 40 Hz. This passband is employed to emphasize the relatively slow and low-frequency neural activity of putative cortical origin. Responses were epoched and averaged with an artifact rejection criterion of ± 65 μV and the first 2000 artifact-free sweeps were averaged in each condition. Cortical response peaks (P1, N1, P2 and N2) were chosen from each subject's averages, providing amplitude and latency information. Strength of neural synchrony in response to a given stimulus was assessed by P1-N1 and P2-N2 peak-to-peak slopes.

Description of brainstem and cortical responses

The brainstem response to a speech syllable mimics stimulus characteristics with high fidelity (Johnson et al. 2005;Kraus & Nicol 2005;Russo et al. 2004). The beginning portion of the brainstem response to speech (~0-30 ms) encodes the onset of sound in a series of peaks, the first 5 of which are analogous to responses obtained in hearing clinics with click or tone stimuli (e.g. Waves I-V) (Hood 1998). With this stimulus, a large peak is also typically observed at ~8-12 ms, called Wave γ (Musacchia et al. 2006;Musacchia et al. 2007). Other laboratories have demonstrated similar relationships between the temporal characteristics of tonal stimuli in the human brainstem response (Akhoun et al. 2008;Galbraith, Arbagey, Branski, Comerci, & Rector 1995;Galbraith, Amaya, de Rivera, Donan, Duong, Hsu, Tran, & Tsang 2004;Galbraith, Jhaveri, & Kuo 1997;Galbraith, Olfman, & Huffman 2003;Krishnan, Xu, Gandour, & Cariani 2005). In the current study, we restricted our peak latency and amplitude analyses to Wave γ because it was the only brainstem peak to sound onset that previously differed between musicians and non-musicians. The voiced portion of the speech syllable evokes an FFR, which reflects neural phase-locking to the stimulus F0. Figure 9 shows the grand average brainstem responses of musicians and non-musicians in A and AV conditions. The grand average FFTs are shown in insets. Grand average cortical responses are shown in Figure 10. Speech stimuli, presented in either the A or AV condition, elicited four sequential peaks of alternating positive and negative polarity and are labeled P1, N1, P2, and N2, respectively. As is typically observed in cortical responses to sound, these components occurred within ~75-250 ms post-acoustic stimulation (Hall 1992). To investigate relationships between musical training and brainstem and cortical processing, Pearson's *r* correlations were run between all measures of musicianship and brainstem and cortical responses.

Results

Differences between musicians and non-musicians

As has been shown in previous studies, musicians had more robust encoding of speech periodicity in the FFR. Musicians had larger F0 peak amplitudes, in both the A ($t = 2.33$, $p = 0.02$) and AV conditions ($t = 2.42$, $p = 0.01$), compared to non-musicians. Group differences were also observed on measures of timbre representation ($t_{H3} = 2.00$, $p = 0.03$; $t_{H4} = 1.784$, $p = 0.045$; $t_{H5} = 1.767$, $p = 0.045$) and onset timing ($t_{\delta \text{Latency}} = 1.95$, $p = 0.03$) in the AV condition.

Overall, P1 and N1 peaks were earlier and larger in the musician group (Figure 10). Musicians had larger amplitudes at P1 in the A condition ($t = 2.106$, $p = 0.046$) and at N1 in the AV condition ($t = 2.099$, $p = 0.047$). P1-N1 slope, our measure of early aggregate cortical timing, was steeper in musicians compared to non-musicians for both the A ($t = 2.90$, $p = 0.01$) and AV conditions ($t = 5.01$, $p < 0.001$). Later timing components, as measured by P2-N2 slope, did not differ between groups.

Relationships between ABR/FFR measures and P1-N2 peaks

Among ABR and FFR measures that differ between musicians and non-musicians, periodicity encoding correlated with measures of P1-N1timing most consistently (Table 3, Figure 11). Across all subjects, larger F0 peak amplitudes of the brainstem response were associated with steeper cortical P1-N1 slopes in both A ($r = -0.47$, $p = 0.02$) and AV ($r = -0.50$, $p = 0.01$) conditions. An ANOVA between the slopes of the regression lines across the two groups showed a significant difference in the A condition ($F = 8.61$, $p < 0.01$). Correlations between brainstem periodicity encoding and early cortical timing were stronger in musicians ($r = -0.70$, $p = 0.01$) than non-musicians ($r = 0.13$, $p = 0.68$). Although the ANOVA test for the correlation data did not reach statistical significance in the AV condition, this same trend was observed. F0 amplitude in the A condition also correlated with measures of later cortical peaks, P2 and N2 ($r_{P2} = -0.49$, $p = 0.01$: $r_{N2} = -0.44$, $p = 0.02$), such that larger F0 amplitudes were associated with steeper slopes. These correlations were not observed in the AV condition.

Peak amplitudes of FFT harmonics reflect encoding of acoustic cues for timbre perception. In the AV condition, H3 peak amplitude correlated with P2 latency ($r = -0.40$, $p = 0.04$) and H4 peak amplitude correlated with N2 latency ($r = 0.42$, $p = 0.03$) (Table 4). Table 5 shows that P1-N1 slope in the audiovisual condition and N2 peak latency in the auditory alone condition correlated positively with brainstem onset timing (Wave δ latency). That is, smaller P1-N1 slope and later latency values correlate with earlier brainstem onset responses.

Relationships between behavior and EP measures

P1-N1 slope related to perceptual measures of tonal memory from the MAT-3 and Seashore tests (Table 6). In both tests, subjects were presented with two successive sequences of tones and asked to choose which tone was different in the second sequence they heard. The Seashore test presented pure tones, while the MAT-3 consisted of musical notes played on the piano. Standardized tonal memory scores correlated with P1-N1 slope measures for both tests in both modalities (A: $r_{\text{MAT-3}} = -0.425$, $p = 0.03$; AV: $r_{\text{MAT-3}} = -0.49$, $p = 0.01$; $r_{\text{SEA}} = -0.47$, $p = 0.02$). Musicians had stronger correlations ($r = -0.52$, $p=0.06$) between Seashore tonal memory scores and P1-N1 slopes in the AV condition ($F = 4.99$, $p < 0.05$) than their non-musician counterparts ($r = 0.11$, $p = 0.74$). Overall, musicians had higher standardized scores on both tests than non-musicians (MAT-3: $M_{\text{musicians}} = 3.14$, $M_{\text{non-musicians}} = -3.65$, $t = 4.50$ $p < 0.001$; Seashore: $M_{\text{musicians}} = 2.66$, $M_{\text{non-musicians}} = -3.10$, $t = 3.44$, $p < 0.001$).

Less consistent correlations between neurophysiological and behavioral measures were observed between Wave δ latency and Seashore's loudness subtest ($r = -0.41$, $p = 0.04$), as well as between H2 peak amplitude and Seashore's test of timbre discrimination ($r = 0.47$, $p = 0.02$) in the A condition (Table 7). The strength of these correlations did not differ across groups.

Relationships between evoked potentials and extent of musical training

F0 amplitude and P1-N1 slope correlated with consistent musical practice while FFT measures harmonic representation correlated with the age that musicians began their training. Consistent practice among musicians was measured by the self-reported number of years, within the last ten, each player practiced his or her instrument (> 3 times per week for > 2 hours per day). This measure of musical training strongly correlated with F0 amplitude and P1-N1 slope in both modalities (Table 8, Figure 12). More years of consistent musical practice was associated with larger F0 peak amplitudes in both conditions ($r_{AA} = 0.78$, $p = 0.001$; $r_{AV} = 0.72$, $p = 0.003$). Similarly, more years of consistent practice was associated with steeper P1-N1 slopes in the A condition ($r = -0.68$, $p = 0.007$). The age that musicians began playing correlated negatively with timbre representation, as measured by H3 and H4 peak amplitude in the A condition (Table 8). That is, earlier beginning age was associated with larger harmonic peak amplitudes ($r_{H3} = -0.54$, $p = 0.047$; $r_{H4} = -0.63$, $p = 0.02$).

Discussion

Musician-related plasticity and corticofugal modulation

The first picture that emerges from our data is that recent musical training improves one's auditory memory and shapes composite (P1-N1) and pitch-specific encoding (F0) in a co-coordinated manner. Our EP and behavior correlations suggest that complex auditory task performance is related to the strength of the P1-N1 response. Both the Seashore and MAT-3 Tonal Memory tests require listeners to hold a sequence of pitches in memory and identify pitch differences in a second sequence. Scores from both tests correlated with P1-N1 slopes in both A and AV modalities such that steeper slopes were associated with higher scores. Not surprisingly, these measures are affected by musicianship: instrumental musicians performed better on the tests and had steeper P1-N1 slopes than non-musicians. Our P1-N1 results corroborate previous work showing that that musical training is associated with earlier and larger P1-N1 peaks (Fujioka et al. 2004b).

However, it was not only the individual tests and measures that were musician-related. Musicians had a statistically stronger *correlation* between this set of brain and behavior measures than non-musicians. While it is well-known that trained musicians outperform untrained controls and have more robust evoked-potentials than non19 musicians, our data show that the accord, or relationship, between brain and behavior is also improved in musicians. Our data steer us one step further, however. Because steeper P1-N1 slopes are associated with more years of musical training, we can speculate that the accord between brain and behavior is strengthened with consistent years of musical training.

Interestingly, variance in the P1-N1 slope measure is also explained by peak amplitude of the fundamental frequency in the FFR across all subjects. This indicates that robust, frequency-specific representations of a sound's pitch are vital to later, composite measures of neural activity. F0 amplitude, like P1-N1 slope, also varies positively with years of consistent musical training. Taken together, the P1-N1, FFR, and Tonal Memory correlations imply that the high cognitive demand of consistent musical training improves auditory acuity and shapes composite and frequency-specific encoding in a coordinated manner.

We can interpret these data in terms of corticofugal mechanisms of plasticity. Playing music involves tasks with high cognitive demands, such as playing one's part in a musical ensemble, as well as detailed auditory acuity, such as monitoring ones intonation while playing the part. It is conceivable that the demand for complex organization and simultaneously detail-oriented information engages cortical mechanisms that are capable of refining the neural code at a basic sensory level. This idea is consistent with models of perceptual learning that involve perceptual weighting with feedback {Nosofsky, 1986 495 /id}. In this case, attention to pitch-relevant cues would increase the perceptual weighting of these dimensions. Positive and negative feedback in the form of harmonious pitch cues and auditory beats could shift the weighting system to represent the F0 more faithfully. Our theory also comports with the Reverse Hierarchy

Theory (RHT) of visual learning {Ahissar, 2001 382 /id}. The RHT suggests that goal-oriented behavior shapes neural circuitry in “reverse” along the neural hierarchy. Applied to our data, this would suggest that the goal of accurately holding successive pitches in auditory memory would first tune complex encoding mechanisms (e.g. cortical), followed by a “backward” search for increased signal-to-noise ratios of pitch related features in sensory systems (e.g. brainstem). Indeed, this interpretation has been invoked by Kraus and colleagues to interpret subcortical changes in subcortical function associated with short-term training and lifelong language and music experience in language-compromised, typical listeners and auditory experts (e.g. Kraus and Banai 2007; Banai et al. 2007; Song et al. in press; Wong et al. 2007; Musacchia et al. 2007). Finally, recent models also suggest that topdown guided plasticity may be mediated by sensory-specific memory systems. Instead of being generated by prefrontal and parietal memory systems, it is thought that sensory memory is directly linked to the sensory system used to encode the information (Pasternak & Greenlee 2005). In this way, enhancements at the sensory encoding level would increase the probability of creating accurate sensory memory traces.

With respect to our other evoked-potential measures, the second concept to emerge is the relationship between auditory discrimination of fine-grained stimulus features, such as timbre, the neural representation of those features in the FFR and the age at which musical training began. Seashore’s test of timbre is a two-alternative forced choice procedure that asks subjects to discriminate whether a second sound differs (in perceived timbre tonality) from the first. Timbre is widely understood to be the sound quality which can distinguish sounds with the same pitch and loudness (e.g., the quality a trumpet versus a violin). Acoustic differences such as harmonic context and sound rise time give rise to this perception (Erickson 1978). In contrast to the previous case where behavior was linked with later, cortical EPs, behavioral scores on timbre discrimination were directly related to harmonic components of the FFR. Specifically, larger H3 and H4 amplitudes were associated with better timbre scores across all subjects. However, like the relationship between behavior and later cortical peak components, the representation of harmonics does seem to simply distinguish musicians as a group from non-musicians because the amplitude of H3 and H4 was positively correlated with the age at which musical training began.

One interpretation of these data is that tasks requiring auditory discrimination of subtle stimulus features depend more heavily upon stimulus-specific encoding mechanisms. Consistent with theories of corticofugal modulation, it is possible that cognitive demands of timbre discrimination tasks progressively tune sensory encoding mechanisms related to harmonic representation. In this case, lifelong experience distinguishing between instruments may strengthen the direct link between the sensory representation of harmonic frequencies and the perception that they subserve. It is important to note that cortical EPs are not completely bypassed in timbre perception. H3 and H4 amplitude in the A condition correlate to P2 and N2 peak latency, respectively. Consequently, these cortical response components may be related to the encoding of these subtle stimulus features. Perhaps this is one of the reasons that timbre discrimination, anecdotally, takes longer to perceive than the pitch of a note. A similar type of mechanism may underlie the correlation between Wave □ latency and loudness discrimination ability. However, the functional relationship between response to sound onset and the perception of a sound’s amplitude is less transparent, although response timing is a common neural reflection of sound intensity (Jacobson 1985).

The continuum between expert and impaired experience

The current study shows how extensive musical training strengthens the relationship between measures of putatively low- and high-levels of neural encoding. On the other end of the experience continuum, previous data in school-aged children indicate that the strength of these relationships can be weakened in the language-impaired system. In normal-learning children, Wible and colleagues demonstrated a relationship between brainstem response timing and cortical response fidelity to signals presented in background noise, which learning-impaired (LI) children fail to show (Wible et al. 2005). The normal pattern of hemispheric asymmetry to speech was also disrupted in LI children with brainstem response abnormalities (Abrams et al. 2006). In addition, children with brainstem response timing deficits showed reduced cortical sensitivity to acoustic change (Banai et al. 2005). Taken together with those findings in language-impaired systems, the current findings suggest a continuum of cohesive brainstem/cortical association that can be disrupted in impaired populations and strengthened by musical training.

Conclusion

Overall, our data indicate that the effects of musical experience on the nervous system include relationships between brainstem and cortical EPs recorded simultaneously in the same subject to seen and heard speech. Moreover, these relationships were related to behavioral measures of auditory perception and were stronger in the audiovisual condition. This implies that musical training promotes plasticity throughout the auditory and multisensory pathways. This includes encoding mechanisms that are relevant for musical sounds as well as for the processing of linguistic cues and multisensory information. This is in line with previous work which has shown that experience which engages cortical activity (language, music, auditory training) shapes subcortical circuitry, likely through corticofugal modulation of sensory function. That is, brainstem activity is affected by lifelong language expertise (Krishnan 2005), its disruption (reviewed in Banai et al. 2007) and music experience (Musacchia et al. 2007;Wong et al. 2007) as well as by short term training (Russo et al. 2005;Song et al. 2008;Russo 2005). Consistent with this notion of reciprocal cortical-subcortical interaction, the current work shows a relationship between sensory representation of stimulus features and cortical peaks. Specifically, we find that musical training tunes stimulus feature-specific (e.g. onset response/FFR) and composite (e.g. P1-N2) encoding of auditory and multi-sensory stimuli in a coordinated manner. We propose that the evidence for corticofugal mechanisms of plasticity [e.g. (Suga and Ma, 2003)] as well as the theories that these data drive (Ahissar, 2004), combined with theories of music acquisition and training [e.g. (Hannon and Trainor, 2007)] provide a theoretical framework for our findings. Further research is needed to determine directly how top-down or bottom-up mechanisms may contribute to music-related plasticity in the cortical/subcortical auditory pathway axis. Experiments, such as recording the time course of brainstem-cortical interactions could prove to be especially fruitful in this area.

Acknowledgements: NSF 0544846 and NIH R01 DC01510 supported this work. The authors wish to thank Scott Lipscomb, Ph.D. for his musical background, Matthew Fitzgerald, Ph.D. and Trent Nicol for their critical commentary, and the subjects who participated in this experiment.

Conclusion

The data presented here show that musical training shapes how our subcortical and cortical systems encode multimodal and linguistic information. Specifically, Chapter 2 shows that professional musicians with extensive training have an enhanced representation of periodicity in the brainstem and more robust cortical response timing to seen and heard speech and musical stimuli, compared non-musicians. One of the main challenges to this work was developing the method to record visual influence on human brainstem activity, which is detailed in Chapter 1.

In addition, Chapter 3 describes a strong relationship between brainstem measures of periodicity and measures of early cortical response timing. Measurements of the brainstem response, described in Appendix 1, were not pervasively related to cortical responses. Instead, we show, specifically, that the representation of the fundamental frequency at lower levels translates to an increase in aggregate neural timing at the cortex. The correlation between these features of brainstem and cortical encoding was significantly stronger in our musicians, implying more efficient communication between high and low-level encoding in this group. Future investigations are needed to determine whether musical training strengthens the link between brainstem and cortical encoding in a top-down or bottom-up manner.

Also from Chapter 3, we conclude that protracted musical practice, and not innate musical disposition, promotes neural plasticity. In musicians, strong correlations were observed between years of musical experience measures of encoding at both brainstem and cortical levels. This suggests that the fidelity of encoding at brainstem and cortical levels is strengthened by practicing a musical instrument consistently.

At the opposite end of the spectrum, we show that degraded auditory input, in the form of hearing loss, is associated with degraded neural representation of auditory and audiovisual speech stimuli. Data from Appendix 2 show that visual influence on auditory processing was significantly less pronounced in older adults with hearing impairment. This indicates diminished audiovisual interaction compared to their normal-hearing counterparts. The clinical implication of these findings is that hearing loss not only impacts the neural response to auditory stimuli, but also how multi-modal stimuli are encoded.

A primary conclusion from this work is that learning-related plasticity associated with musical training generalizes to speech encoding. Musicians had particularly enhanced brainstem responses to auditory and audiovisual linguistic cues. Whereas speech and music are known to be left- and right-lateralized, respectively, in the cortex, here we show shared neural resources at the subcortical level.

Taken together, our work suggests multimodal experience impacts subcortical and cortical encoding mechanisms as well as the relationship between them. A continuum between neural integrity and the extent of multi-modal input, which is high in musicians and low in hearing-impaired individuals, is also implied. Future investigations are needed to determine which aspects of musical training, such as length of training, cognitive demand, attentional focus or multi-modal stimulation, contribute significantly to neural response enhancement. It may be that plasticity results as a combination of these factors. Even without knowing which aspects of musical training specifically contribute to enhanced neural activity, we can conclude that it is beneficial to the CANS. This information may be useful in designing training paradigms for people with communication difficulties and, in particular, implies that musical training may be an effective remediation tool for those with auditory or multi-sensory encoding deficits.

Abrams, D. A., Nicol, T., Zecker, S. G., & Kraus, N. 2006, "Auditory brainstem timing predicts cerebral asymmetry for speech", *J.Neurosci.*, vol. 26, no. 43, pp. 11131-11137.

Ahissar, M. 2001, "Perceptual training: a tool for both modifying the brain and exploring it", *Proc.Natl.Acad.Sci.U.S.A*, vol. 98, no. 21, pp. 11842-11843.

Ahissar, M. & Hochstein, S. 2004, "The reverse hierarchy theory of visual perceptual learning", *Trends Cogn Sci.*, vol. 8, no. 10, pp. 457-464.

Akhoun, I., Gallego, S., Moulin, A., Menard, M., Veuillet, E., Berger-Vachon, C., Collet, L., & Thai-Van, H. 2008, "The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults", *Clin Neurophysiol*, vol. 119, no. 4, pp. 922-933.

Armstrong, B. A., Neville, H. J., Hillyard, S. A., & Mitchell, T. V. 2002, "Auditory deprivation affects processing of motion, but not color", *Brain Res.Cogn Brain Res.*, vol. 14, no. 3, pp. 422-434.

Banai, K., Abrams, D., & Kraus, N. 2007, "Sensory-based learning disability: Insights from brainstem processing of speech sounds", *Int.J.Audiol.*, vol. 46, no. 9, pp. 524-532.

Banai, K., Nicol, T., Zecker, S. G., & Kraus, N. 2005, "Brainstem timing: implications for cortical processing and literacy", *J.Neurosci.*, vol. 25, no. 43, pp. 9850-9857.

Bavelier, D., Tomann, A., Hutton, C., Mitchell, T., Corina, D., Liu, G., & Neville, H. 2000, "Visual attention to the periphery is enhanced in congenitally deaf individuals", *J.Neurosci.*, vol. 20, no. 17, p. RC93.

Bergan, J. F., Ro, P., Ro, D., & Knudsen, E. I. 2005, "Hunting increases adaptive auditory map plasticity in adult barn owls", *J.Neurosci.*, vol. 25, no. 42, pp. 9816-9820.

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. 2004, "Bimodal speech: early suppressive visual effects in human auditory cortex", *Eur.J.Neurosci.*, vol. 20, no. 8, pp. 2225-2234.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., & Possing, E. T. 2000, "Human temporal lobe activation by speech and nonspeech sounds", *Cereb.Cortex*, vol. 10, no. 5, pp. 512-528.

Bradlow, A. R., Kraus, N., Nicol, T. G., Mcgee, T. J., Cunningham, J., Zecker, S. G., & Carrell, T. D. 1999, "Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children", *J.Acoust.Soc.Am.*, vol. 106, no. 4 Pt 1, pp. 2086-2096.

Brochard, R., Dufour, A., & Despres, O. 2004, "Effect of musical expertise on visuospatial abilities: evidence from reaction times and mental imagery", *Brain Cogn*, vol. 54, no. 2, pp. 103-109.

Burnett, L. R., Stein, B. E., Chaponis, D., & Wallace, M. T. 2004, "Superior colliculus lesions preferentially disrupt multisensory orientation", *Neuroscience*, vol. 124, no. 3, pp. 535-547.

Burton, M. W., Small, S. L., & Blumstein, S. E. 2000, "The role of segmentation in phonological processing: an fMRI investigation", *J.Cogn Neurosci.*, vol. 12, no. 4, pp. 679-690.

Bushara, K. O., Grafman, J., & Hallett, M. 2001, "Neural correlates of auditory-visual stimulus onset asynchrony detection", *J.Neurosci.*, vol. 21, no. 1, pp. 300-304.

Bushara, K. O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., & Hallett, M. 2003, "Neural correlates of cross-modal binding", *Nat.Neurosci.*, vol. 6, no. 2, pp. 190-195.

Calvert, G. A. 2001, "Crossmodal processing in the human brain: insights from functional neuroimaging studies", *Cereb.Cortex*, vol. 11, no. 12, pp. 1110-1123.

Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S. 1999, "Response amplification in sensory-specific cortices during crossmodal binding", *Neuroreport*, vol. 10, no. 12, pp. 2619-2623.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., & David, A. S. 1997, "Activation of auditory cortex during silent lipreading", *Science*, vol. 276, no. 5312, pp. 593-596.

Calvert, G. A., Campbell, R., & Brammer, M. J. 2000, "Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex", *Curr.Biol.*, vol. 10, no. 11, pp. 649-657.

Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., Brammer, M. J., & David, A. S. 2001, "Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning)", *Brain Res.Cogn Brain Res.*, vol. 12, no. 2, pp. 233-243.

Celesia, G. G. 1968, "Auditory evoked responses. Intracranial and extracranial average evoked responses", *Arch.Neurol.*, vol. 19, no. 4, pp. 430-437.

Chomsky, N. 1985, *The Logical Structure of Linguistic Theory* The University of Chicago Press, Chicago, IL.

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. 2001, "Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement", *Clin.Neurophysiol.*, vol. 112, no. 5, pp. 758-767.

Dalton, D. S., Cruickshanks, K. J., Klein, B. E. K., Klein, R., Wiley, T. L., & Nondahl, D. M. 2003, "The impact of hearing loss on quality of life in older adults", *Gerontologist*, vol. 43, no. 5, pp. 661-668.

Doupe, A. J. 1997, "Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development", *J.Neurosci.*, vol. 17, no. 3, pp. 1147-1167.

Drake, C. & Palmer, C. 1993, "Accent Structures in Music Performance", *Music Perception*, vol. 10, no. 3, pp. 343-378.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. 1995, "Increased cortical representation of the fingers of the left hand in string players", *Science*, vol. 270, no. 5234, pp. 305-307.

Erber, N. 1972, "Auditory, visual and auditory-visual recognitionof consonants by children with normal and impaired hearing", *Journal of Speech and Hearing Research*, vol. 15, pp. 413-422.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. 2004, "Musical training enhances automatic encoding of melodic contour and interval structure", *J.Cogn Neurosci.*, vol. 16, no. 6, pp. 1010-1021.

Galbraith, G. C., Amaya, E. M., de Rivera, J. M., Donan, N. M., Duong, M. T., Hsu, J. N., Tran, K., & Tsang, L. P. 2004, "Brain stem evoked response to forward and reversed speech in humans", *Neuroreport*, vol. 15, no. 13, pp. 2057-2060.

Galbraith, G. C., Arbagey, P. W., Branski, R., Comerci, N., & Rector, P. M. 1995, "Intelligible speech encoded in the human brain stem frequency-following response", *Neuroreport*, vol. 6, no. 17, pp. 2363-2367.

Galbraith, G. C., Chae, B. C., Cooper, J. R., Gindi, M. M., Ho, T. N., Kim, B. S., Mankowski, D. A., & Lunde, S. E. 2000, "Brainstem frequency-following response and simple motor reaction time", *Int.J.Psychophysiol.*, vol. 36, no. 1, pp. 35-44.

Galbraith, G. C. & Doan, B. Q. 1995, "Brainstem frequency-following and behavioral responses during selective attention to pure tone and missing fundamental stimuli", *Int.J.Psychophysiol.*, vol. 19, no. 3, pp. 203-214.

Galbraith, G. C., Jhaveri, S. P., & Kuo, J. 1997, "Speech-evoked brainstem frequency-following responses during verbal transformations due to word repetition", *Electroencephalogr.Clin Neurophysiol*, vol. 102, no. 1, pp. 46-53.

Galbraith, G. C., Olfman, D. M., & Huffman, T. M. 2003, "Selective attention affects human brain stem frequency-following response", *Neuroreport*, vol. 14, no. 5, pp. 735-738.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. 1996, "Action recognition in the premotor cortex", *Brain*, vol. 119 (Pt 2), pp. 593-609.

Gardi, J., Merzenich, M., & McKean, C. 1979, "Origins of the scalp recorded frequency-following response in the cat", *Audiology*, vol. 18, no. 5, pp. 358-381.

Gaser, C. & Schlaug, G. 2003c, "Brain structures differ between musicians and non-musicians", *J.Neurosci.*, vol. 23, no. 27, pp. 9240-9245.

Gaser, C. & Schlaug, G. 2003a, "Brain structures differ between musicians and non-musicians", *Journal of Neuroscience*, vol. 23, no. 27, pp. 9240-9245.

Gaser, C. & Schlaug, G. 2003b, "Brain structures differ between musicians and non-musicians", *Journal of Neuroscience*, vol. 23, no. 27, pp. 9240-9245.

Giard, M. H. & Peronnet, F. 1999, "Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study", *J.Cogn Neurosci.*, vol. 11, no. 5, pp. 473-490.

Giraud, A. L., Price, C. J., Graham, J. M., Truy, E., & Frackowiak, R. S. 2001, "Cross-modal plasticity underpins language recovery after cochlear implantation", *Neuron*, vol. 30, no. 3, pp. 657-663.

Grant, K. W. & Seitz, P. F. 2000, "The use of visible speech cues for improving auditory detection of spoken sentences", *J.Acoust.Soc.Am.*, vol. 108, no. 3 Pt 1, pp. 1197-1208.

Green, K. P. 1987, "The perception of speaking rate using visual information from a talker's face", *Percept.Psychophys.*, vol. 42, no. 6, pp. 587-593.

Hall, J. W. I. I. 1992, *Handbook of Auditory Evoked Responses* Allyn and Bacon, Needham Heights, MA .

Hannon, E. E. & Trainor, L. J. 2007, "Music acquisition: effects of enculturation and formal training on development", *Trends Cogn Sci*, vol. 11, no. 11, pp. 466-472.

Hauser, M. D., Chomsky, N., & Fitch, W. T. 2002, "The faculty of language: what is it, who has it, and how did it evolve?", *Science*, vol. 298, no. 5598, pp. 1569-1579.

Hayes, E. A., Warrier, C. M., Nicol, T. G., Zecker, S. G., & Kraus, N. 2003, "Neural plasticity following auditory training in children with learning problems", *Clin.Neurophysiol.*, vol. 114, no. 4, pp. 673-684.

Hebb, D. O. 1949, *The Organization of Behavior* Wiley, New York.

Heil, P., Bronchti, G., Wollberg, Z., & Scheich, H. 1991, "Invasion of visual cortex by the auditory system in the naturally blind mole rat", *Neuroreport*, vol. 2, no. 12, pp. 735-738.

Hillyard, S. A., Woldorff, M., Mangun, G. R., & Hansen, J. C. 1987, "Mechanisms of early selective attention in auditory and visual modalities", *Electroencephalogr.Clin.Neurophysiol.Suppl*, vol. 39, pp. 317-324.

Hodges, D. A., Hairston, W. D., & Burdette, J. H. 2005, "Aspects of multisensory perception: the integration of visual and auditory information in musical experiences", *Ann.N.Y.Acad.Sci.*, vol. 1060, pp. 175-185.

Hood, L. J. 1998, *Clinical Applications of the Auditory Brainstem Response* Singular, San Diego.

Hoormann, J., Falkenstein, M., & Hohnsbein, J. 1994, "Effect of selective attention on the latency of human frequency-following potentials", *Neuroreport*, vol. 5, no. 13, pp. 1609-1612.

Hoormann, J., Falkenstein, M., Hohnsbein, J., & Blanke, L. 1992, "The human frequency-following response (FFR): normal variability and relation to the click-evoked brainstem response", *Hear.Res.*, vol. 59, no. 2, pp. 179-188.

Huffman, R. F. & Henson, O. W., Jr. 1990, "The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus", *Brain Res.Brain Res.Rev.*, vol. 15, no. 3, pp. 295-323.

Humes, L. E. & Roberts, L. 1990, "Speech-recognition difficulties of the hearing-impaired elderly: the contributions of audibility", *J.Speech Hear.Res.*, vol. 33, no. 4, pp. 726-735.

Hyde, P. S. & Knudsen, E. I. 2002, "The optic tectum controls visually guided adaptive plasticity in the owl's auditory space map", *Nature*, vol. 415, no. 6867, pp. 73-76.

Hyde, P. S. & Knudsen, E. I. 2001, "A topographic instructive signal guides the adjustment of the auditory space map in the optic tectum", *J.Neurosci.*, vol. 21, no. 21, pp. 8586-8593.

Jacobson, J. T. 1985, *The Auditory Brainstem Response* College-Hill Press, San Diego.

Jancke, L., Schlaug, G., & Steinmetz, H. 1997, "Hand skill asymmetry in professional musicians", *Brain Cogn*, vol. 34, no. 3, pp. 424-432.

Jiang, W., Jiang, H., & Stein, B. E. 2006, "Neonatal cortical ablation disrupts multisensory development in superior colliculus", *J.Neurophysiol*, vol. 95, no. 3, pp. 1380-1396.

Jiang, W. & Stein, B. E. 2003, "Cortex controls multisensory depression in superior colliculus", *J.Neurophysiol.*, vol. 90, no. 4, pp. 2123-2135.

Jing, H. & Benasich, A. A. 2006, "Brain responses to tonal changes in the first two years of life", *Brain Dev.*, vol. 28, no. 4, pp. 247-256.

Johnson, K. L., Nicol, T. G., & Kraus, N. 2005, "Brain stem response to speech: a biological marker of auditory processing", *Ear Hear.*, vol. 26, no. 5, pp. 424-434.

Kelly, J. P. & Wong, D. 1981, "Laminar connections of the cat's auditory cortex", *Brain Res.*, vol. 212, no. 1, pp. 1-15.

Kent, R. D. 1984, "Psychobiology of speech development: coemergence of language and a movement system", *Am.J.Physiol*, vol. 246, no. 6 Pt 2, p. R888-R894.

Kincaid, A. E., Duncan, S., & Scott, S. A. 2002, "Assessment of fine motor skill in musicians and nonmusicians: differences in timing versus sequence accuracy in a bimanual fingering task", *Percept.Mot.Skills*, vol. 95, no. 1, pp. 245-257.

King, C., Warrier, C. M., Hayes, E., & Kraus, N. 2002, "Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems", *Neurosci.Lett.*, vol. 319, no. 2, pp. 111-115.

Klatt, D. H. 1980, "Software for a cascade/parallel formant synthesizer", *J.Acoust.Soc.Am.*, vol. 67, no. 3, pp. 971-995.

Klucharev, V., Mottonen, R., & Sams, M. 2003, "Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception", *Cogn Brain Res.*, vol. 18, no. 1, pp. 65-75.

Klucharev, V. & Sams, M. 2004, "Interaction of gaze direction and facial expressions processing: ERP study", *Neuroreport*, vol. 15, no. 4, pp. 621-625.

Korte, M. & Rauschecker, J. P. 1993, "Auditory spatial tuning of cortical neurons is sharpened in cats with early blindness", *J.Neurophysiol.*, vol. 70, no. 4, pp. 1717-1721.

Kral, A. & Eggermont, J. J. 2007, "What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity", *Brain Res.Rev.*, vol. 56, no. 1, pp. 259-269.

Kraus, N. & Nicol, T. 2005, "Brainstem origins for cortical 'what' and 'where' pathways in the auditory system", *Trends Neurosci.*, vol. 28, no. 4, pp. 176-181.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. 2005, "Encoding of pitch in the human brainstem is sensitive to language experience", *Brain Res.Cogn Brain Res.*, vol. 25, no. 1, pp. 161-168.

Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. 2006, "Enhanced multisensory integration in older adults", *Neurobiol.Aging*, vol. 27, no. 8, pp. 1155-1163.

Lim, H. H. & Anderson, D. J. 2007, "Antidromic activation reveals tonotopically organized projections from primary auditory cortex to the central nucleus of the inferior colliculus in guinea pig", *J.Neurophysiol*, vol. 97, no. 2, pp. 1413-1427.

Linkenkaer-Hansen, K., Palva, J. M., Sams, M., Hietanen, J. K., Aronen, H. J., & Ilmoniemi, R. J. 1998, "Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by mag", *Neurosci.Lett.*, vol. 253, no. 3, pp. 147-150.

Lu, S. T., Hamalainen, M. S., Hari, R., Ilmoniemi, R. J., Lounasmaa, O. V., Sams, M., & Vilkman, V. 1991, "Seeing faces activates three separate areas outside the occipital visual cortex in man", *Neuroscience*, vol. 43, no. 2-3, pp. 287-290.

MacDonald, J. & McGurk, H. 1978, "Visual influences on speech perception processes", *Percept.Psychophys.*, vol. 24, no. 3, pp. 253-257.

MacLeod, A. & Summerfield, Q. 1987, "Quantifying the contribution of vision to speech perception in noise", *Br.J.Audiol.*, vol. 21, no. 2, pp. 131-141.

Magne, C., Schon, D., & Besson, M. 2006, "Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches", *J.Cogn Neurosci.*, vol. 18, no. 2, pp. 199-211.

Marks, L. E. 2004, "Cross-Modal Interactions in Speeded Classification," in *The Handbook of Multisensory Processes*, G. Calvert, C. Spence, & B. E. Stein, eds., MIT Press, Cambridge.

Marks, L. E. 1982, "Bright sneezes and dark coughs, loud sunlight and soft moonlight", *J.Exp.Psychol.Hum.Percept.Perform.*, vol. 8, no. 2, pp. 177-193.

Marques, C., Moreno, S., Luis, C. S., & Besson, M. 2007, "Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence", *J.Cogn Neurosci.*, vol. 19, no. 9, pp. 1453-1463.

Marsh, J. T., Brown, W. S., & Smith, J. C. 1975, "Far-field recorded frequency-following responses: correlates of low pitch auditory perception in humans", *Clin.Neurophysiol.*, vol. 38, pp. 113-119.

Massaro, D. W. 1998, *Perceiving talking faces: From speech perception to a behavioral principle* MIT Press, Cambridge, MA.

Massaro, D. W. & Cohen, M. M. 1983, "Evaluation and integration of visual and auditory information in speech perception", *J.Exp.Psychol.Hum.Percept.Perform.*, vol. 9, no. 5, pp. 753-771.

McGurk, H. & MacDonald, J. 1976, "Hearing lips and seeing voices", *Nature*, vol. 264, no. 5588, pp. 746-748.

Meredith, M. A. & Stein, B. E. 1986a, "Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration", *J.Neurophysiol.*, vol. 56, no. 3, pp. 640-662.

Meredith, M. A. & Stein, B. E. 1986b, "Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration", *J.Neurophysiol.*, vol. 56, no. 3, pp. 640-662.

Moller, A. R. 1999, "Neural mechanisms of BAEP", *Electroencephalogr.Clin Neurophysiol Suppl*, vol. 49, pp. 27-35.

Moore, B. C. J. 2003, *Introduction to the psychology of hearing* Academic, London.

Mottonen, R., Krause, C. M., Tiippana, K., & Sams, M. 2002, "Processing of changes in visual speech in the human auditory cortex", *Brain Res.Cogn Brain Res.*, vol. 13, no. 3, pp. 417-425.

Moushegian, G., Rupert, A. L., & Stillman, R. D. 1973, "Laboratory note. Scalp-recorded early responses in man to frequencies in the speech range", *Electroencephalogr.Clin Neurophysiol*, vol. 35, no. 6, pp. 665-667.

Munte, T. F., Nager, W., Beiss, T., Schroeder, C., & Altenmuller, E. 2003, "Specialization of the specialized: electrophysiological investigations in professional musicians", *Ann.N.Y.Acad.Sci.*, vol. 999, pp. 131-139.

Musacchia, G., Sams, M., Nicol, T., & Kraus, N. 2006b, "Seeing speech affects acoustic information processing in the human brainstem", *Experimental Brain Research*, vol. 168, no. 1-2, pp. 1-10.

Musacchia, G., Sams, M., Nicol, T., & Kraus, N. 2006a, "Seeing speech affects acoustic information processing in the human brainstem", *Exp.Brain Res.*, vol. 168, no. 1-2, pp. 1-10.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. 2007a, "Musicians have enhanced subcortical auditory and audiovisual processing of speech and music", *Proc.Natl.Acad.Sci.U.S.A.*

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. Musicians have enhanced auditory and audiovisual brainstem responses to speech and music. Proc.Natl.Acad.Sci.U.S.A . 2007b.
Ref Type: In Press

Musiek, F. & Berge, B. 1998, "CAPD: Mostly Management," in *Central Auditory Processing Disorders: Mostly Management*, J. Katz, ed., Williams & Wilkins, Baltimore, pp. 15-31.

Nishitani, N. & Hari, R. 2002, "Viewing lip forms: cortical dynamics", *Neuron*, vol. 36, no. 6, pp. 1211-1220.

Oates, P. A., Kurtzberg, D., & Stapells, D. R. 2002, "Effects of sensorineural hearing loss on cortical event-related potential and behavioral measures of speech-sound processing", *Ear and Hearing*, vol. 23, no. 5, pp. 399-415.

Oatman, L. C. & Anderson, B. W. 1977, "Effects of visual attention on tone burst evoked auditory potentials", *Exp.Neurol.*, vol. 57, no. 1, pp. 200-211.

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., Katoh, A., & Imabayashi, E. 2001, "Functional anatomy of musical perception in musicians", *Cereb.Cortex*, vol. 11, no. 8, pp. 754-760.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. 1998, "Increased auditory cortical representation in musicians", *Nature*, vol. 392, no. 6678, pp. 811-814.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. 2001, "Timbre-specific enhancement of auditory cortical representations in musicians", *Neuroreport*, vol. 12, no. 1, pp. 169-174.

Pantev, C., Ross, B., Fujioka, T., Trainor, L. J., Schulte, M., & Schulz, M. 2003, "Music and learning-induced cortical plasticity", *Ann.N.Y.Acad.Sci.*, vol. 999, pp. 438-450.

Peretz, I. & Zatorre, R. J. 2005, "Brain organization for music processing", *Annu.Rev.Psychol.*, vol. 56, pp. 89-114.

Perrault, T. J., Jr., Vaughan, J. W., Stein, B. E., & Wallace, M. T. 2003, "Neuron-specific response characteristics predict the magnitude of multisensory integration", *J.Neurophysiol,* vol. 90, no. 6, pp. 4022-4026.

Picton, T. W. & Hillyard, S. A. 1974, "Human auditory evoked potentials. II. Effects of attention", *Electroencephalogr.Clin.Neurophysiol.*, vol. 36, no. 2, pp. 191-199.

Popper, R. F. 1992, *The Mammalian Auditory Pathway: Neurophysiology.* Springer, New York .

Rauschecker, J. P. & Korte, M. 1993, "Auditory compensation for early blindness in cat cerebral cortex", *J.Neurosci.*, vol. 13, no. 10, pp. 4538-4548.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. 1993, "Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys", *J.Neurosci.*, vol. 13, no. 1, pp. 87-103.

Roder, B., Stock, O., Bien, S., Neville, H., & Rosler, F. 2002, "Speech processing activates visual cortex in congenitally blind humans", *Eur.J.Neurosci.*, vol. 16, no. 5, pp. 930-936.

Russo, N., Larson, C., & Kraus, N. 2008, "Audio-vocal system regulation in children with autism spectrum disorders", *Exp.Brain Res.*

Russo, N., Nicol, T., Musacchia, G., & Kraus, N. 2004, "Brainstem responses to speech syllables", *Clin.Neurophysiol.*, vol. 115, no. 9, pp. 2021-2030.

Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. 2005, "Auditory training improves neural timing in the human brainstem", *Behav.Brain Res.*, vol. 156, no. 1, pp. 95-103.

Saito, D. N., Yoshimura, K., Kochiyama, T., Okada, T., Honda, M., & Sadato, N. 2005, "Cross-modal binding and activated attentional networks during audio-visual speech integration: a functional MRI study", *Cereb.Cortex*, vol. 15, no. 11, pp. 1750-1760.

Saldana, E., Feliciano, M., & Mugnaini, E. 1996a, "Distribution of descending projections from primary auditory neocortex to inferior colliculus mimics the topography of intracollicular projections", *J.Comp Neurol.*, vol. 371, no. 1, pp. 15-40.

Saldana, E., Feliciano, M., & Mugnaini, E. 1996b, "Distribution of descending projections from primary auditory neocortex to inferior colliculus mimics the topography of intracollicular projections", *J.Comp Neurol.*, vol. 371, no. 1, pp. 15-40.

Saldana, H. M. & Rosenblum, L. D. 1993, "Visual influences on auditory pluck and bow judgments", *Percept.Psychophys.*, vol. 54, no. 3, pp. 406-416.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., & Simola, J. 1991, "Seeing speech: visual information from lip movements modifies activity in the human auditory cortex", *Neurosci.Lett.*, vol. 127, no. 1, pp. 141-145.

Schlaug, G., Jancke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. 1995, "Increased corpus callosum size in musicians", *Neuropsychologia*, vol. 33, no. 8, pp. 1047-1055.

Schmithorst, V. J. & Holland, S. K. 2004, "The effect of musical training on the neural correlates of math processing: a functional magnetic resonance imaging study in humans", *Neurosci.Lett.*, vol. 354, no. 3, pp. 193-196.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. 2002, "Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians", *Nat.Neurosci.*, vol. 5, no. 7, pp. 688-694.

Schofield, B. R. & Coomes, D. L. 2005, "Auditory cortical projections to the cochlear nucleus in guinea pigs", *Hear Res.*, vol. 199, no. 1-2, pp. 89-102.

Schon, D., Magne, C., & Besson, M. 2004, "The music of speech: music training facilitates pitch processing in both music and language", *Psychophysiology*, vol. 41, no. 3, pp. 341-349.

Scott, S. K. & Johnsrude, I. S. 2003, "The neuroanatomical and functional organization of speech perception", *Trends Neurosci.*, vol. 26, no. 2, pp. 100-107.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. 2003, "Auditory-visual speech perception examined by fMRI and PET", *Neurosci.Res.*, vol. 47, no. 3, pp. 277-287.

Smith, J. C., Marsh, J. T., & Brown, W. S. 1975b, "Far-field recorded frequency-following responses: evidence for the locus of brainstem sources", *Electroencephalogr.Clin Neurophysiol*, vol. 39, no. 5, pp. 465-472.

Smith, J. C., Marsh, J. T., & Brown, W. S. 1975a, "Far-field recorded frequency-following responses: evidence for the locus of brainstem sources", *Electroencephalogr.Clin Neurophysiol Suppl* no. 39, pp. 465-472.

Song, J. H., Skoe, E., Wong, P. C., & Kraus, N. 2008, "Plasticity in the Adult Human Auditory Brainstem following Short-term Linguistic Training", *J.Cogn Neurosci.*

Stein, B. E., Huneycutt, W. S., & Meredith, M. A. 1998, "Neurons and Behavior: The same rules of multisensory integration apply", *Brain Res.*, vol. 448, pp. 355-358.

Stein, B. E., Jiang, W., Wallace, M. T., & Stanford, T. R. 2001, "Nonvisual influences on visual-information processing in the superior colliculus", *Prog.Brain Res.*, vol. 134, pp. 143-156.

Stein, B. E., Wallace, M. W., Stanford, T. R., & Jiang, W. 2002, "Cortex governs multisensory integration in the midbrain", *Neuroscientist.*, vol. 8, no. 4, pp. 306-314.

Suga, N. & Ma, X. 2003, "Multiparametric corticofugal modulation and plasticity in the auditory system", *Nat.Rev.Neurosci.*, vol. 4, no. 10, pp. 783-794.

Suga, N., Xiao, Z., Ma, X., & Ji, W. 2002, "Plasticity and corticofugal modulation for hearing in adult animals", *Neuron*, vol. 36, no. 1, pp. 9-18.

Sumby, W. H. & Pollack, I. 1954, "Visual contribution to speech intelligibility in noise.", *J.Acoust.Soc.Am.*, vol. 26, pp. 212-215.

Summerfield, Q. 1979, "Use of visual information for phonetic perception", *Phonetica*, vol. 36, no. 4-5, pp. 314-331.

Teder-Salejarvi, W. A., McDonald, J. J., Di, R. F., & Hillyard, S. A. 2002, "An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings", *Brain Res.Cogn Brain Res.*, vol. 14, no. 1, pp. 106-114.

Tervaniemi, M. & Hugdahl, K. 2003, "Lateralization of auditory-cortex functions", *Brain Res.Brain Res.Rev.*, vol. 43, no. 3, pp. 231-246.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schroger, E. 2005, "Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study", *Exp.Brain Res.*, vol. 161, no. 1, pp. 1-10.

Thompson, R. F. 1986, "The neurobiology of learning and memory", *Science*, vol. 233, no. 4767, pp. 941-947.

Thompson, W. F., Graham, P., & Russo, F. A. 2005, "Seeing music performance: Visual influences on perceptiopn and experience", *Semiotica* no. 156, pp. 203-227.

Thompson, W. F., Schellenberg, E. G., & Husain, G. 2003, "Perceiving prosody in speech. Effects of music lessons", *Ann.N.Y.Acad.Sci.*, vol. 999, pp. 530-532.

Trainor, L. J. 2005, "Are there critical periods for musical development?", *Dev.Psychobiol.*, vol. 46, no. 3, pp. 262-278.

Trehub, S. E. 2003, "The developmental origins of musicality", *Nat.Neurosci.*, vol. 6, no. 7, pp. 669-673.

Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. 1997, "Central auditory system plasticity: generalization to novel stimuli following listening training", *J.Acoust.Soc.Am.*, vol. 102, no. 6, pp. 3762-3773.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. 2001, "Central auditory plasticity: changes in the N1-P2 complex after speech-sound training", *Ear Hear.*, vol. 22, no. 2, pp. 79-90.

Tremblay, K. L., Billings, C., & Rohila, N. 2004, "Speech evoked cortical potentials: effects of age and stimulus presentation rate", *J.Am.Acad.Audiol.*, vol. 15, no. 3, pp. 226-237.

Tremblay, K. L. & Kraus, N. 2002, "Auditory training induces asymmetrical changes in cortical neural activity", *J.Speech Lang Hear.Res.*, vol. 45, no. 3, pp. 564-572.

Tremblay, K. L., Piskosz, M., & Souza, P. 2003, "Effects of age and age-related hearing loss on the neural representation of speech cues", *Clin.Neurophysiol.*, vol. 114, no. 7, pp. 1332-1343.

Tye-Murray, N., Sommers, M. S., & Spehar, B. 2007, "Audiovisual integration and lipreading abilities of older adults with normal and impaired hearing", *Ear Hear.*, vol. 28, no. 5, pp. 656-668.

van Wassenhove, V., Grant, K. W., & Poeppel, D. 2005, "Visual speech speeds up the neural processing of auditory speech", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 4, pp. 1181-1186.

Van, P. H., Kempermann, G., & Gage, F. H. 2000, "Neural consequences of environmental enrichment", *Nat.Rev.Neurosci.*, vol. 1, no. 3, pp. 191-198.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. 2005, "Cross-modal interactions in the perception of musical performance", *Cognition*.

Wallace, M. T., Meredith, M. A., & Stein, B. E. 1998, "Multisensory integration in the superior colliculus of the alert cat", *J.Neurophysiol.*, vol. 80, no. 2, pp. 1006-1010.

Wallace, M. T., Meredith, M. A., & Stein, B. E. 1993, "Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus", *J.Neurophysiol*, vol. 69, no. 6, pp. 1797-1809.

Wallace, M. T., Perrault, T. J., Jr., Hairston, W. D., & Stein, B. E. 2004b, "Visual experience is necessary for the development of multisensory integration", *J.Neurosci.*, vol. 24, no. 43, pp. 9580-9584.

Wallace, M. T., Perrault, T. J., Jr., Hairston, W. D., & Stein, B. E. 2004a, "Visual experience is necessary for the development of multisensory integration", *J.Neurosci.*, vol. 24, no. 43, pp. 9580-9584.

Watkins, K. & Paus, T. 2004, "Modulation of motor excitability during speech perception: the role of Broca's area", *J.Cogn Neurosci.*, vol. 16, no. 6, pp. 978-987.

Wible, B., Nicol, T., & Kraus, N. 2004, "Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems", *Biol.Psychol.*, vol. 67, no. 3, pp. 299-317.

Wible, B., Nicol, T., & Kraus, N. 2005, "Correlation between brainstem and cortical auditory processes in normal and language-impaired children", *Brain*, vol. 128, no. Pt 2, pp. 417-423.

Winer, J. A., Larue, D. T., Diehl, J. J., & Hefti, B. J. 1998, "Auditory cortical projections to the cat inferior colliculus", *J.Comp Neurol.*, vol. 400, no. 2, pp. 147-174.

Woldorff, M., Hansen, J. C., & Hillyard, S. A. 1987, "Evidence for effects of selective attention in the mid-latency range of the human auditory event-related potential", *Electroencephalogr.Clin.Neurophysiol.Suppl*, vol. 40, pp. 146-154.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. 2007a, "Musical experience shapes human brainstem encoding of linguistic pitch patterns", *Nat.Neurosci.*, vol. 10, no. 4, pp. 420-422.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. 2007b, "Musical experience shapes human brainstem encoding of linguistic pitch patterns", *Nat.Neurosci.*, vol. 10, no. 4, pp. 420-422.

Xu, Y., Krishnan, A., & Gandour, J. T. 2006, "Specificity of experience-dependent pitch representation in the brainstem", *Neuroreport*, vol. 17, no. 15, pp. 1601-1605.

Yan, W. & Suga, N. 1998, "Corticofugal modulation of the midbrain frequency map in the bat auditory system", *Nat.Neurosci.*, vol. 1, no. 1, pp. 54-58.

Yvert, B., Fischer, C., Guenot, M., Krolak-Salmon, P., Isnard, J., & Pernier, J. 2002, "Simultaneous intracerebral EEG recordings of early auditory thalamic and cortical activity in human", *Eur.J.Neurosci.*, vol. 16, no. 6, pp. 1146-1150.

Zatorre, R. J. 1998b, "Functional specialization of human auditory cortex for musical processing", *Brain*, vol. 121, pp. 1817-1818.

Zatorre, R. J. 1998a, "Functional specialization of human auditory cortex for musical processing", *Brain*, vol. 121 (Pt 10), pp. 1817-1818.

Zhou, X. & Jen, P. H. 2007, "Corticofugal modulation of multi-parametric auditory selectivity in the midbrain of the big brown bat", *J.Neurophysiol*, vol. 98, no. 5, pp. 2509-2516.

Tables

Chapter 1 Tables

Table 1. Wave γ latency (ms)

Subject	Unimodal Acoustic	AV Concordant	AV Conflicting	UA+UV Concordant	UA+UV Conflicting
1	10.70	10.65	10.45	10.65	10.25
2	10.10	10.20	10.00	9.90	10.20
3	10.95	11.15	11.15	11.05	10.85
4	11.05	11.30	11.00	11.25	11.20
5	11.25	11.90	12.00	11.25	11.10
6	9.40	11.10	11.10	8.85	8.60
7	9.65	11.55	11.45	10.20	9.50
8	10.35	10.45	10.50	10.85	10.65
9	9.80	12.90	13.20	9.25	9.35
10	10.35	13.75	13.60	10.30	10.35
Mean	10.3600	11.4950	11.4450	10.3550	10.2050
Std. Dev.	.6271	1.1067	1.1741	.8258	.8321

Table 1. Individual Wave γ latencies (ms) in Unimodal Acoustic, $AV_{Concordant}$, $AV_{Conflicting}$ and the sum of Unimodal Acoustic + Unimodal Visual responses.

Table 2. Onset RMA (μV)

Subject	Unimodal Acoustic	AV Concordant	AV Conflicting
1	.21	.20	.21
2	.22	.17	.18
3	.51	.30	.34
4	.35	.20	.19
5	.18	.17	.22
6	.20	.10	.10
7	.12	.16	.21
8	.26	.17	.20
9	.23	.24	.21
10	.32	.15	.21
Mean	.26	.19	.21
Std. Dev.	.11	.05	.06

Table 2. RMA (μV) of individual onset responses in Unimodal Acoustic, $AV_{Concordant}$ and $AV_{Conflicting}$ conditions.

		FFR Periodicity Encoding	
		A F0 Amplitude	AV F0 Amplitude
Cortcial	P1-N1 Slope	-0.47*	-0.50**
	P2 Latency	0.24	-0.49*
	N2 Latency	0.21	-0.44*

Table 3 (In manuscript, Table 1). Pearson correlation coefficients for relationships between measures of FFR periodicity and late EP measures in all subjects (*p < 0.05, ** p<0.01).

		FFR Harmonic Encoding	
		AV H3 Amplitude	A H4 Amplitude
Cortcial	P1-N1 Slope	-0.27	-0.05
	P2 Latency	-0.40*	0.21
	N2 Latency	-0.36	0.42*

Table 4 (In manuscript, Table 2). Pearson correlation coefficients for relationships between FFR harmonic encoding and late EP measures across all subjects (*p < 0.05, ** p<0.01).

		ABR Onset Timing	
		A Delta Latency	AV Delta Latency
Cortcial	P1-N1 Slope	0.01	0.51**
	P2 Latency	0.3	0.26
	N2 Latency	0.50**	0.18

Table 5 (In manuscript, Table 3). Pearson correlation coefficients for relationships between peaks of the ABR to sound onset and late EP measures across all subjects (*p < 0.05, ** p<0.01).

		Cortical	
		A P1-N1 slope	AV P1-N1 slope
Test Scores	Loudness	0.18	-0.16
	Timbre	0.15	0.16
	SEA Tonal Mem	-0.34	-0.47*
	MAT Tonal Mem	-0.43*	-0.50*

Table 6 (In manuscript, Table 4). Pearson correlation coefficients for relationships between late EP measures and perceptual scores across all subjects (*p < 0.05, ** p<0.01).

		Brainstem Response	
		Timing	Harmonic
		A Delta Latency	A H2 Amplitude
Test Scores	Loudness	-0.41*	0.25
	Timbre	-0.04	0.47*
	SEA Tonal Mem	-0.37	0.23
	MAT Tonal Mem	-0.36	0.11

Table 7 (In manuscript, Table 5). Pearson correlation coefficients for relationships between brainstem response measures and perceptual scores across all subjects (*p < 0.05, ** p<0.01).

	ABR/FFR					Cortical
	Timing	Periodicity		Harmonics		Timing
	AV Delta Latency	A F0 Amplitude	AV F0 Amplitude	AA H3 Amplitude	AA H4 Amplitude	AA P1-N1 slope
Age Began	0.27	-0.41	-0.26	-0.60*	-0.63*	0.37
Musical Practice	-0.72**	0.79**	0.72**	0.38	0.40	-0.68*

Table 8 (In manuscript, Table 6). Pearson correlation coefficients for relationships between EP measures and musical training across all subjects (*$p < 0.05$, ** $p<0.01$).

Figures

Figure 1.

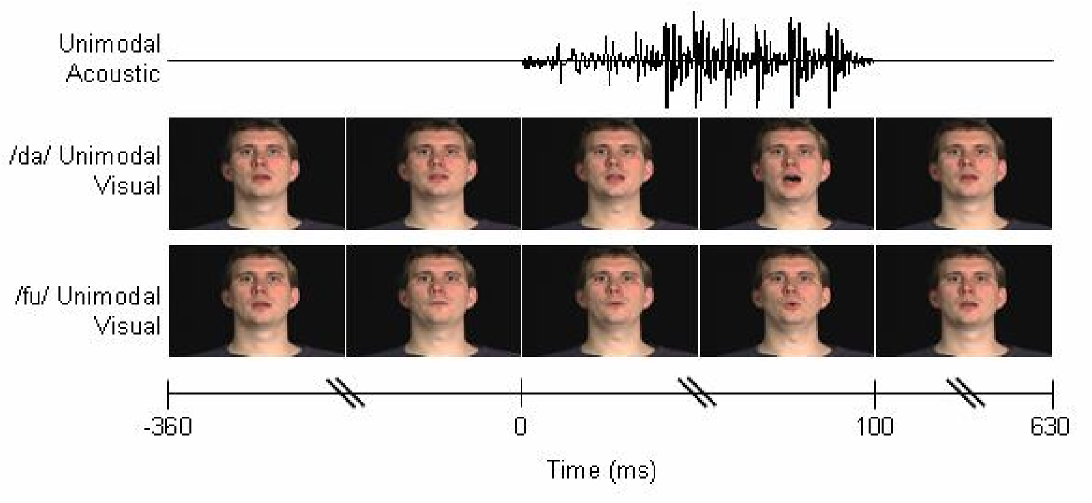

Figure 1. Acoustic and visual stimuli. Compressed timelines of two visual stimuli and the uncompressed acoustic stimulus are shown. Each unimodal visual utterance (/da/, /fu/ and /du/) was digitized from a recording of a male speaker. All three clips began and ended with the same neutral frame, but differed over the length of the utterance. The release of the consonant was edited to occur at frame 11 for all three visual tokens. For audiovisual presentation, the speech stimulus was paired with each visuofacial movement and acoustic onset occurred at time 0.

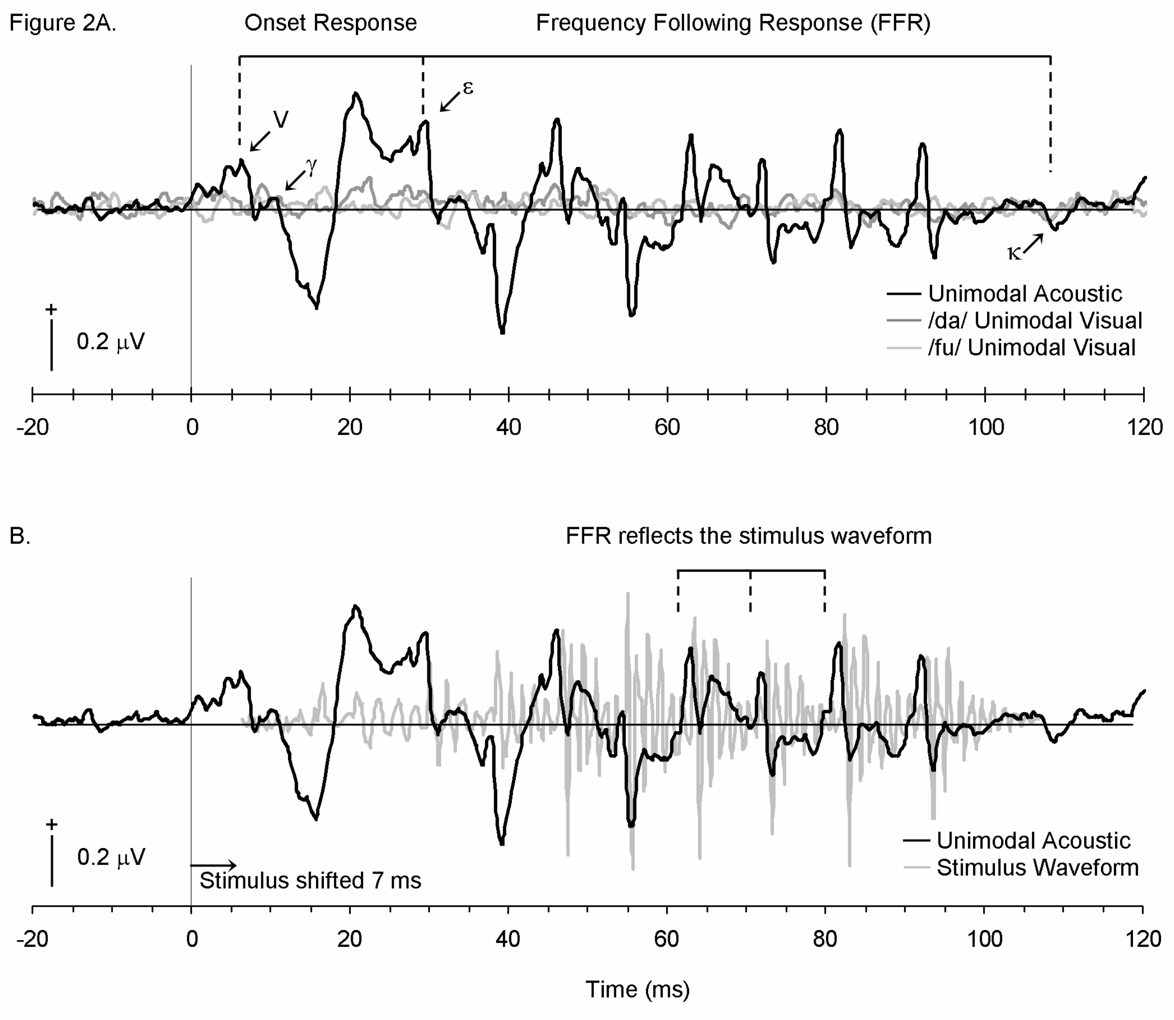

Figure 2. Stimulus waveform and unimodal grand average responses. Time 0 = acoustic stimulus onset. A. Prominent peaks of the Unimodal Acoustic response (black) to speech onset include Wave V followed by a positive deflection called Wave γ. The periodic portion of the response, the frequency following response, beginning at Wave ε and ending at Wave κ, is the region in which time between peaks reflects phase-locking to the stimulus waveform. Replicable waves were not observed in the unimodal visual /da/ (dark gray) or /fu/ (light gray) conditions. B. The grand average unimodal acoustic response is overlayed on the stimulus waveform. The onset of the stimulus has been shifted in time to correspond to response onset. Peaks of the periodic portion of the stimulus waveform can be seen to correspond to peaks of the frequency following response.

Figure 3.

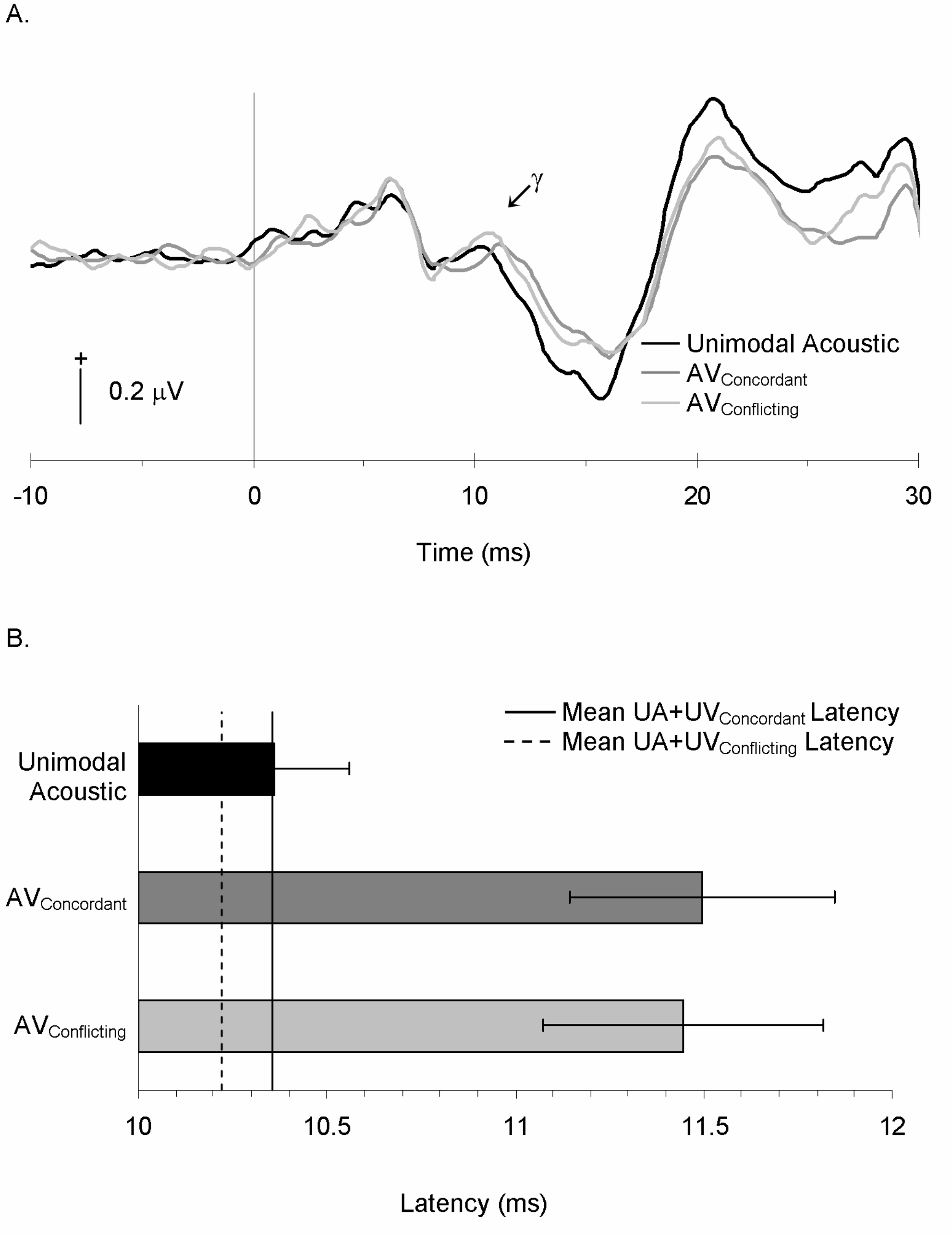

Figure 3. Onset responses in unimodal acoustic and the two audiovisual conditions. A. Grand average onset responses to UA (black), $AV_{Concordant}$ (dark gray) and $AV_{Conflicting}$ (light gray) are shown. The size of both AV responses is noticeably smaller than that of the UA response from approximately 10 to 30ms. Wave γ latency was prolonged, relative to the Unimodal Acoustic latency in both $AV_{Concordant}$ and $AV_{Conflicting}$ conditions. B. Mean Wave γ latencies are shown for UA and the two AV responses. Error bars show the standard error.

Figure 4.

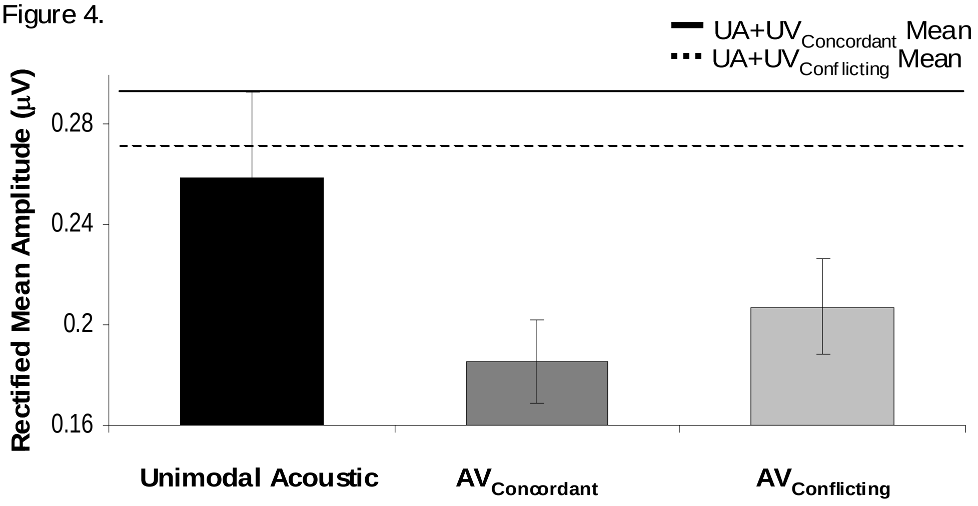

Figure 4. Unimodal acoustic and audiovisual onset response magnitude. The rectified mean amplitude (RMA, μV) of the Unimodal Acoustic response over the onset region (Wave V to ε) was larger than both the AV$_{Concordant}$ and AV$_{Conflicting}$ responses. Audiovisual RMA values were smaller than their computed counterparts (as indicated by lines) and the AV$_{Concordant}$ response was smaller than that of the AV$_{Conflicting}$.

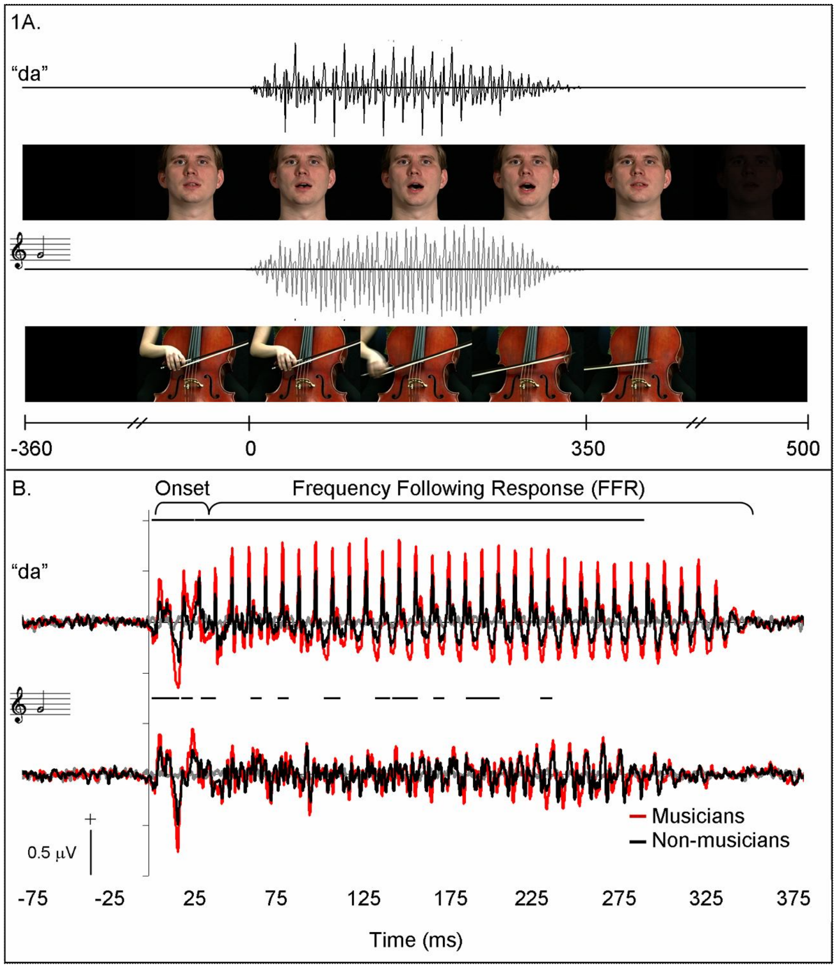

Figure 5 (In manuscript, Fig. 1). A. Auditory and visual components of speech and music stimuli. Visual components were digitized from videos of a speaker uttering "da" and a musician bowing a G note on the cello. Acoustic onset for both speech and music occurred 350 ms after the first video frame and 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 (top) and cello (bottom) stimuli. Group amplitude differences were assessed 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. Unimodal visual speech and music stimuli elicited little activity, as indicated by the grey traces.

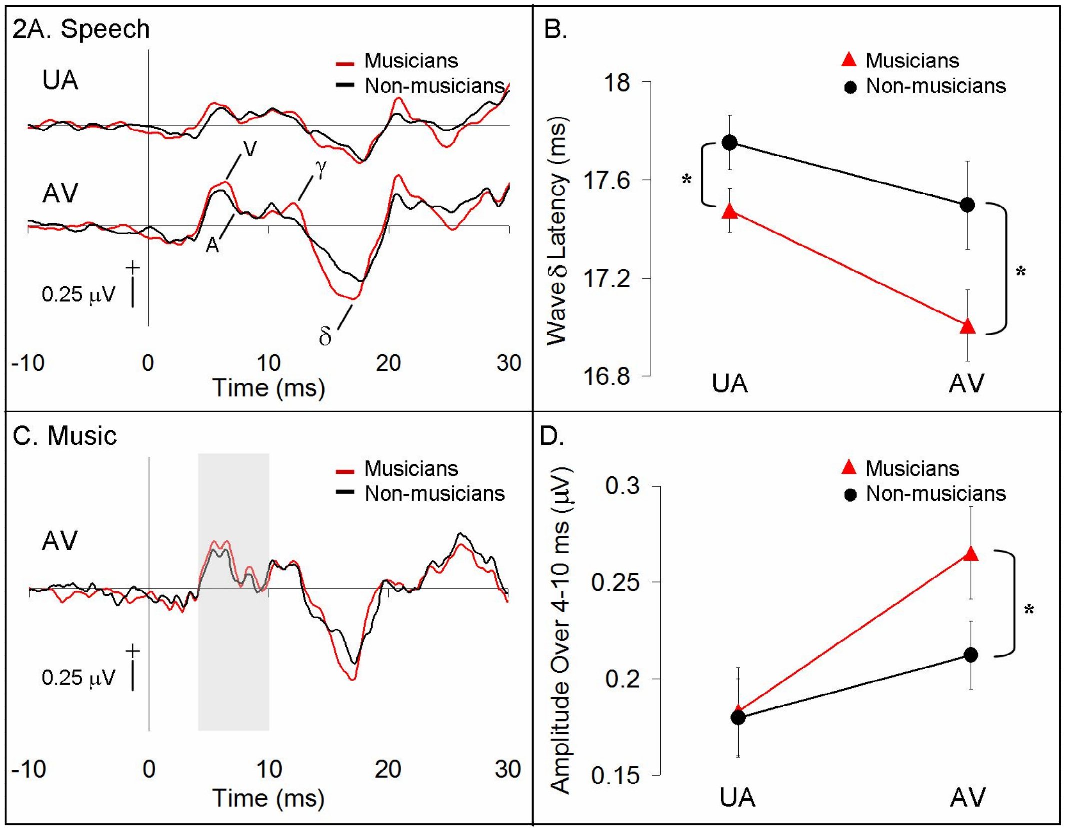

Figure 6 (In manuscript, Fig. 2). A. Grand average onset responses of the musicians and control subjects to the AV (top) and UA (bottom) speech stimuli. Unimodal visual speech and music stimuli elicited little activity, as indicated by the grey traces. Prominent peaks of the onset response (V, A, γ, 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 RMA values were calculated over 4-10 ms (shaded grey) to test whether musicians (red) had larger response magnitude early in the subcortical stream, prior to 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.

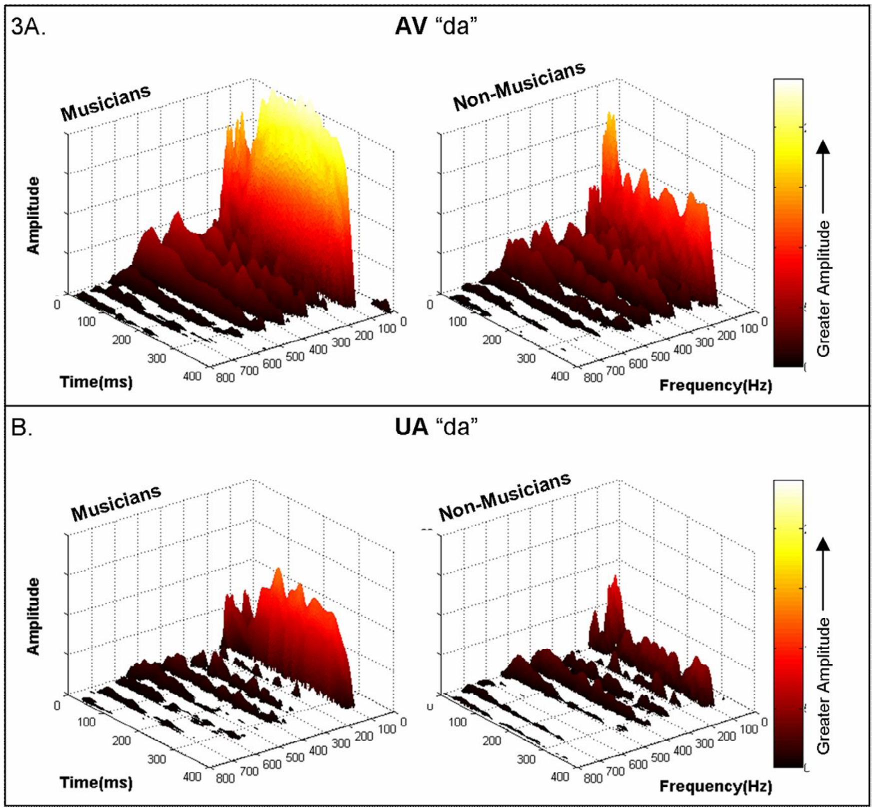

Figure 7 (In manuscript, Fig. 3). Narrowband spectrograms were calculated over the entire response to produce time-frequency plots (1 ms resolution) for musician and non-musician responses to audiovisual (top) and unimodal (bottom) speech. Lighter colors indicate greater amplitudes. Musicians have greater spectral energy over the duration of the response than controls, this difference being most pronounced at the 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.

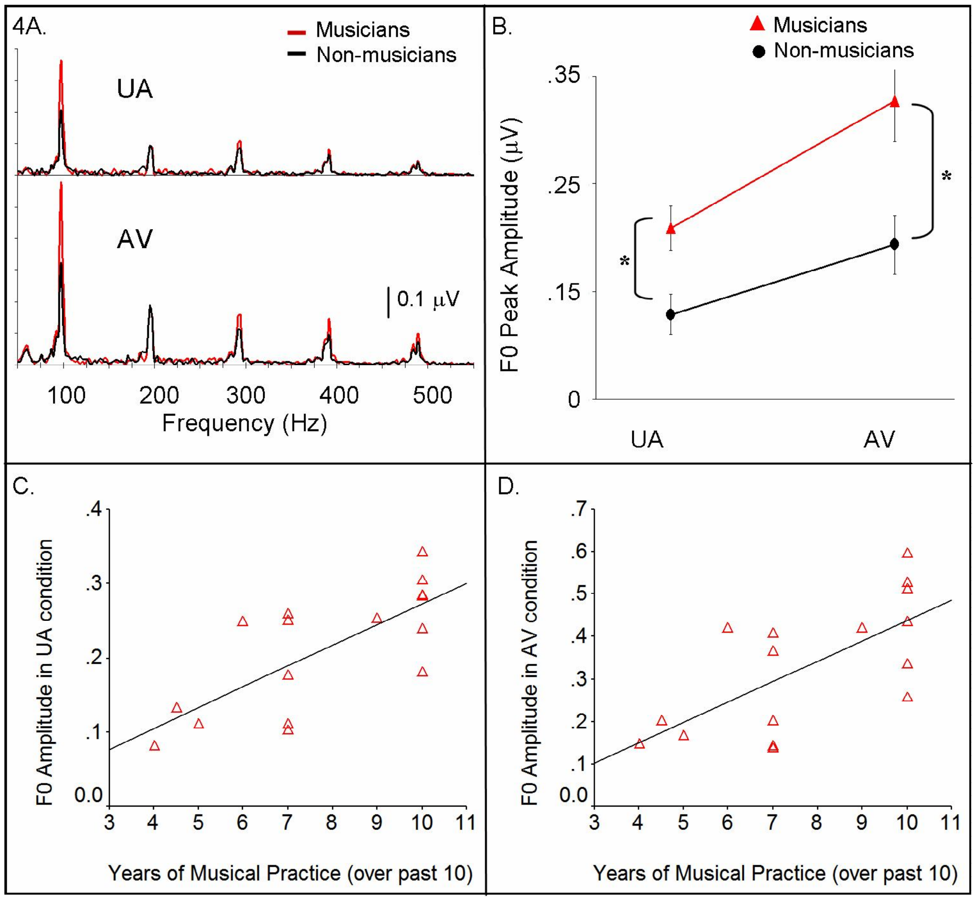

Figure 8 (In manuscript, Fig. 4). A. FFT analysis of the entire FFR period (30-350 ms) shows that musicians have more robust F0 peak amplitudes to both unimodal and audiovisual speech stimuli. B. The mean F0 peak amplitudes (±SEMs) were significantly larger in musicians than controls for both unimodal auditory and audiovisual stimuli. C & D. Years of consistent musical practice (>3days/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.

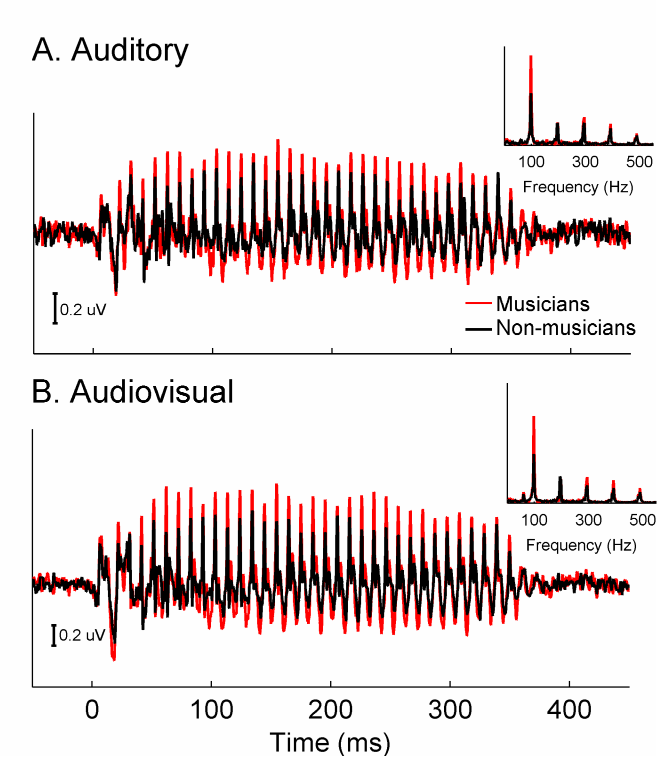

Figure 9 (In manuscript, Fig. 1). Grand average brainstem responses to speech. A. Musicians (red) have more robust responses than non-musicians (black) in the Audiovisual (Panel A) condition. Initial peaks of deflection (0-30 ms) represent the brainstem response to sound onset. Wave Delta of the response to sound onset are noted. The subsequent periodic portion reflects phase-locking to stimulus periodicity (frequency following response). Seeing a speaker say “da” elicited little brainstem activity, as illustrated by the musician’s Visual Alone grand average (grey). Non-musicians showed the same type of visual response, but for clarity, are not shown. B. The same musician-related effect is observed in the Auditory condition. Frequency spectra of the group averages, as assessed by Fast Fourier Transforms, are inset in each panel.

Figure 2.

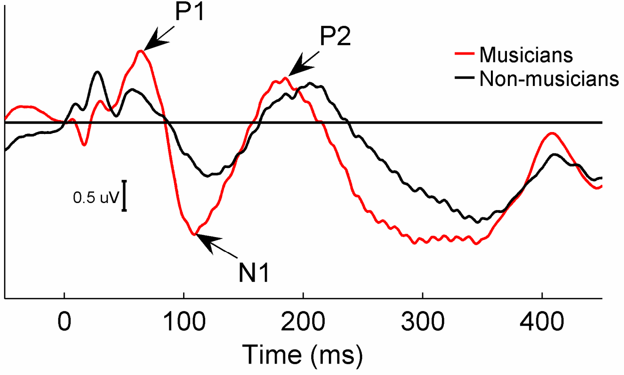

Figure 10 (In manuscript, Fig. 2). Musician and non-musician grand average cortical responses to speech in the AV condition. The speech syllable “da” in both A and AV conditions elicited four sequential peaks of alternating positive and negative deflections labeled P1, N1, P2, and N2, respectively. The slope between P1 and N1 was calculated to assess the synchrony of positive to negative deflection in the early portion of the cortical response. Peaks of cortical activity were earlier and larger in musicians (grey) than in non-musicians (black). In addition, P1-N1 slope was steeper in musicians compared to non-musicians. Similar effects were seen in the A condition.

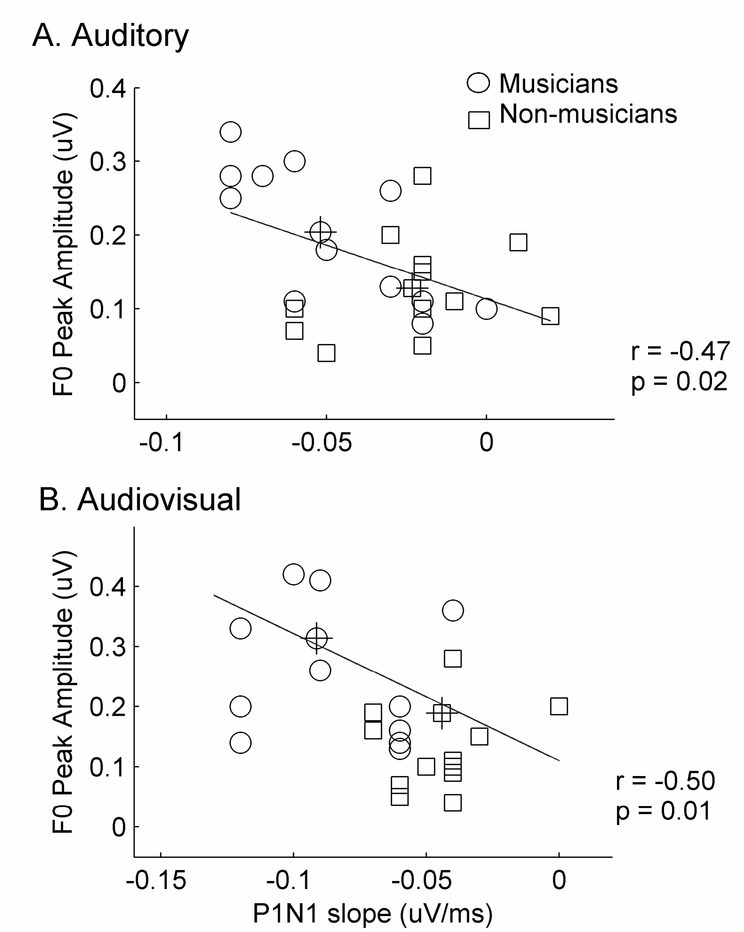

Figure 11 (In manuscript, Fig. 3). Relationship between P1-N1 slope and FFR encoding of pitch cues. A. Peak amplitude of the fundamental frequency (F0) correlated negatively with P1-N1 slope, indicating an association of larger F0 amplitude with steeper P1-N1 slope. Overall, musicians (circles) had larger F0 amplitudes and steeper slopes than non-musicians (squares). B. This relationship was also observed in the Audiovisual condition. Group means (crossed symbols) show that musicians have larger F0 amplitudes and steeper P1-N1 slopes than non-musicians in both stimulus conditions. Figure 4. Relationships between neurophysiological measures and musical training in musicians. A. More years of consistent musical practice were associated with steeper P1-N1 slope values in the Auditory condition (r = -0.68, p = 0.007). B. Years of consistent musical practice also correlated with brainstem measures of F0 amplitude in the Auditory and Audiovisual conditions (r_A =0.78, p = 0.001; r_{AV} = 0.72, p = 0.003). Only data from the Auditory condition are depicted in panel B.

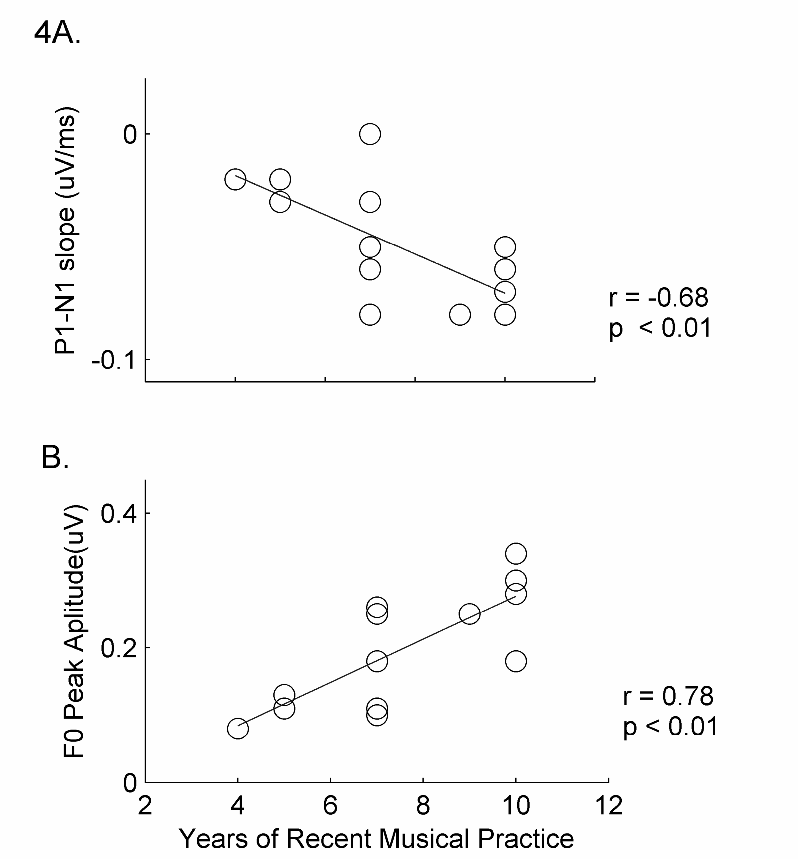

Figure 12 (In manuscript,Fig. 4). Relationships between neurophysiological measures and musical training in musicians. A. More years of consistent musical practice were associated with steeper P1-N1 slope values in the Auditory condition ($r = -0.68$, $p = 0.007$). B. Years of consistent musical practice also correlated with brainstem measures of F0 amplitude in the Auditory and Audiovisual conditions ($r_A = 0.78$, $p = 0.001$; $r_{AV} = 0.72$, $p = 0.003$). Only data from the Auditory condition are depicted in panel B.

Appendix 1

Manuscript - Brainstem responses to speech syllables

Clin Neurophysiol. 2004 Sep;115(9):2021-30.

Nicole Russoa,b,c,*, Trent Nicola,c, Gabriella Musacchiaa,c, Nina Krausa,b,c,d,e

a. Auditory Neuroscience Laboratory, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA b. Institute for Neuroscience, Northwestern University, Evanston, IL 60208, USA c. Department of Communication Sciences, Northwestern University, Evanston, IL 60208, USA d. Department of Neurophysiology and Physiology, Northwestern University, Evanston, IL 60208, USA e. Department of Otolaryngology, Northwestern University, Evanston, IL 60208, USA

Abstract

Objective: To establish reliable procedures and normative values to quantify brainstem encoding of speech sounds.

Methods: Auditory brainstem responses to speech syllables presented in quiet and in background noise were obtained from 38 normal children. Brainstem responses consist of transient and sustained, periodic components—much like the speech signal itself. Transient peak responses were analyzed with measures of latency, amplitude, area, and slope. Magnitude of sustained, periodic frequency-following responses was assessed with root mean square, fundamental frequency, and first formant amplitudes; timing was assessed by stimulus-to-response and quiet-to-noise inter-response correlations.

Results: Measures of transient and sustained components of the brainstem response to speech syllables were reliably obtained with high test–retest stability and low variability across subjects. All components of the brainstem response were robust in quiet. Background noise disrupted the transient responses whereas the sustained response was more resistant to the deleterious effects of noise.

Conclusions: The speech-evoked brainstem response faithfully reflects many acoustic properties of the speech signal. Procedures to quantitatively describe it have been developed.

Significance: Accurate and precise manifestation of stimulus timing at the auditory brainstem is a hallmark of the normal perceptual system. The brainstem response to speech sounds provides a mechanism for understanding the neural bases of normal and deficient attention-independent auditory function.

Keywords: Speech syllable response; Brainstem response; Auditory brainstem response; Frequency-following response; Effects of noise

1. Introduction

The neural encoding of sound begins in the auditory nerve and travels to the auditory brainstem. Brainstem responses to simple stimuli (e.g., clicks, tones) are widely used in clinical practice in the evaluation of auditory pathway integrity (Møller, 1999; Starr and Don, 1988). Less well-defined is how the brainstem responds to complex stimuli. Describing auditory encoding of speech sounds provides insight into some of the central auditory processes involved in normal communication. Furthermore, this knowledge may be applied to understanding effects of the aging process on hearing, as well as to a broad range of other circumstances, including hearing and communication in individuals with learning problems, peripheral hearing impairments, cochlear implants, or auditory neuropathies.

1.1. Background and significance

Some people have normal peripheral hearing, but still cannot perceive speech well. Previous studies have shown that the disruption of neural timing at the cortex is linked to auditory perceptual deficits (Kraus et al., 1996; Nagarajan et al., 1999; Tonnquist-Uhlen, 1996; Wible et al., 2002). In addition, abnormal electrophysiological responses to speech syllables at the brainstem level have been associated with a wide spectrum of diagnosed learning problems (King et al., 2002; Wible et al., in press). These abnormalities include a temporally delayed response to the onset of a consonant and deficient spectral representation of harmonic aspects of the speech signal. Disruptions of neural encoding in both the brainstem and cortex were exacerbated when speech was presented in background noise (Cunningham et al., 2001).

Part of the difficulty in perceiving consonants in noisy situations is that they are rapid, relatively low-amplitude transient features of speech. Stop consonants, such as /d/, are known to be particularly vulnerable to disruption by background noise in normal and clinical populations (Brandt and Rosen, 1980). The perception of vowels, however, is more resistant to the effects of noise because they are periodic, sustained signals, and generally louder than consonants.

Brainstem responses provide direct information about how the sound structure of a speech syllable is encoded by the auditory system. It is particularly compelling to consider that specific aspects of the sound structure of the acoustic signal are maintained and reflected in the neural code. Similar to the speech syllable itself, the brainstem response to a speech syllable can be divided into transient and sustained portions, namely the onset response and the frequency-following response (FFR) (Boston and Møller, 1985). Onset responses are transient, with peak durations lasting tenths of milliseconds, thus we will refer to these rapid deflections as transient responses. Within the FFR are discrete peaks corresponding to the periodic peaks in the stimulus waveform. However, this region can be considered as a whole, as it contains a periodic signal sustained for tens or hundreds of milliseconds. Although peaks within the FFR may be thought of as successive onsets, for descriptive purposes, we will use the term FFR to refer to the later portion of the response evoked by the harmonic vowel structure of the stimulus. There is a parallel effect of noise on the brainstem response, similar to the disruption of speech perception, in that transient onsets were more affected by the noise, sometimes even eliminated, while the sustained portion remained intact (Cunningham et al., 2001).

1.2. Specific aims

The specific aims of this study were: (1) to delineate measures of the timing and magnitude of the brainstem response to the speech syllable /da/ in quiet and background noise; (2) to establish normative values for these features; and (3) to determine the test–retest reliability of these measures.

2. Methods

2.1. Subjects

Thirty-eight children, ages 8–12 years (21 male, 17 female) participated in the primary focus of this study, which established normative values for the brainstem response to speech syllables. Eight children (four male, four female) were part of the retest reliability portion of the study. None of the children had a history of medical or learning problems and all performed within normal limits on laboratory-internal standardized measures of learning and academic achievement. These measures consisted of selected subtests of Woodcock Johnson, Woodcock Johnson—Revised, and Wide Range Achievement Tests that have been described in detail elsewhere (Kraus et al., 1996). All of the subjects had normal click-evoked auditory brainstem response latencies and normal hearing thresholds at or below 20 dB HL for octaves from 500 to 4000 Hz. Consent and assent were obtained from the parents (or legal guardians) and the children involved in the study. The Institutional Review Board of Northwestern University approved all research.

2.2. Stimulus and recording parameters

Because stop consonants provide considerable phonetic information and their perception is particularly vulnerable to background noise in both normal and clinical populations, a five-formant synthesized /da/ was chosen for the stimulus (Klatt, 1980). The stimulus duration was 40 milliseconds (ms). The consonant contained an initial 10 ms burst; the frequencies of which were centered around the beginning frequencies of formants 3–5, thus in the range of 2580–4500 Hz. Additional details of the speech synthesis parameters can be found in King et al. (2002). The syllable /da/ was presented monaurally, in alternating polarities, at 80 dB SPL to the right ear via insert earphones (ER-3, Etymotic Research, Elk Grove Village, IL), with an inter-stimulus interval of 51 ms.

During testing, children watched a videotape with the sound level set at , 40 dB SPL in the non-test ear. Responses were collected with Ag–AgCl scalp electrodes, differentially recorded from Cz (active) to ipsilateral earlobe (reference), with the forehead as ground. Three blocks of 1000 sweeps per block were collected at each polarity and in each of two different signal-to-noise conditions, quiet and ipsilateral white Gaussian noise (þ5 dB SNR).

Waveforms were averaged online in Neuroscan (Compumedics, El Paso, TX) with a

recording time window spanning 10 ms prior to the onset and 20 ms after the offset of the stimulus. The sampling rate was 20,000 Hz and responses were online bandpass filtered from 100–2000 Hz, 6 dB/octave. Trials with eye-blinks greater than 35 mV were rejected online. Responses of alternating polarity were then added together to isolate the neural response by minimizing stimulus artifact and cochlear microphonic (Gorga et al., 1985).

2.3. Description of the brainstem response to speech syllables

The electrophysiological brainstem response to a speech sound is a complex waveform (Fig. 1). This waveform includes transient peaks as well as sustained elements that comprise the FFR. The response to the onset of the speech stimulus /da/ includes a positive peak (wave V), likely analogous to the wave V elicited by click stimuli, followed immediately by a negative trough (wave A). In most subjects, positive peaks corresponding to click-evoked auditory brainstem response waves I and III are also visible. Following the onset response, peaks C and F are present in the FFR. While other peaks are discernable in this region, a previous study (King et al., 2002) determined that peaks C and F were the most stable for this stimulus across individuals, having latencies with standard deviations (SDs) less than 0.5 ms across a normal population. The defining feature of the sustained portion of the response is its periodicity, which follows the frequency information contained in the stimulus (Marsh et al., 1974; Smith et al., 1975). The timing and magnitude of both the transient and sustained aspects of the response waveform were evaluated with the measures described below and summarized in Table 1.

2.4. Discrete peak measures

Measures of both timing and magnitude were utilized to assess the discrete peaks. The onset response waves V and A occur at latencies before 10 ms, while peaks C and F occur at latencies of about 18 and 40 ms, respectively. Three experienced raters picked peaks V, A, C, and F and their latencies and amplitudes were measured. The VA complex was further investigated by measuring its inter-peak interval, amplitude, slope, and area. These measures were taken from the raw waveforms of the responses collected in quiet. When background noise was introduced with the syllable, peaks were often obscured in the raw waveform. Therefore, a wavelet-denoising technique adapted from Quian Quiroga and Garcia (2003) was applied to the waveforms collected in noise before obtaining transient response measures. Nevertheless, some peaks were still imperceptible after the denoising procedure. These peaks were designated as having 0 mV for amplitude and area and were omitted from latency, slope, and inter-peak interval analyses. Additionally, some peaks were not eliminated completely, but if there was not a consensus among peak pickers regarding the actual presence and location of the peak, it was omitted from statistical analyses. These omissions are reflected in the subject numbers listed in Tables 2–4.

2.5. Frequency-following response measures

The region following the onset response was defined as the FFR. The FFR analysis window was chosen to incorporate as much of the sustained response as possible, while avoiding the refractory period of the onset response and any offset transients. Five analysis techniques were employed to analyze the FFR: (1) root mean square amplitude (RMS amp); (2) amplitude of the spectral component corresponding to the stimulus fundamental frequency (F0 amp); (3) amplitude of the spectral component corresponding to first formant frequencies of the stimulus (F1 amp); (4) stimulus-to-response (S–R) correlations, and (5) inter-response

(I–R) correlations between the responses collected in the quiet and noise conditions. These composite FFR measures describe the sustained portion of the response as whole.

2.5.1. Root mean square amplitude

This measure reflects the averaged magnitude of activation of the neural population over an 11.5–46.5 ms epoch of the sustained response. Responses were de-meaned and, to correct for varying amounts of internal (e.g., myogenic) noise among subjects, the RMS amplitude of the response was divided by the RMS amplitude of the pre-stimulus period.

2.5.2. Amplitudes of the fundamental frequency and first formant

The FFR consists of energy at the fundamental frequency of the stimulus and its harmonics (Worden and Marsh, 1968). Fourier analysis was performed on the 11.5–46.5 ms epoch of the FFR in order to assess the amount of activity occurring over two frequency ranges. Activity occurring in the frequency range of the response corresponding to the fundamental frequency of the speech stimulus (103 –121 Hz) was calculated for each subject. The response activity corresponding to the first formant frequencies of the stimulus (220–720 Hz) was also measured. The F0 amp provides a gauge of the specific portion of the sustained response devoted to encoding the fundamental frequency of the speech sound, while the F1 amp is devoted to encoding the first formant (Fig. 2). A 2 ms on–2 ms off Hanning ramp was applied to the waveform. Zero-padding was employed to increase the number of frequency points where spectral estimates were obtained. A subject’s response was required to be above the noise floor in order to be included in the analyses. This calculation was performed by comparing the spectral magnitude of the pre-stimulus period to that of the response[1]. If the

---

[1] The FFR period is 3.5 times longer than the pre-stimulus period, so in order to make an accurate comparison, the spectral magnitude of the 10 ms pre-stimulus period was compared

quotient of the magnitude of the F0 or F1 frequency component of the FFR divided by that of the pre-stimulus period was greater than or equal to one, the response was deemed above the noise floor. The raw amplitude value of the F0 or F1 frequency component of the response FFR was then measured. Only the F0 and F1 frequencies of the response were above the noise floor. Response frequencies corresponding to higher stimulus formants were not significantly above the noise floor in either quiet or background noise.

2.5.3. Stimulus-to-response correlation

The stimulus-to-response (S–R) correlation reflects how faithfully the response waveform mimics the stimulus waveform, and provides a measure of phase locking that excludes the non-periodic activity inherent in the RMS amp measure. Each response was cross-correlated to the 10–40 ms portion of the stimulus that includes the harmonic segment of the syllable. Due to the time it takes for neural impulses to propagate through the nervous system, the response lagged behind the stimulus by approximately 7–10 ms. Thus, maximal correlation within this range was recorded.

2.5.4. Quiet-to-noise inter-response correlation

The inter-response (I–R) correlation reflects the fidelity of the response morphology recorded in noise to that of the response recorded in quiet, providing a way to quantify the effects of background noise on the timing of the sustained response. The I–R correlation was calculated similarly to the S–R correlation measure. However, because the addition of noise can delay

---

against the average of three 10 ms ranges (12.5–22.5, 22.5–32.5, and 32.5–42.5 ms) within the FFR. This ensured that the same number of points and therefore the same frequency ranges were compared. The pre-stimulus period and each of the 10 ms ranges were de-meaned to zero before performing spectral analyses.

the brainstem response by a couple of milliseconds, a cross-correlation was performed and the quiet response was allowed to lead the noise response by up to 2 ms.

Mean Pearson's r-values were reported for normative descriptive purposes of S–R and I–R correlations, although Fisher's transformation was used to convert r-values to z0-scores for all statistical computations.

3. Results

Based on our evaluation of 38 subjects' responses recorded in quiet and 36 subjects' responses recorded in background noise, normative values for the aforementioned brainstem measures were established. Table 2 shows means and SDs for discrete peak measures obtained in quiet and background noise. Tables 3 and 4 provide timing and magnitude values, respectively, for the FFR.

3.1. The normal response in quiet

Responses were highly replicable both within and across subjects. Fig. 3 shows three 1000-sweep blocks obtained from a representative subject (top), as well as responses obtained from another subject on two separate test sessions (bottom). Peaks V, A, and C were detectable in all subjects, and peak F was detectable in all but two subjects. The onset response waves V and A were largest in magnitude, followed by FFR peak F and then peak C. Consistent with other neurophysiological responses, as latency increases, so does its variability. The SD of latency was smallest for the early onset response waves V and A (0.25 and 0.34 ms, respectively), and increased with latency (up to 0.61 ms).

The FFR was evident in all subjects. Timing of the FFR was indicated by S–R and I–R correlations. The magnitude of the response was evaluated with RMS, F0 and F1 amplitudes. As shown in Fig. 2, the greatest amount of energy is present in the F0 region.

3.1.1. Relationships among brainstem response measures

Relationships among brainstem response measures recorded in quiet were explored using Pearson's correlations. Correlations among measures exceeding ^0.30 and P , 0:05 criterion were considered significant. Tables 5–7 show the relationships among and between transient and sustained measures. Transient measures, especially those describing the VA complex and wave C, were largely related to each other (Table 5), while the timing of peak F was relatively independent of the timing of other peaks. Composite sustained measures did not exhibit strong relationships with each other, indicating that each measure described a unique quality of the FFR (Table 6). Transient onset responses were largely independent of sustained measures, with the exception that a number of wave V and A measures were related to F1 amplitude. Discrete peaks within the FFR were related to almost all of the composite FFR measures (Table 7).

3.2. The normal response in noise

The addition of background noise interfered with normal brainstem encoding of the speech stimulus /da/. Fig. 4 shows the effects of noise. Table 2B shows normative values for the transient response measures in noise. Most affected were the onset responses V and A, which were severely degraded and completely obscured in more than 40% of the subjects. Peaks C and F, however, remained present in noise in most subjects (100 and 86%, respectively). Peak amplitudes also were affected by the presence of noise; all peaks were reduced in size (P , 0:001; all tests).

When not eliminated, latencies of onset peaks V and A, and FFR peak C were delayed in comparison to quiet (P , 0:01; all tests). In contrast, peak F showed resilience to background noise in that its latency did not change with the addition of the noise P . 0:12and remained easily identifiable in most subjects. As expected, the introduction of background noise increased the variability in the latencies of all peaks.

Although reduced, the composite FFR remained relatively intact and was discernible in noise. RMS amp and S–R correlations showed significant reductions in noise (P , 0:002; both tests). F0 and F1 amp were also significantly affected by the presence of background noise (P , 0:002; both tests). The addition of noise obscured onset peaks in the responses of many subjects, thus it was not possible to calculate the relationships between onset and FFR measures in noise.

3.3. Test–retest stability

In order to determine whether the variables described here are stable over time, eight of the children were retested after a 2–10-month interval. Test–retest reliability is illustrated in the waveforms shown in a representative subject in Fig. 3 (bottom) and at the group level in Fig. 4. Two-tailed, paired t tests were calculated for all brainstem response measures. A significance criterion of P , 0:05 was used. Most brainstem measures did not change significantly over the test–retest time interval P . 0:09; exceptions included the amplitude and slope of the VA complex in quiet and wave C latency in noise P , 0:02; all). Sustained measures were stable from test to retest (P . 0:30; all tests).

4. Discussion

The ability to quantify a brainstem response elicited by speech sounds provides a powerful tool for research and clinical use. The speech-evoked brainstem response faithfully reflects many acoustic properties of the speech signal. In the normally perceiving auditory system, stimulus timing, on the order of fractions of milliseconds, is accurately and precisely represented at the level of the brainstem. Overall, the brainstem response provides a mechanism for understanding the neural bases of normal and deficient auditory function, by providing a quantifiable measure of an individual's attention-independent neural encoding of speech sounds.

This study described explicit methods to record and quantify the brainstem response to /da/ in quiet and in background noise and provided a normative data set which can be used to assess the integrity of speech signal encoding in normal and clinical populations. Measures of timing and magnitude were identified for both transient and sustained aspects of the responses. Transient response measures included latency and amplitudes of peaks V, A, C, and F, as well as inter-peak interval, slope, area and amplitude of the VA complex as a unit. Sustained measures included RMS amplitude, F0 and F1 amplitudes, S–R correlations, and I–R correlations. In quiet, these brainstem encoding parameters can be obtained nearly 100% of the time; variability is low and test–retest stability is high. The addition of background noise often eliminated the onset response (waves V and A) or resulted in non-uniform latency delays. Because robust responses are necessary for accurate encoding, this disruption could underlie perceptual difficulties. Although the latencies of waves V, A, and C were delayed in noise, peak F remained stable. Thus, while it appeared that background noise induced a delay in responding to the onset of a sound, compensatory mechanisms may correct for this lag throughout the neural pathway. F0 remained robust in background noise, while other sustained measures, despite often being reduced in magnitude, also showed more resilience to the effects of noise.

Overall, test–retest stability was high for responses obtained in both quiet and background noise. Although minimal variability may exist due to placement of electrodes or the insert earphone, the test–retest measures described in this study showed considerable stability over time

4.1. Interpreting the brainstem response: transient versus sustained

In as much as it may be an oversimplification to equate features of speech, such as consonants and vowels, with transient and sustained evoked responses, there are certain parallels. The transient portions of the brainstem response reflect the encoding of rapid temporal changes inherent in consonants. The sustained FFR encodes the harmonic and periodic sound structure of vowels. In quiet conditions, both the transient and sustained components of the speech syllable /da/ are robustly encoded. In noise, just as vowels are less affected than consonants, the FFR is less degraded than the onset response.

A major difference between the onset and FFRs measured here was that under a stressed circumstance— background noise—neural encoding of onset features was severely degraded, whereas the sustained FFR features remained relatively unaltered. Onset waves V and A were eliminated in almost half of the subjects, while peaks C and F, and the FFR region as a whole, remained stable. Consequently, the perceptual problems associated with consonant identification in background noise could be attributed to the decreased neural synchrony reflected in the onset, while the intact encoding reflected in the sustained region enabled accurate vowel perception.

F0 amplitude remained robust in noise. Encoding of the fundamental frequency is important for identifying the speaker and emotional tone of voice. Meanwhile, the degradation of F1; which provides phonetic information, coupled with the loss of the onset burst, further degrades perception of the speech signal in noise. These data provide evidence to support the observation that speaker identity and speaker tone of voice is more resistant to noise than the phonetic content of what is being said.

However, another possible explanation is that the /da/ stimulus is smaller in amplitude at its onset than towards the end. Thus, the elimination of waves V and A, and the maintenance of the FFR, may be due to the relative differences in which aspects of the stimulus did or did not exceed the level of the acoustic masking noise. Future studies incorporating different types of background noise, such as pink noise or multi-speaker babble maskers (which more closely resemble naturally occurring noise and

N. Russo et al. / Clinical Neurophysiology 115 (2004) 2021–2030

the spectrum of speech itself) likely will contribute to further understanding the encoding of speech in background noise.

The overall resistance of the FFR versus the disruption of the onset response in noise suggests a relative independence of brainstem encoding processes. Furthermore, the independence of the transient versus sustained responses was apparent in the relationships among measures. That is, correlations were strong within transient and sustained measures separately, whereas fewer, weaker relationships existed between these classes of measures. Although transient measures within the FFR (e.g., waves C and F) showed relationships to the composite sustained response measures, transient onset and composite FFR measures demonstrated few relationships, reinforcing that they are neither wholly separate nor wholly related measures. It is interesting to note that brainstem responses that reflect prosodic aspects of speech (F0 and RMS amplitude) are largely independent from the internally related measures (waves V, A, and C latency and F1 amplitude), which represent phonetic information of the stimulus.

4.2. Practical applications

4.2.1. Individual versus group data

Most physiological and imaging approaches for assessing the functional integrity of sensory systems require group data and can be time-intensive. Collecting the brainstem response to a speech sound can be done in a few minutes, requires few electrodes, and is passively acquired. Normal variability of response attributes is low. Furthermore, the brainstem response is stable over time. Consequently, the measures reported here lend themselves to the assessment of the encoding of sound structure in individual subjects.

4.2.2. Identification of auditory-based learning disabilities

The data provided here serve as a metric for determining normal brainstem function in response to speech sounds. Deficits in neural timing and magnitude in response to speech syllables at the brainstem level have been previously found to occur in certain children with auditory-based learning problems (Cunningham et al., 2001; King et al., 2002). Timing abnormalities in waves V, A, and C have been identified (King et al., 2002). Decreases in S–R correlations and the reduced magnitude of the FFR, specifically in the frequency composition of F1; have also been found in children with learning problems (Cunningham et al., 2001).

4.2.3. Predictors of future language impairment

A recent publication by Benasich and Tallal (2002) reported that behavioral measures of central auditory function, obtained in children under 1 year of age (mean age ¼ 7.5 months), can serve as predictors for subsequent specific language impairments (SLI) and other developmental language delays. Due to the early maturation of the brainstem response, the brainstem measures described in this paper, might provide a biological marker for early detection of central auditory deficits that may dovetail with these behavioral findings. Further research is needed to determine which specific manifestations of brainstem abnormalities may facilitate the early prediction of SLI. The normative data provided here can serve as an objective index for early diagnosis and identification of deficits in the neural encoding of sound structure in the brainstem. Intervention could be applied before the behavioral aspects of their impairment impact a child.

4.2.4. Predictors for success with auditory training

Neural encoding of sound structure in the auditory brainstem may provide a predictive index for success with auditory training regimens. Children with learning problems and brainstem-encoding deficits have been shown to benefit from auditory training (Hayes et al., 2003). Specifically, trained children with a delayed brainstem onset latency (wave A) in quiet showed greater improvements in the timing and magnitude of cortical responses. Additionally, behavioral improvements were seen in tests of phoneme decoding (Sound Blending and Auditory Processing) in these children. Thus, children with brainstem encoding deficits particularly appeared to benefit from auditory training. These data support the idea that early analysis of the brainstem response could predict which children would benefit from auditory training.

5. Conclusions

Brainstem response timing and magnitude measures provide reliable information about the neural encoding of speech sounds. This study outlined specific measures of brainstem function that may be used to characterize neural encoding of speech sounds for clinical and research applications. Transient and sustained measures provide information regarding auditory pathway encoding of brief and periodic aspects of the stimulus. Some of the data suggest that transient and sustained responses represent independent mechanisms. A better understanding of brain-stem encoding may assist in early diagnosis and intervention of auditory disorders, as well as in measuring the success of training programs.

The current study is a springboard for further examination of brainstem activity to complex speech stimuli, as well as for identifying abnormalities in clinical populations such as aging, peripheral hearing impairment, cochlear implant, auditory neuropathy, and non-native listener populations, in which the assessment of auditory function is relevant. Future parametric studies (e.g., of the effects of different speech stimuli, methods of presentation or types of background noise) may enhance the potential clinical use of the brainstem response to speech sounds.

Acknowledgements

The National Institute of Health NIDCD R01-01510 supported this research. We thank the children and their families who participated in this study. We also thank Steven Zecker, for statistical consultation; Erika Skoe, for the development of programs to analyze the data; and also members of the Kraus laboratory who tested subjects and gave their support throughout the study.

Tables and Figures

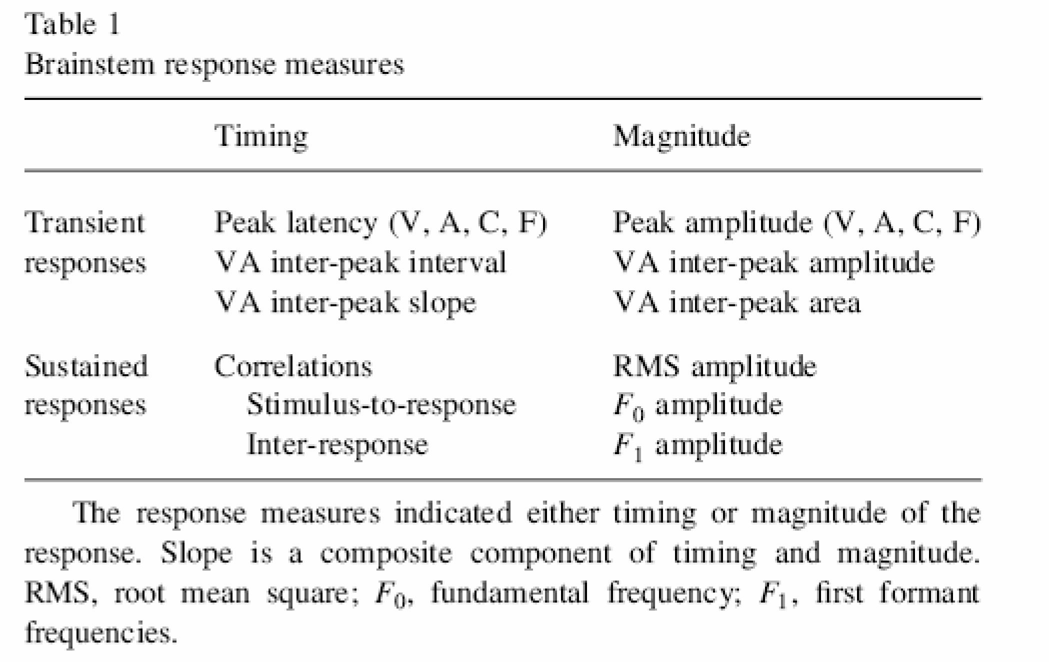

Table 1
Brainstem response measures

	Timing	Magnitude
Transient responses	Peak latency (V, A, C, F)	Peak amplitude (V, A, C, F)
	VA inter-peak interval	VA inter-peak amplitude
	VA inter-peak slope	VA inter-peak area
Sustained responses	Correlations	RMS amplitude
	Stimulus-to-response	F_0 amplitude
	Inter-response	F_1 amplitude

The response measures indicated either timing or magnitude of the response. Slope is a composite component of timing and magnitude. RMS, root mean square; F_0, fundamental frequency; F_1, first formant frequencies.

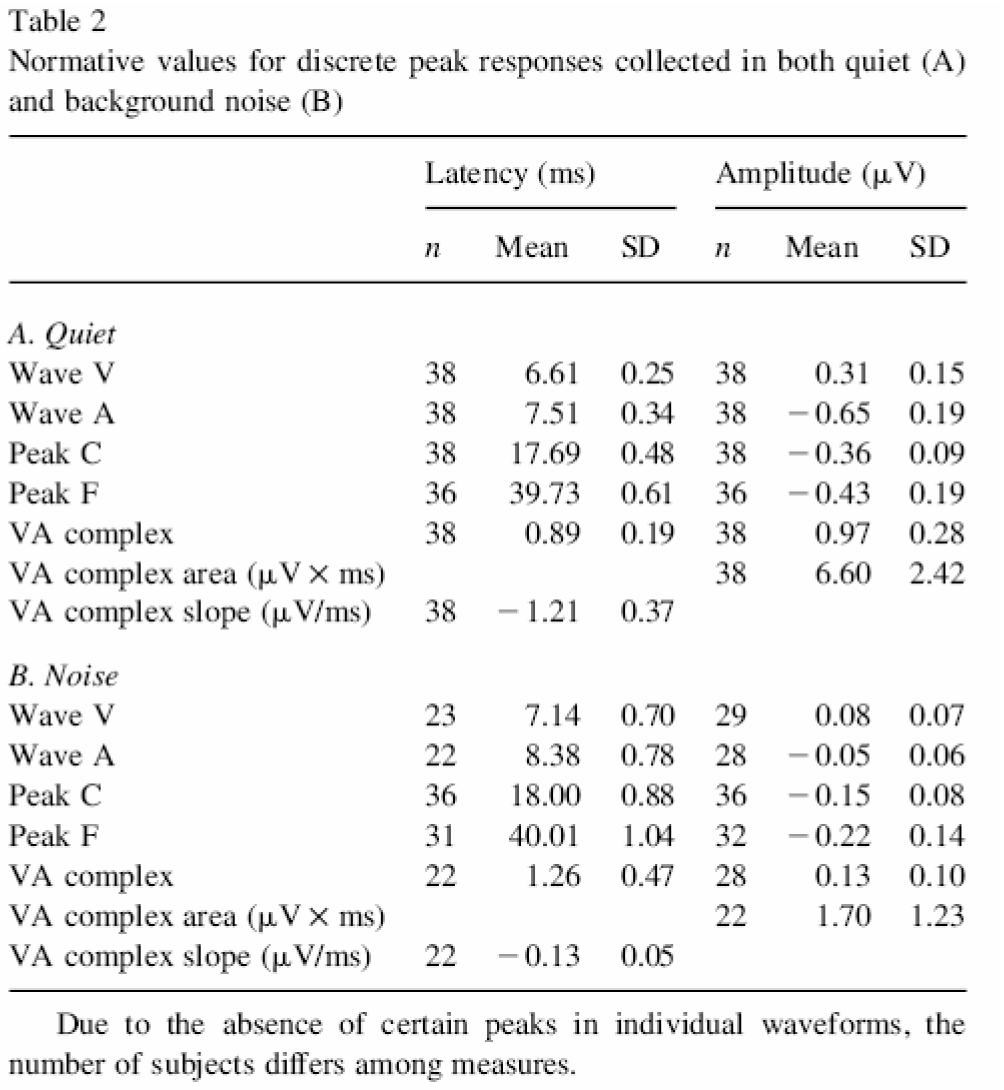

Table 2
Normative values for discrete peak responses collected in both quiet (A) and background noise (B)

	Latency (ms)			Amplitude (μV)		
	n	Mean	SD	*n*	Mean	SD
A. Quiet						
Wave V	38	6.61	0.25	38	0.31	0.15
Wave A	38	7.51	0.34	38	−0.65	0.19
Peak C	38	17.69	0.48	38	−0.36	0.09
Peak F	36	39.73	0.61	36	−0.43	0.19
VA complex	38	0.89	0.19	38	0.97	0.28
VA complex area (μV × ms)				38	6.60	2.42
VA complex slope (μV/ms)	38	−1.21	0.37			
B. Noise						
Wave V	23	7.14	0.70	29	0.08	0.07
Wave A	22	8.38	0.78	28	−0.05	0.06
Peak C	36	18.00	0.88	36	−0.15	0.08
Peak F	31	40.01	1.04	32	−0.22	0.14
VA complex	22	1.26	0.47	28	0.13	0.10
VA complex area (μV × ms)				22	1.70	1.23
VA complex slope (μV/ms)	22	−0.13	0.05			

Due to the absence of certain peaks in individual waveforms, the number of subjects differs among measures.

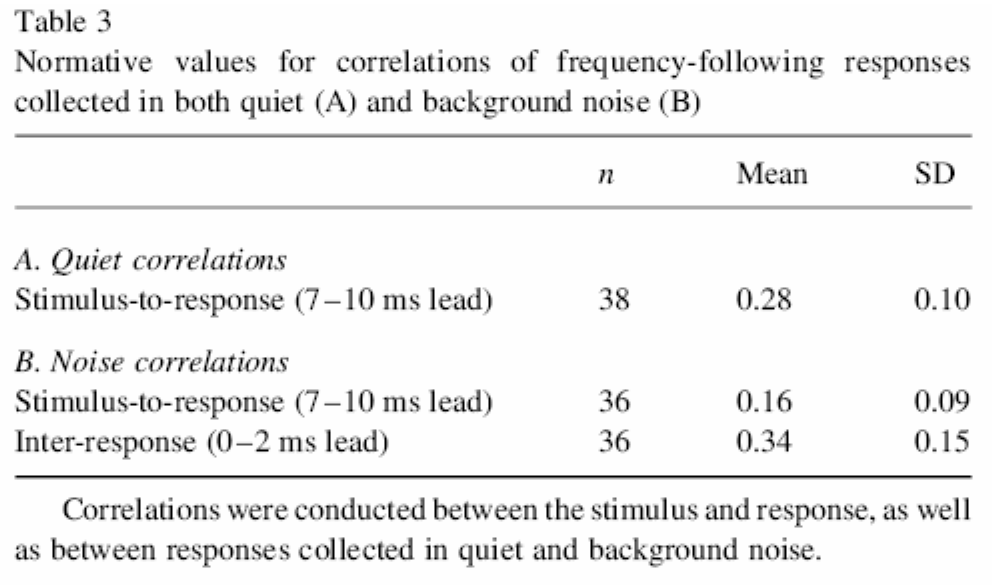

Table 3
Normative values for correlations of frequency-following responses collected in both quiet (A) and background noise (B)

	n	Mean	SD
A. Quiet correlations			
Stimulus-to-response (7–10 ms lead)	38	0.28	0.10
B. Noise correlations			
Stimulus-to-response (7–10 ms lead)	36	0.16	0.09
Inter-response (0–2 ms lead)	36	0.34	0.15

Correlations were conducted between the stimulus and response, as well as between responses collected in quiet and background noise.

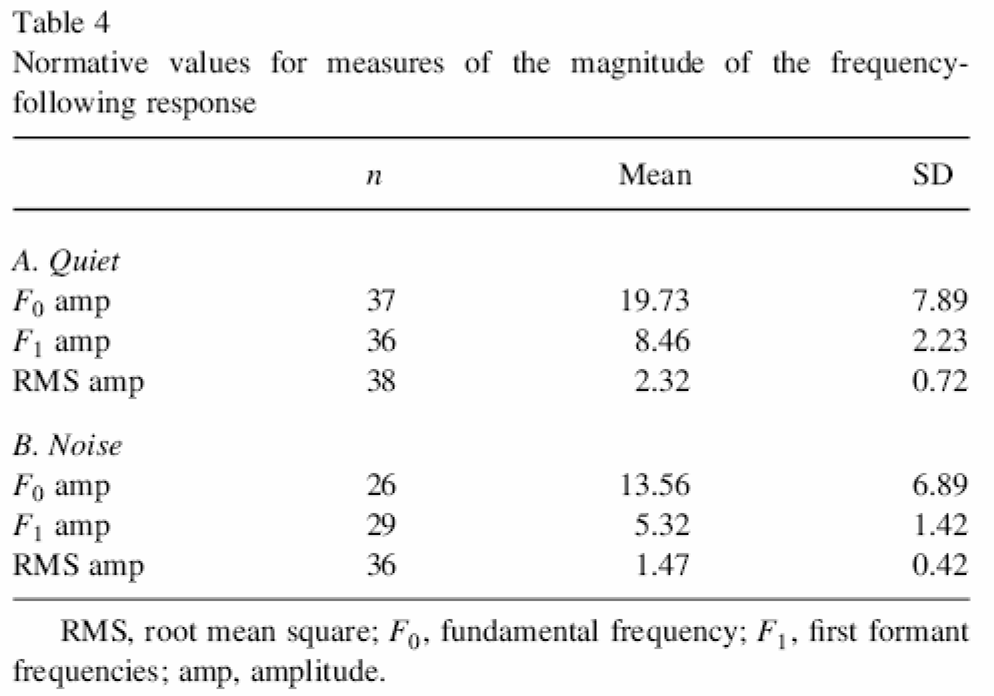

Table 4
Normative values for measures of the magnitude of the frequency-following response

	n	Mean	SD
A. Quiet			
F_0 amp	37	19.73	7.89
F_1 amp	36	8.46	2.23
RMS amp	38	2.32	0.72
B. Noise			
F_0 amp	26	13.56	6.89
F_1 amp	29	5.32	1.42
RMS amp	36	1.47	0.42

RMS, root mean square; F_0, fundamental frequency; F_1, first formant frequencies; amp, amplitude.

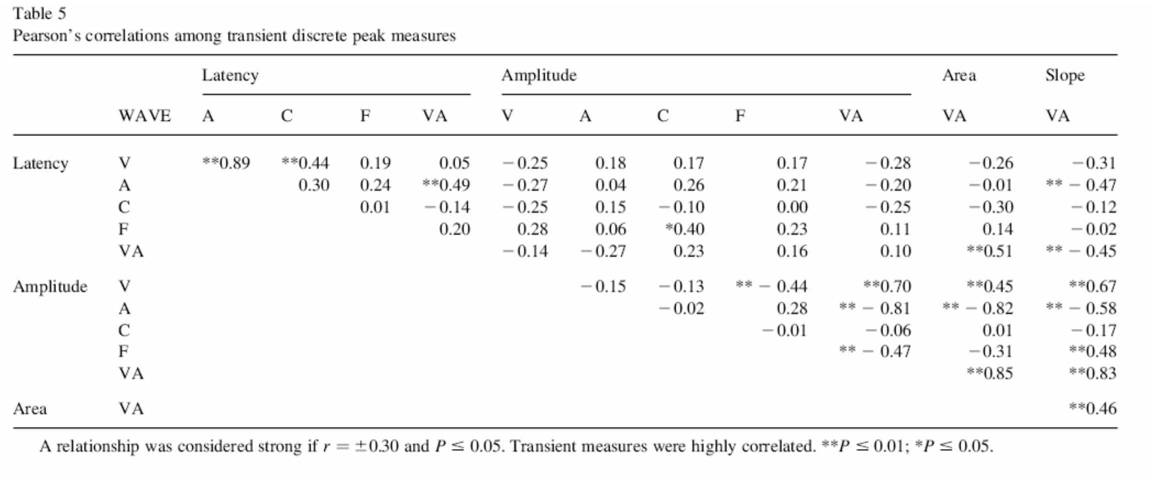

Table 5
Pearson's correlations among transient discrete peak measures

		Latency				Amplitude					Area	Slope
	WAVE	A	C	F	VA	V	A	C	F	VA	VA	VA
Latency	V	**0.89	**0.44	0.19	0.05	−0.25	0.18	0.17	0.17	−0.28	−0.26	−0.31
	A		0.30	0.24	**0.49	−0.27	0.04	0.26	0.21	−0.20	−0.01	** − 0.47
	C			0.01	−0.14	−0.25	0.15	−0.10	0.00	−0.25	−0.30	−0.12
	F				0.20	0.28	0.06	*0.40	0.23	0.11	0.14	−0.02
	VA					−0.14	−0.27	0.23	0.16	0.10	**0.51	** − 0.45
Amplitude	V						−0.15	−0.13	** − 0.44	**0.70	**0.45	**0.67
	A							−0.02	0.28	** − 0.81	** − 0.82	** − 0.58
	C								−0.01	−0.06	0.01	−0.17
	F									** − 0.47	−0.31	**0.48
	VA										**0.85	**0.83
Area	VA											**0.46

A relationship was considered strong if $r = \pm 0.30$ and $P \leq 0.05$. Transient measures were highly correlated. $^{**}P \leq 0.01$; $^{*}P \leq 0.05$.

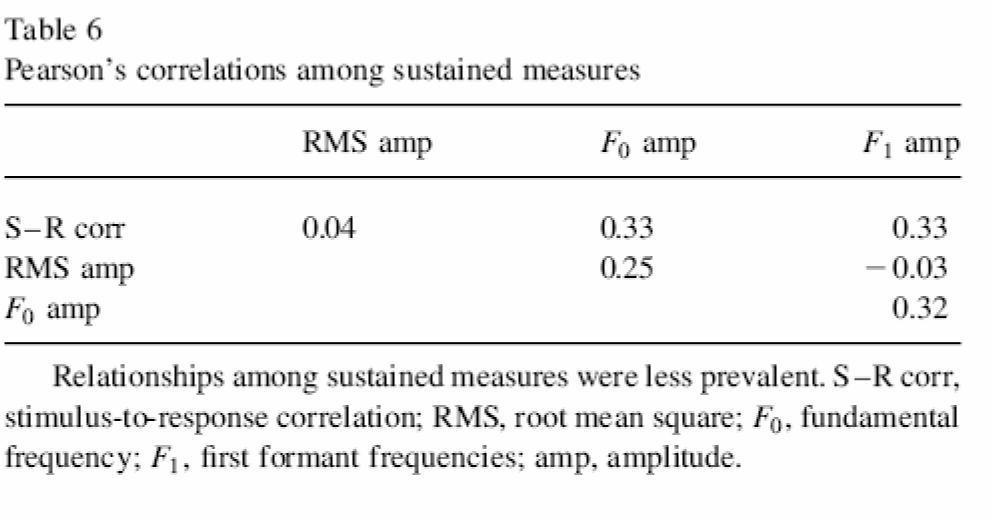

Table 6

Pearson's correlations among sustained measures

	RMS amp	F_0 amp	F_1 amp
S–R corr	0.04	0.33	0.33
RMS amp		0.25	−0.03
F_0 amp			0.32

Relationships among sustained measures were less prevalent. S–R corr, stimulus-to-response correlation; RMS, root mean square; F_0, fundamental frequency; F_1, first formant frequencies; amp, amplitude.

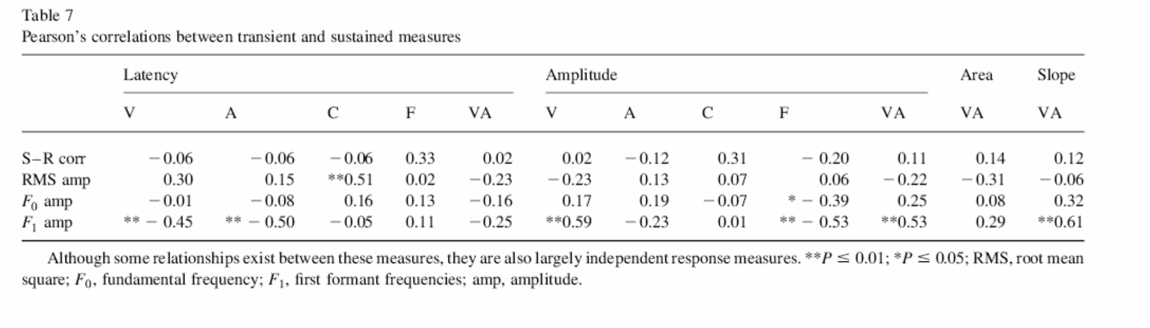

Table 7
Pearson's correlations between transient and sustained measures

	Latency					Amplitude					Area	Slope
	V	A	C	F	VA	V	A	C	F	VA	VA	VA
S–R corr	−0.06	−0.06	−0.06	0.33	0.02	0.02	−0.12	0.31	−0.20	0.11	0.14	0.12
RMS amp	0.30	0.15	**0.51	0.02	−0.23	−0.23	0.13	0.07	0.06	−0.22	−0.31	−0.06
F_0 amp	−0.01	−0.08	0.16	0.13	−0.16	0.17	0.19	−0.07	* −0.39	0.25	0.08	0.32
F_1 amp	** −0.45	** −0.50	−0.05	0.11	−0.25	**0.59	−0.23	0.01	** −0.53	**0.53	0.29	**0.61

Although some relationships exist between these measures, they are also largely independent response measures. $^{**}P \leq 0.01$; $^{*}P \leq 0.05$; RMS, root mean square; F_0, fundamental frequency; F_1, first formant frequencies; amp, amplitude.

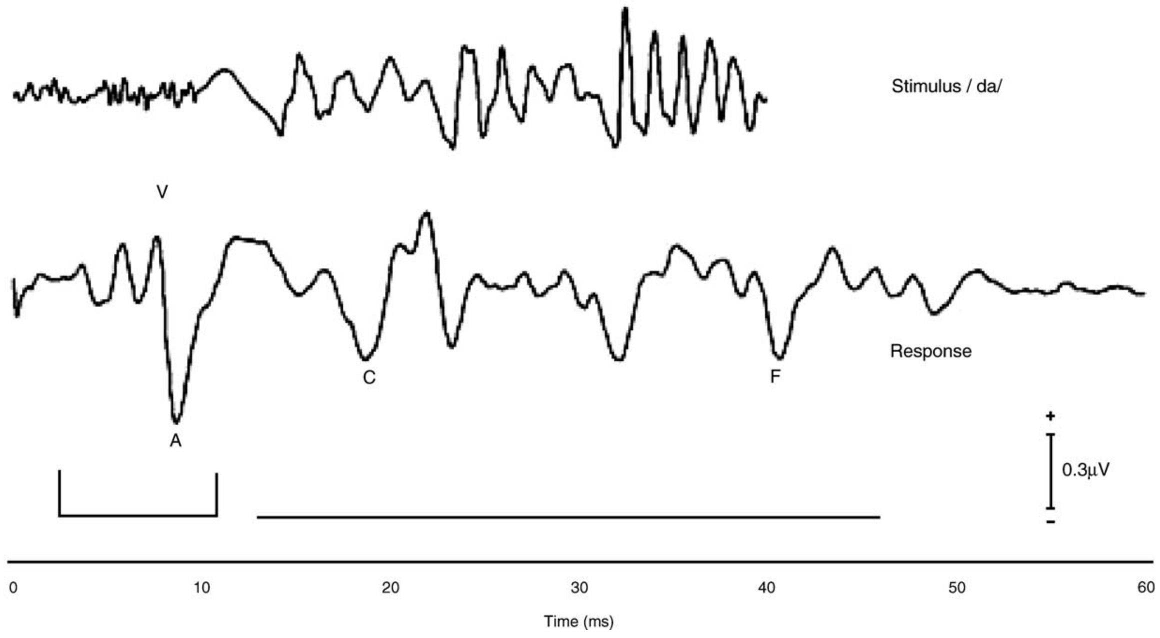

Fig. 1. Stimulus waveform (top) and grand average brainstem response in quiet (bottom). Three reliable negative peaks, waves A, C, and F, follow wave V. The onset response is bracketed, while the region containing the frequency-following response is indicated with a horizontal line.

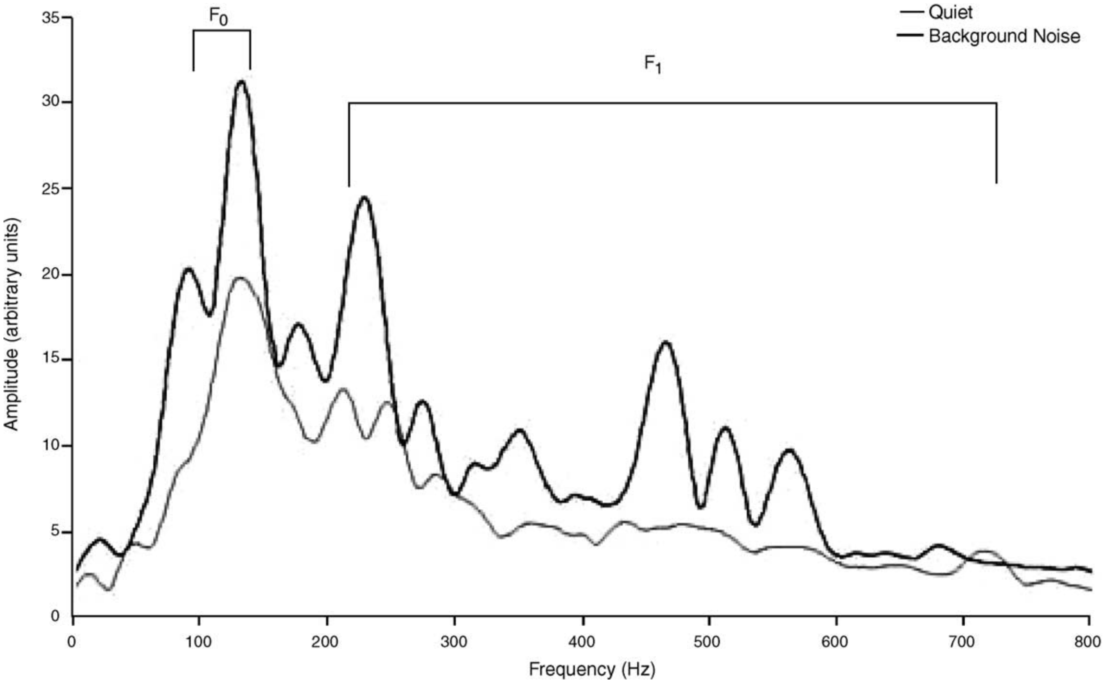

Fig. 2. Grand average frequency content in responses collected in quiet n ¼ 36and background noise n ¼ 22: Analysis of the responses indicated that only the fundamental frequency and first formant (F0 103 – 121 Hz; F1 220 – 720 Hz) were measurable, whereas the higher frequency formants were not above

the noise floor.

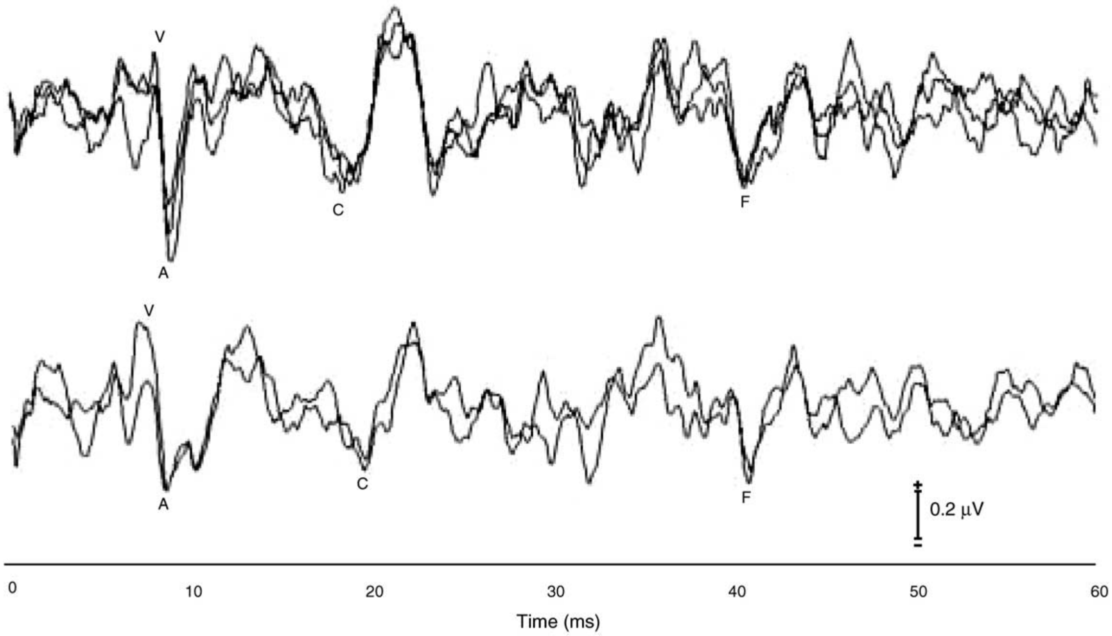

Fig. 3. Top: intra-subject, intra-test session reliability. Illustrated are three 1000-sweep subaverages that contributed to the final 3000-sweep response obtained for a representative subject. Bottom: intra-subject, inter-test session reliability. In another subject, two 3000-sweep averages were obtained on different test dates.

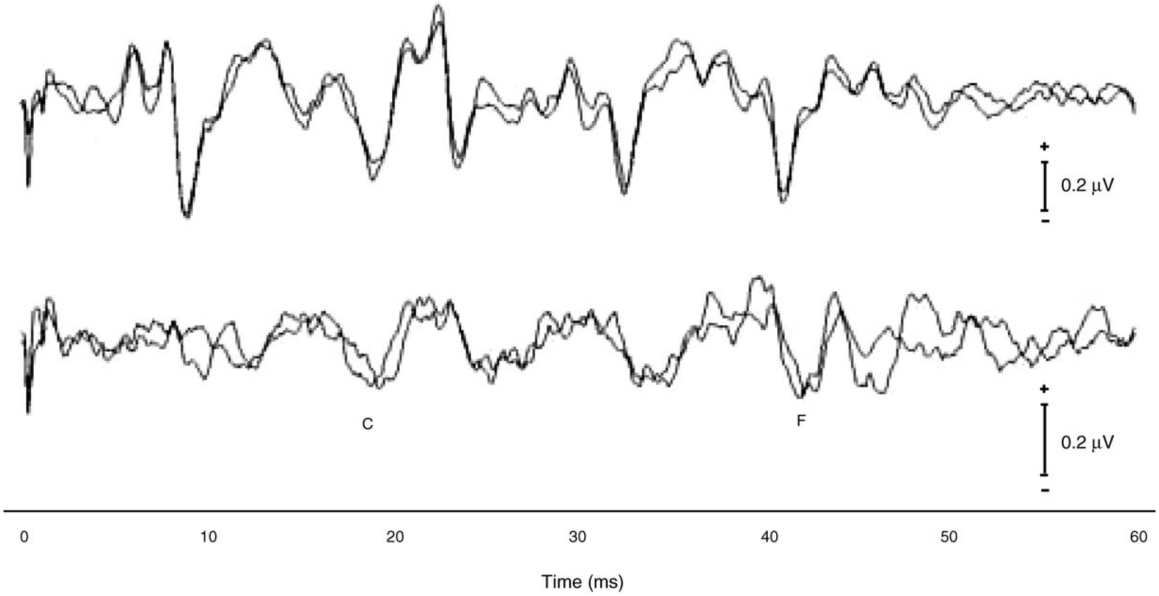

Fig. 4. Test–retest reliability. Grand average response waveforms collected in quiet (top) and background noise (bottom) at two different test sessions. Background noise effectively disrupts the onset response, while the frequency-following response remains intact.

Appendix 2

Manuscript - Audiovisual deficits in older adults with hearing impairment: biological evidence

Laurie Arum[1], Gabriella Musacchia[1], Trent Nicol[1], Dean Garstecki[1], and Nina Kraus[1,2]

[1]Auditory Neuroscience Laboratory, Department of Communication Sciences and
[2]Departments of Neurobiology and Physiology and Otolaryngology, Northwestern
University, Evanston, Illinois.

*Corresponding author
Gabriella Musacchia
Email: g-musacchia@northwestern.edu
Address: Department of Communication Sciences; Northwestern University, 2240 Campus Dr., Evanston, IL
60208, USA
Tel: 847.491.2465
Fax: 847.491.2523

Abstract

Objective: To examine the impact of hearing loss on audiovisual processing of speech. We hypothesized that hearing loss has a pervasive effect on sensory processing, extending beyond the auditory domain. In older adults with hearing loss, the ability to combine auditory with visual input is compromised such that multisensory processing is impaired.

Design: Cortical responses to speech were recorded from 24 adults (mean age 71 years), with speech presented in three conditions: when hearing a syllable "bi" (auditory), when viewing a person say "bi" (visual), and when seeing and hearing the syllables simultaneously (audiovisual). Twelve of these individuals had mild/moderate sloping sensorineural hearing loss (HI) and the other 12 had age-normal hearing thresholds (NH). All completed a vision screening, a hearing evaluation, and the Wechsler Abbreviated Scale of Intelligence.

Results: In the audiovisual condition, NH participants showed a clear and consistent decrease in P1 and N1 latencies as well as a reduction in P1 amplitude, compared to the auditory condition (A), and the sum of the unimodal components (A+V). These audiovisual interaction effects were absent or less consistent in HI participants. Effect sizes for visual modulation of auditory responses (A vs. AV) and audiovisual interaction (AV vs. A+V) were larger in the NH group, compared to their HI counterparts.

Conclusions: Hearing loss alters audiovisual processing in the central auditory system. Despite controlling for auditory sensation level, visual influence on auditory processing was significantly less pronounced in older adults with hearing impairment, indicating diminished audiovisual interaction in this population. These data corroborate animal studies showing diminished multisensory interaction following deprivation of unisensory stimulation. The clinical implication of these findings is that the effects of hearing loss extend beyond increased hearing thresholds.

Key Words: Auditory, Visual, Language, Speech, Aging

Introduction

Approximately 30% of Americans over the age of 65 have a hearing loss (Better Hearing Institute, 2005) that reduces quality of life (Dalton et al. 2003). Hearing loss primarily results in a decline in auditory speech perception (Humes & Roberts 1990). However, mounting evidence suggests that adaptation to profound hearing loss also affects cross-modal processing. Deaf individuals have an enhanced response to visual motion (Armstrong et al. 2002), and in the case of cochlear implantation following profound hearing loss, activity in both auditory and visual cortices has been shown to develop as a function of time (Giraud et al. 2001). Several lines of animal studies demonstrate a degradation in multisensory activity areas following unimodal deprivation (Heil et al. 1991;Korte & Rauschecker 1993;Rauschecker & Korte 1993;Wallace et al. 2004a). While recent evidence shows that older adults can perform as well as young adults on multisensory tasks (Laurienti et al. 2006), it is unclear whether intermediate auditory deprivation in older adults, in the form of hearing loss, impacts multisensory processing. In the current study, we explore whether hearing impaired (HI) older adults who do not use auditory amplification process the sight and sound of speech in the same way as their normal hearing (NH) counterparts.

Seeing speech has been shown to impact neural activity of the auditory brainstem and primary auditory and higher-level multisensory cortex in NH young adults (Calvert 2001;Musacchia et al. 2006b;Sams, Aulanko, Hamalainen, Hari, Lounasmaa, Lu, & Simola 1991). Evoked-potential data, which provide temporal resolution of interaction effects, consistently show that seeing speech decreases the latency of auditory cortical responses occurring at ~70-200 ms (e.g., P1, N1, and P2) (Mottonen, Krause, Tiippana, & Sams 2002;van Wassenhove, Grant, & Poeppel 2005) and also decreases amplitudes (Besle et al.

2004). Imaging data have revealed that when paired with auditory stimuli, simultaneous videos yield a superadditive increase in the blood oxygen level in auditory and multisensory areas (Calvert, Campbell, & Brammer 2000).

Although, to our knowledge, audiovisual evoked-potential studies have not been conducted in older HI adults, we do know that hearing impairment is associated with prolonged peaks and decreased amplitude of speech- evoked auditory HI young adults relative to NH counterparts (Oates, Kurtzberg, & Stapells 2002). In addition, aging impacts how auditory stimuli are processed. Older adults with and without hearing loss exhibit a degradation of N1 and P2 responses to fast temporal cues, such as voice onset time (Tremblay, Piskosz, & Souza 2003), and rapid stimulus presentation rates (Tremblay, Billings, & Rohila 2004), compared to younger adults. In order to tease out effects of age and hearing loss, we sought to control for audibility by presenting sounds at +30 dB sensation level (SL) for both HI and NH older adults. It was the subject of our investigation to determine whether HI older adults show more or less audiovisual interaction than their NH counterpart despite having equalized auditory input.

Specifically, we sought to determine 1) whether older adults with HI show audiovisual interaction to seen and heard speech and 2) whether these effects differ from the responses observed in the NH group. We hypothesized that hearing loss has a pervasive effect on sensory processing, extending beyond the auditory domain. In older adults with hearing loss, the ability to combine auditory with visual input is compromised such that multisensory processing is impaired. An alternative outcome is that audiovisual interaction is enhanced via strengthening of visual function.

We recorded cortical evoked potentials to speech in three conditions: 1) when subjects only heard the acoustic speech (auditory, A); 2) when they saw a video projection of a speaker articulating only (visual, V); and 3) when they saw and heard these tokens paired synchronously (audiovisual, AV).

Methods

Subjects

Twelve older adults with NH (mean age 70.8, SD = 4.5) and twelve older adults with HI (mean age 72.2, SD = 3.7) were recruited from the Buehler Center on Aging research registry, and from visitors, staff, and continuing education attendants of the Northwestern University campus. All subjects were native English speakers. Subjects who had worn hearing aids in the past five years, had a history of neurological disorders (e.g., seizures, cerebral palsy, spina bifida, or any syndrome associated with central or peripheral nervous system), and exhibited below normal cognitive function were excluded. Testing required a 3.5-hour session, which most participants completed in one day. Each participant signed an informed consent form prior to the commencement of the experiment in accordance with the Institutional Review Board procedures at Northwestern University and was compensated for his or her participation.

Cognitive function was assessed with the Wechsler Abbreviated Scale of Intelligence (WASI) Full-4 to measure verbal, nonverbal, and general cognitive function. Mean intelligence quotient (IQ) scores for older adults with NH and HI were 123.0 (SD = 10.9) and 115.3 (SD = 10.7), respectively. IQ scores did not differ between groups ($t=1.75$, $p=0.093$).

Subjects whose normal or corrected-to-normal visual acuity, as assessed with a Snellen 10-foot eye chart, exceeded 10/15 were excluded from participation. The audiologic assessment included pure-tone air-conduction and bone-conduction thresholds, as well as binaural speech audiometry. Hearing sensitivity was established for a broad range of frequencies in each ear using the GSI 61 Audiometer (Grason-Stadler), in a sound-treated room and with sound delivered through insert earphones (Etymotic Research, ER-3A). Bone conduction testing was performed to confirm NH sensitivity and to differentiate between sensorineural and conductive hearing loss. Subjects with a conductive hearing loss were excluded from the study. Speech testing consisted of binaural word recognition test presented at a subjectively determined most comfortable level (around 30-40 dB sensation level (SL) above pure tone average). A binaural speech awareness threshold was determined using the same auditory stimulus used for the neurophysiological recordings (“bi”). The participants were instructed to say “yes” when the syllable became audible and distinct.

Subjects with either normal or a symmetric sloping sensorineural hearing loss no greater than a moderately-severe degree (<80 dB HL) in the speech frequencies (500-4000 Hz) were selected for this study. For this study, subjects were considered hearing impaired if pure-tone thresholds exceeded the normative sensitivity values in at least two or more frequencies set forth by the ISO Standard on hearing by age and sex (ISO 7029:2000 Acoustics). Pure-tone thresholds were assessed for both groups at 250, 500, 1000, 2000, 4000, and 8000 Hz (Table 1). Larger standard deviations in the HI group are due to the sloping configurations of their hearing loss. Because the groups diverged most in the 2-8 kHz range, a mean threshold value over this frequency region was calculated for subsequent correlational analyses.

Stimuli and Presentation Sequence

The acoustic stimulus consisted of a five-formant, 430 ms, synthetic speech syllable, “bi”, created with a DH Klatt synthesizer. The visual stimulus was a digital recording of a male speaker saying the "bi" syllable in a clear speaking style (Figure 1). The articulation was contained within 37 frames that began and ended with the same neutral resting position. Stimulus onset asynchrony for the visual stimuli was 1.3 seconds. Frames 16-18 depicted the release of the consonant, which coincided with auditory onset at 570 ms. Hence, time 0 will be referenced to acoustic onset. The visual stimuli were projected 1.8 meters in front of the subject, with a visual angle of 38 degrees, and the auditory stimulus, "bi", was presented bilaterally through insert earphones (Etymotic Research, ER-3A). Delivery of the audio and visual stimuli was controlled by Presentation software (Neurobehavioral Systems, Inc., CA). Stimuli were presented in the A, AV, and V conditions randomly throughout the testing session. The auditory stimulus was presented at 30 dB SL above the speech awareness threshold determined during the hearing evaluation. Presentation at a sensation level was employed to ensure equal audibility across all participants. In order to control for attention, participants were asked to silently count the number of catch trials (15% deviant rates). Catch trials consisted of sporadic visual projections of a red asterisk and were interspersed randomly (15%) throughout the presentation sequence. No difference was seen in catch trial percent error across groups ($M_{NH} = 17.5$, $SD = 10.76$; $M_{HI} = 13.16$ $SD = 5.98$; $t = 1.219$, $p = 0.236$). Eight blocks of 75 stimulus repetitions were presented to each subject, with a short break between each block. A long break was provided midway through the testing session. In all, 200 responses were recorded in each of the three conditions.

Recording Parameters

Neurophysiological recordings were conducted in a sound-attenuated booth. Cortical responses to speech stimuli were acquired with Neuroscan 4.3 (Compumedics, El Paso, TX, USA) using Ag-AgCl electrodes (impedance <5 kΩ). Reference, ground, and eye blink monitor electrodes were placed on the nose, forehead, and superior (VEOG) and outer

(HEOG) canthus of the left eye, respectively. Seven active electrodes, placed according to the 10-20 International system, were positioned on F3, F4, Fz, Cz, and Pz. Continuous EEG was recorded with a band pass filter from 0.5 to 30 Hz at a sampling rate of 1000 Hz. Offline processing included dividing continuous EEG into epochs from -150 to 1000 ms post-acoustic onset. An artifact criterion was applied to HEOG and VEOG channels in order to reject those epochs that contained myogenic and eye-blink artifacts. Any epoch with a voltage exceeding +/- 65 µV was omitted from the average. The artifact-free epochs (M=418 sweeps, SD=14, per condition**)** were then averaged, according to stimulus type.

Response Measurements

Signal-to-noise ratio (SNR) was calculated by dividing the root mean square (RMS) amplitude of the post-stimulus period by the RMS amplitude of the pre-stimulus period. Only responses with >3 SNR were used for data analysis. This eliminated 1 HI and 1 NH subjects for subsequent analyses. Peaks P1, N1, and P2 were chosen by visual inspection for all participants, in all conditions, and by two raters “blind” to group and condition. The timing and size of cortical responses were quantified by peak latency and amplitude measures, respectively. To assess the impact of visual stimuli on auditory responses, effect sizes (A vs AV and A+V vs AV) were calculated. Effect size not only provides information on how distinct AV latencies and amplitudes were from those of the auditory alone or the summed auditory and visual responses, but also the direction of the differences based on the sign of the effect size. A positive effect size would indicate earlier or smaller peaks in the AV condition.

Statistical Analysis

Comparisons were made between the physiological responses of the NH and HI groups across conditions. Specifically, modulation and interaction effects were investigated.

Modulation effects are changes in the acoustic response due to the addition of visual stimuli; thus, how AV response differs from A. Interaction effects are differences between the AV responses and the summed unimodal responses, such that the measured AV activity is nonlinear (AV≠A+V). Repeated measure 2X2X5 ANOVAs (independent factor: group (NH, HI); repeated factors: condition (A, AV or A+V, AV) and electrode (F3, F4, Fz, Cz, and Pz)) and *post hoc* t-tests, when applicable, were employed with P1, N1, and P2 latency and amplitude measures to test whether responses in the A, AV, and A+V conditions differed. In order to assess relationships between audiovisual processing and hearing loss, we correlated composite measures of visual modulation (AV-A) on peak latencies and amplitudes with average hearing thresholds.

Results

In all subjects, the first prominent peak, P1, was followed by a negative trough, N1, and a positive peak, P2. P1, N1, and P2 latencies were similar to previously reported normative values (Hall 1992). No group differences were observed for latency or amplitude measures of unimodal auditory responses (Table 2).

Audiovisual modulation and interaction of peak latencies in NH and HI older adults

To assess the impact of visual stimuli on the auditory response in NH and HI subjects, 2X2X5 repeated ANOVAs were performed for each peak latency (P1, N1, and P2). The same analyses were performed with AV and A+V as the condition factor to determine whether AV responses could be solely accounted for by the superposition of auditory and visual evoked potentials.

P1 latency within-subject effects included a main effect of condition ($F = 7.251$, $p = 0.013$), an interaction between condition and group ($F = 5.020$, $p = 0.035$) and a triple interaction between condition, electrode and group ($F = 5.922$, $p < 0.001$). This indicated that seeing a speaker impacted P1 latencies in the NH group differently than in the HI group and that the degree of modulation was different across electrodes. Post-hoc paired t-tests showed that NHs had earlier P1 latencies in the AV condition, relative to the A, at F4, Fz, Cz and Pz, whereas the HI group did not (Table 3, Figure 3A). A main effect of condition was observed for the AV vs. A+V contrast ($F = 12.546$, $p = 0.002$), indicating that the effects of visual influence on P1 latency were non-additive interactions. NHs had earlier P1 latencies in the AV condition, relative to the A+V, at F4, Fz, Cz and Pz (Table 3). The HI group failed to show an AV interaction at any electrode (Table 3, Figure 3B).

The same analyses described above were performed with N1 latencies. There was a main effect of condition (F = 21.618, $p < 0.001$). However, no interaction effect of group was observed. An interaction was observed between condition and electrode for N1 latencies (F = 2.959, p = 0.024). This indicated that the visual influence impacted auditory processing at N1 differently across electrodes. NH and HI groups showed a similar pattern of modulation at N1, with earlier latencies in the AV condition for most electrodes (Table 4, Figure 4). A within-subject main effect of condition was also observed for N1 latencies in the AV and A+V contrast (F = 26.616, $p < 0.001$). No interaction effect of group or electrode was observed with this ANOVA. This indicated that N1 latencies in the AV condition differed from those obtained in the summed A+V condition in a similar way across electrodes and in both groups (Table 4, Figure 4).

A main effect of condition was seen in the AV vs. A contrast for P2 latency (4.530, p = 0.045). Post-hoc paired t-tests at P2 showed that the NH group had earlier latencies in the AV condition, compared to the A, at F4 (t = 2.518, p = 0.029) and Fz (t = 2.238, p = 0.047). The HI group showed no modulation effect at any electrode. A main effect of condition was seen in the AV vs. A+V contrast for P2 latency (F = 9.784, p = 0.005), and post-hoc paired t-tests showed that the difference at F4 in the NH group was non-additive (t = 2.780, p = 0.018).

Audiovisual modulation and interaction of peak amplitudes in NH and HI older adults

As above, 2X2X5 repeated ANOVAs were performed with group as the between-subjects factor for each peak amplitude (P1, N1, and P2). Audiovisual modulation was assessed by comparing A with AV amplitudes and interaction was determined by the AV vs. A+V contrast.

For P1 amplitude, there were main effects of condition (F = 24.722, p < 0.001) and electrode (F = 11.231, p < 0.001), as well as a trend toward the interaction of condition and group (F = 4.171, p = 0.053). Paired t – tests showed that NHs had reduced P1 amplitudes in the AV condition for all electrodes, whereas the HI group showed the same effect only at Fz (Table 3, Figure 5). The ANOVA for the AV vs. A+V contrast showed that the differences at F3 and F4 were non-additive in the NH group. No AV vs A+V comparisons were significant in the HI group (Table 3, Figure 5).

The ANOVA analysis for N1 amplitudes revealed main effects of condition (F = 22.000, p < 0.001) and electrode (F = 19.000, p < 0.001), as well as an interaction between condition and electrode (F = 19.000, p = 0.018). Visual stimuli suppressed N1 amplitudes for both groups: at all electrodes in NHs and at all but F4 in the HIs (Table 4). The ANOVA for the AV vs. A+V contrast showed main effects of condition (F = 18.312, p < 0.001) and electrode (F = 9.214, p < 0.001) and an interaction between electrode and group (F = 2.610, p = 0.041). The post-hoc paired t-test comparisons for this contrast revealed that only the NHs had non-additive effects for N1 peak amplitudes (Table 4).

The ANOVA results for visual modulation of P2 amplitudes showed main effects of condition (F = 5.223, p = 0.032) and electrode (F = 6.039, p < 0.001). Post-hoc paired t-tests revealed that the NH group had suppressed P2 amplitudes in the AV condition at the midline electrodes ($t_{Fz} = 2.362$, p = 0.038; $t_{Cz} = 3.489$, p = 0.005; $t_{Pz} = 5.440$, p < 0.001). The HI group showed no such effect. The effect in the NH group at the midline electrodes was non-additive, according to the AV vs. A+V ANOVA (main effect of condition: $F = 36.178$, $p < 0.001$; triple interaction between condition, electrode and group: $F = 6.458$, $p < 0.001$) and post-hoc paired t-tests ($t_{Fz} = 3.665$, $p = 0.004$; $t_{Cz} = 5.128$, $p < 0.001$; $t_{Pz} = 6.112$, $p < 0.001$).

Modulation and interaction effect sizes in NH and HI older adults

Overall, AV-A effect sizes were medium (>0.5) or large (>0.8) (Cohen 1969, p 23) in the NH group and at least one magnitude category larger than HIs at most electrodes for P1 latency, N1 latency and P1 amplitude measures (Figure 6). In addition, 83% of interaction effect sizes (AV vs A+V) were medium or large for these measures in the NH group whereas the only had 33% of HI measures showed medium or large effect sizes. It is important to note that these differences persisted despite a lack of group differences in the auditory alone condition across all latency and amplitude measures.

Relationships between audiometric thresholds and composite measures of audiovisual modulation

In order to compare visual influence with audiometric thresholds groups, two composite measures of audiovisual modulation were computed. First, P1 latencies in the AV condition were subtracted from those in the A condition at each electrode to produce a measure of AV latency decrease. AV latency decrease was then averaged across all electrodes. Second, AV amplitude values were subtracted from A amplitudes to yield a measure of AV amplitude suppression for each electrode. Since visual influence on P1 amplitude was most significant over the frontal electrodes, AV amplitude suppression was averaged across F4, F3, Fz, and Cz electrodes. Analysis of these two composite measures showed that visual impact on auditory responses was greater in NH subjects than in HIs, for both AV latency decrease ($t=2.37$, $p=0.04$) and AV amplitude suppression ($t=2.19$, $p=0.04$).

Relationships between hearing thresholds for high-frequency pure tones (2, 4, and 8 kHz) and composite measures of visual influence were observed. Right ear average thresholds for 2-8 kHz pure tones correlated with P1 latency decrease ($r_{RE} = -0.43$, $p<0.05$) (Figure 7A). Better (lower) hearing thresholds were associated with greater visual influence on latency (i.e., a larger latency decrease from A to AV conditions). Left ear thresholds exhibited the same trend but did not reach statistical significance. Right and left ear thresholds over these same frequencies showed the same relationship with AV amplitude suppression ($r_{LE} = -0.46$, $p<0.05$; $r_{RE} = -0.43$, $p<0.05$) (Figure 7B & C), such that better thresholds were associated with greater visual influence on peak amplitude (i.e., more suppression of AV amplitude, compared to A).

Discussion

Many older adults with hearing impairment rely on visual cues to compensate for the degraded auditory signal (Erber 1972). Recent data, however, suggest that hearing-impairment in older adults is not necessarily associated with better audiovisual integration abilities (Tye-Murray, Sommers, & Spehar 2007). Here we show that the underlying neural mechanisms of audiovisual interaction in HI older adults are impaired, relative to NHs in the same age group. Specifically, NH older adults consistently exhibit decreased peak latencies and decreased amplitudes to bimodal speech, relative to auditory and the linear summation of unimodal responses. Although HI older adults demonstrate a similar pattern of interaction effects, these differences are smaller, less consistent, and different from the A+V responses less often. The difference between NH and HI older adults in their audiovisual modulation and interaction is most salient for early peaks of the cortical response.

These findings indicate that hearing loss in the aging population impacts neural mechanisms of audiovisual interaction. It has been suggested that auditory mechanisms of excitation and inhibition may be degraded in the aging system, especially when faced with fast time-varying acoustics (Tremblay, Piskosz, & Souza 2003). Here, we show that visually-induced inhibition and latency decrease of auditory responses is diminished in hearing impaired older adults. This suggests that visually-mediated mechanisms of excitation and inhibition are impaired in this population. Our data suggest that hearing-related factors may compound AV processing and could be responsible for some of the perceptual difficulties experienced by older adults with hearing loss.

One hypothesis for how this may occur is that degraded unimodal input debilitates multisensory mechanisms over time. As older adults lose their hearing, the unimodal acoustic input becomes increasingly less audible, or degraded. In animal studies, profound unimodal degradation, in the form of total deprivation, changes the functional properties of neurons in multisensory regions and unisensory regions of the intact modality. Moderate hearing loss over time may engender similar mechanisms of plasticity and likewise impact multisensory regions. Conversely, lifelong experience with language and music enhances both auditory and audiovisual processing (Gaser & Schlaug 2003a;Musacchia et al. 2007b;Wong et al. 2007b;Zatorre 1998b).

An alternative interpretation is that attentional mechanisms may be excessively impacted by hearing impairment in older adults. For example, visual articulatory motion that precedes acoustic onset cue a listener to focus their attention when the speaker releases the consonant. In this case, seeing a speaker prepare to say “bi” would increase expectation, and hence attention to the auditory signal. This hypothesis is unlikely in this experimental paradigm because we observed decreased amplitudes with the addition of visual stimuli whereas increases in attention to auditory stimuli are typically associated with increases in amplitude. In addition, attentional differences across groups would likely be reflected through differences in task accuracy, whereas, no difference was seen in task accuracy between NH and HI groups.

Because the high frequency thresholds are most impaired in our HI group, these results may also reflect a difficulty hearing high-frequency components of the consonant “b”. High-frequency information in consonant-vowel syllables is concentrated in the consonant portion of the sound, where visual articulation gestures differ most. It is possible that HI subjects received less high-frequency information despite the +30 dB SL presentation level. In this case, abnormal audiovisual interaction in HI subjects may still be related to degraded input. However, audiovisual integration has been shown to operate inversely to stimulus salience, such that lower unimodal salience results in larger integration effects (Meredith & Stein 1986a). Given this, one would predict enhanced AV effects in the HI group. Thus, we conclude that despite possible differences in high-frequency input levels between the two groups, audiovisual mechanisms are degraded in older adults with hearing loss.

Implications for rehabilitation and research

The clinical implication of these findings is that hearing loss has a deleterious effect on how the sight and sound of speech are integrated in the older adult nervous system. It is well known that auditory and visual speech cues are essential to hearing impaired populations, especially in noisy conditions. A remaining question is whether auditory-visual training can improve perception and processing in this population. Although audiovisual training has not been investigated to date, auditory training has been shown to impact early cortical components of auditory evoked potentials in young adults and language-impaired children. Normal hearing adults who were trained to discriminate small differences in syllable contrasts, such as VOT durations, showed training-related increases in P1, N1 peak amplitudes and mismatch negativity components(Tremblay et al. 1997;Tremblay et al. 2001). A subsequent study suggested that the capacity for short-term training related plasticity may be greatest in the right hemisphere (Tremblay & Kraus 2002). These data are particularly encouraging given that visual influence on auditory processing was more prevalent at F4 than F3.

Currently, there are a handful of tests and training materials available for auditory-visual communication assessment and treatment. Noteworthy examples include the following. Tyler, Preece, & Tye-Murray (1986) developed the Iowa Consonant Confusion Test, which measures auditory-visual perception of phonemes and everyday sentences. Plant (2001) created the Auditrain program containing analytic and synthetic sentence materials for development of auditory-visual perception skills in cochlear implant users. Seeing and Hearing Speech (Sensimetrics Corporation, Somerville, MA) is a computerized home-study program that emphasizes combined auditory and visual cues of everyday communication in varied levels of background noise. Conversations Made Easy (Tye-Murray, 2002), a computerized program distributed by the Central Institute for the Deaf (St. Louis, MO), presents sentences and everyday scenarios in closed set format for training in an auditory-visual mode or a visual mode alone. Further studies are needed to understand how buttressing such auditory-visual training programs with auditory and audiovisual cortical response measures can help understand and assess multi-sensory training effects.

Acknowledgements

Numerous people in the Auditory Neuroscience Laboratory have assisted in the progress of this project. Special thanks go to Lauri Olivier, Ph.D., Erika Skoe, Erin Hayes, Ph.D., and Judy Song. Appreciation is given to Lynne Bernstein, Ph.D. for the use of her stimuli. This work was supported by NIH R01 DC01510 and NSF 0544846

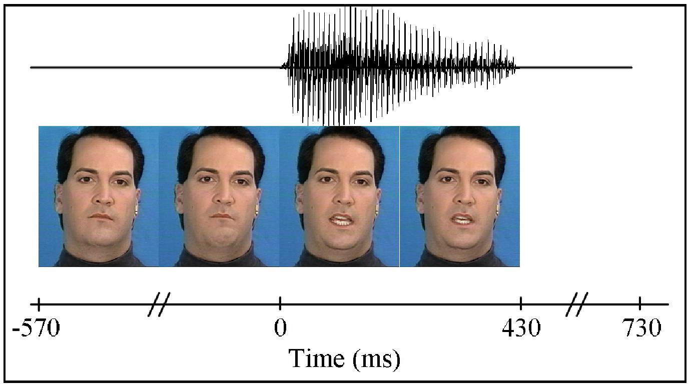

Figure 1. Auditory and visual components of speech stimuli. The waveform of the acoustic stimulus (top) and visual articulation (bottom) are depicted. The visual component was digitized from a video of a speaker uttering “bi”. Acoustic onset occurred 570 ms after the first video frame and simultaneously with the release of consonant closure. Duration of the acoustic stimulus was 430 milliseconds.

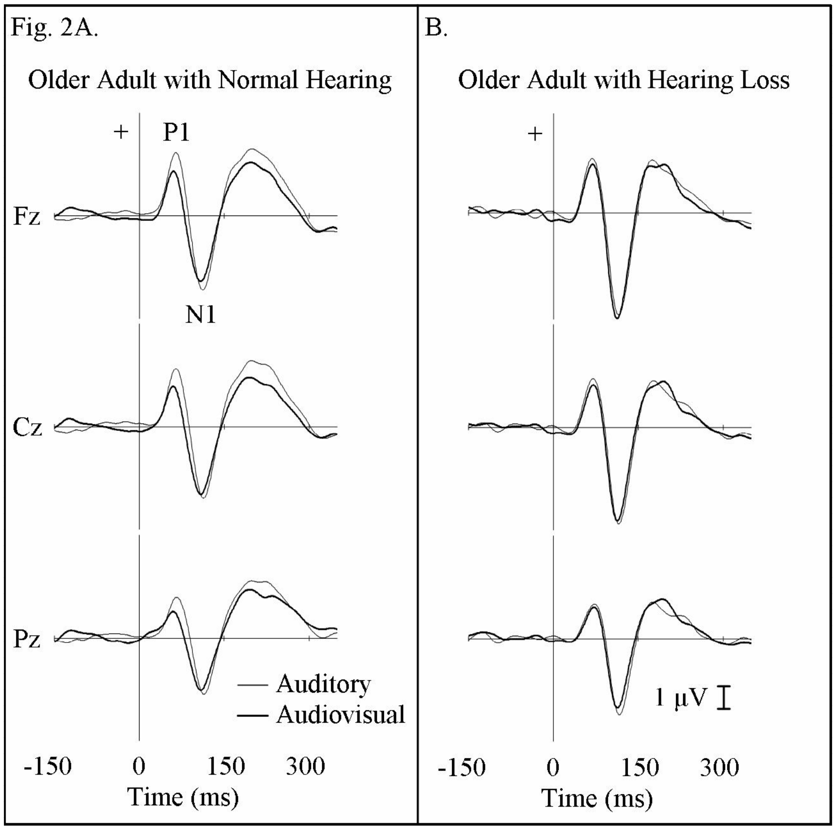

Figure 2. Example waveforms from typical normal hearing and hearing impaired subjects at the midline electrodes. A. This NH subject shows that P1 peak latency is earlier and amplitude is smaller in the Audiovisual (black) response, compared to Auditory (grey), across all midline electrodes. B. The HI subject shows little difference between audiovisual and auditory responses.

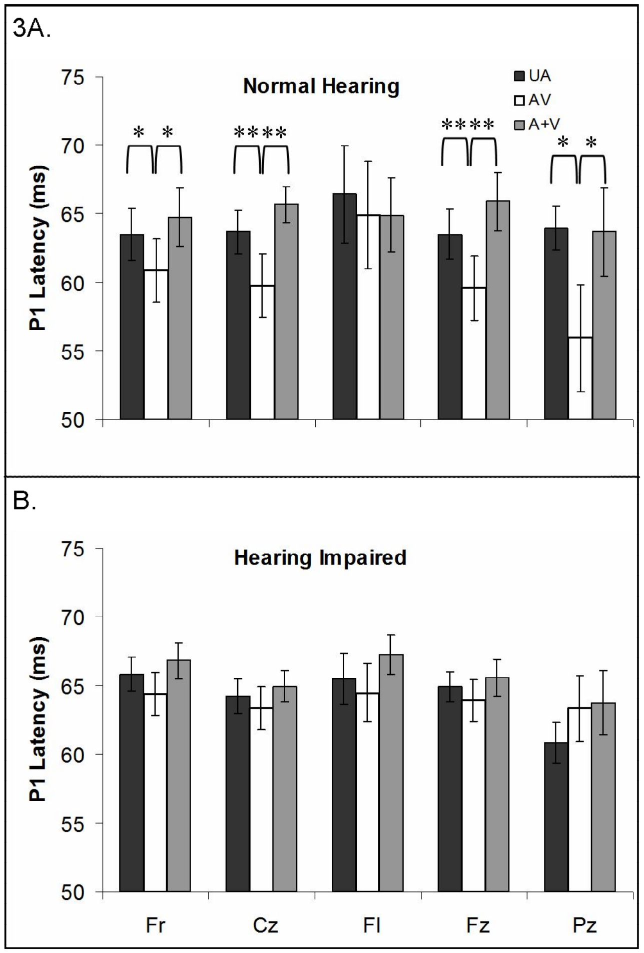

Figure 3. Mean P1 latencies in auditory (A), audiovisual (AV) and summed (A+V) conditions for NH and HI groups. A. NH subjects showed earlier P1 latencies in the AV condition, compared to both A and A+V conditions at F4 and midline electrode sites. B. No P1 latency modulation (A vs. AV) or interaction (AV vs. A+V) effects were observed in the HI group.

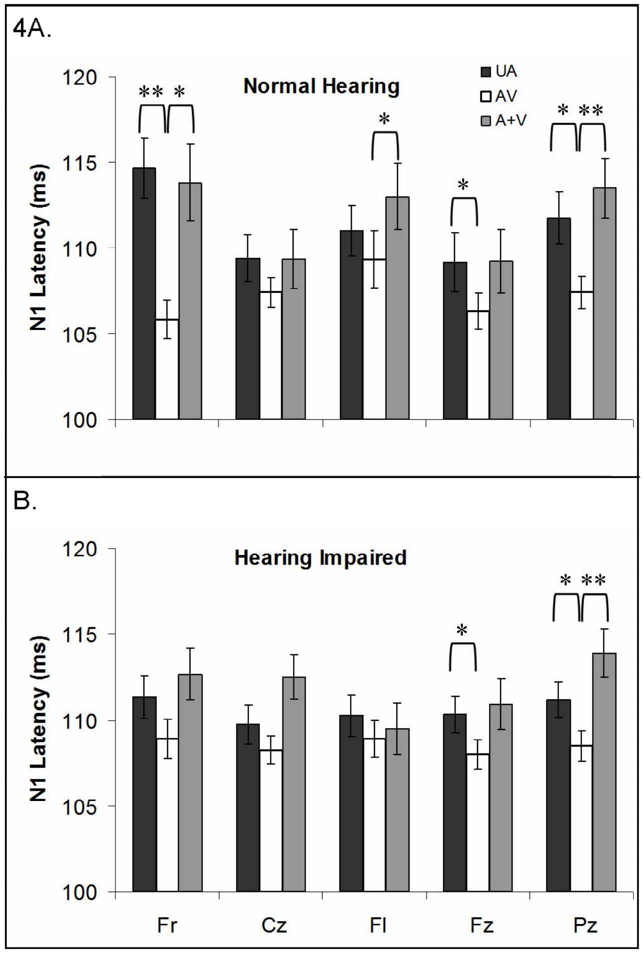

Figure 4. Mean N1 latencies in auditory (A), audiovisual (AV) and summed (A+V) conditions for NH and HI groups. A. NH subjects had earlier N1 latencies in the AV condition, compared to both A and A+V conditions at F4 and Pz. AV latencies were earlier than the sum of the unimodal components at F3. At Fz, AV latencies were earlier than those in the A condition, however, AV latencies did not differ from the summed components at this electrode site. B. HI subjects showed modulation (A vs. AV) and interaction (AV vs. A+V) effects at Pz only. The exhibited modulation only at Fz; interaction only at F4.

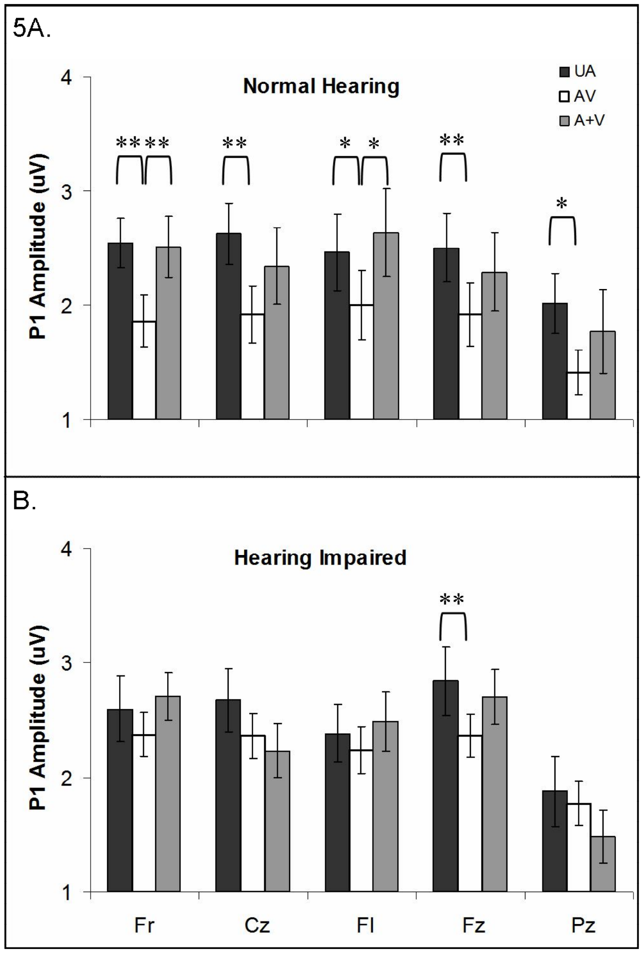

Figure 5. Mean P1 amplitudes in auditory (A), audiovisual (AV) and summed (A+V) conditions for NH and HI groups. A. NH subjects showed smaller P1 amplitudes in the AV condition, compared to both A and A+V conditions at F3 and F4 electrode sites. AV amplitudes were smaller than A at the midline electrodes, however, no interaction effects (AV ≠ A+V) were observed. B. No interaction effects were observed in the HI group. Only Fz showed smaller AV amplitude than A in the HI group.

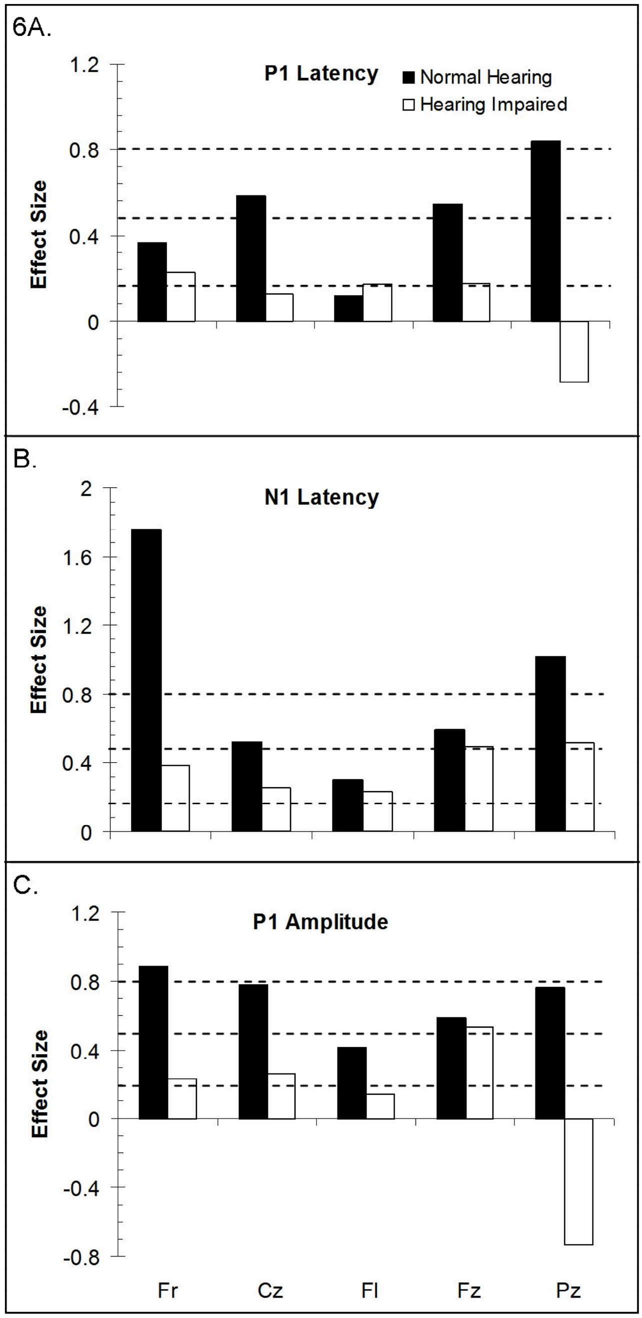

Figure 6. Modulation and interaction effect sizes for NH and HI groups. A. P1 latency modulation (solid bars) and interaction (lined bars) effect sizes were larger in the NH group (black bars) than in the HI group (grey bars) at F4 and the midline electrodes. B. N1 latency modulation effect sizes were at least one size category larger for NHs than HIs at F4, and the midline electrodes. AV interaction effect sizes showed this same pattern at all channels except for Pz. C. NHs showed medium or large modulation effect sizes for P1 amplitude at all electrodes, except for F3, whereas the HI group showed two medium effect sizes (F3 and F4). Interaction effects were larger in NHs than in the HIs at all but the Fz electrode.

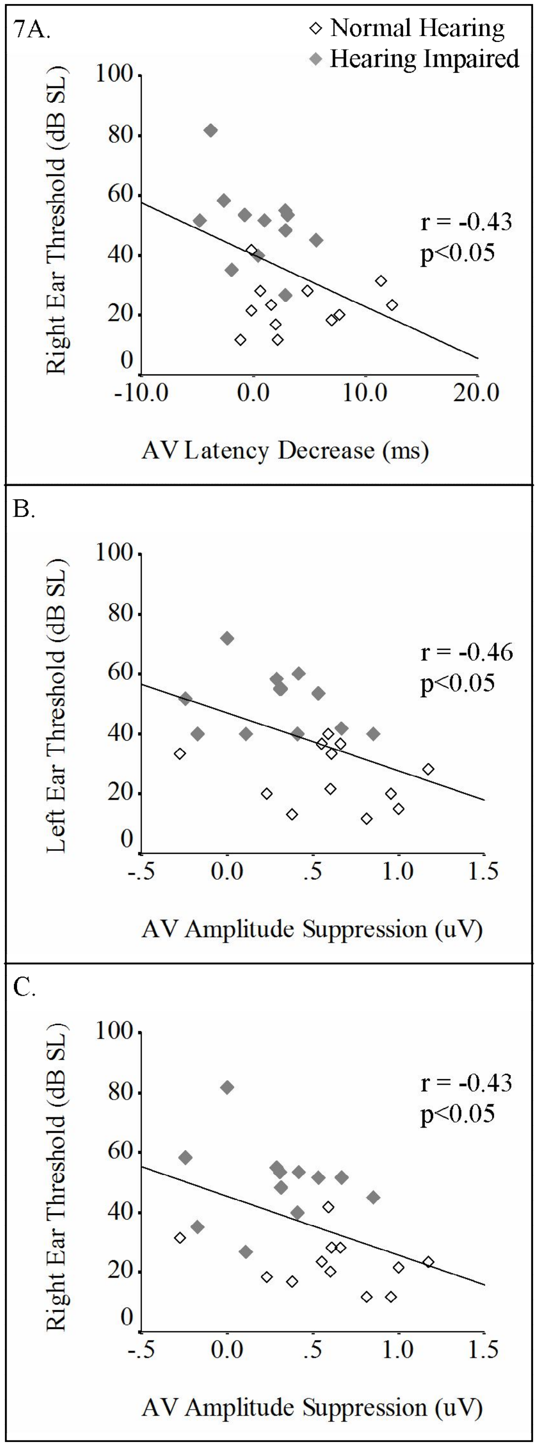

Figure 7. Correlations between composite measures of visual influence and audiometric thresholds. A. AV latency decrease at P1 (AV latency - A latency) was calculated for each individual at each electrode and averaged across all electrodes. A negative correlation was seen between the size of AV latency decrease and right ear threshold. B & C. AV amplitude suppression at P1 (A amplitude - AV amplitude) was calculated for each individual at each electrode and averaged across the frontal electrodes. A negative correlation was seen between the size of AV amplitude suppression and threshold for both left (B) and right (C) ears.

Curriculum Vitae

GABRIELLA A.E. MUSACCHIA

PERSONAL INFORMATION

Address	Auditory Neuroscience Laboratory Northwestern University 2240 Campus Drive Evanston, IL 60613
Phone	847.491.2465
E-mail	g-musacchia@northwestern.edu
Position	Doctoral Candidate, Northwestern University
Defense Date:	September 2007

EDUCATION

1996	A.S. in Biology, City College of San Francisco, S.F., CA Dean's List Academic Honor, 1995, 1996 City College of San Francisco
2000	B.A. in Psychobiology, University of California at Santa Cruz, CA
2001-present	Ph.D., Double Major in Communication Sciences and Disorders\Language and Cognition, Northwestern University, Evanston, IL

PROFESSIONAL EXPERIENCE

6\00 - 1\01	Research Technician I for Stephen G. Lisberger, Ph.D. Keck Center for Integrative Neuroscience, U.C.S.F.
1\01 - 9\01	Research Technician II for Leonard Trejo, Ph.D. and director, Chuck Jorgensen, Ph.D. N.A.S.A. Ames Research Center, Mountain View CA

RESEARCH EXPERIENCE

6\00 - 1\01	Experimental analysis of single unit responses to apparent motion in non-human primates. Keck Center for Integrative Neuroscience, U.C.S.F., under the direction of Stephen G. Lisberger, Ph.D.
9\01 - 12\01	Cl- ion channel health in prevention of sensorineural hearing loss and preservation of Otoacoustic Emissions (OAE) in gerbils. Northwestern University, under the direction of Jonathon Siegel, Ph.D.
1\02 - 9\02	Effect of Cl- channel blocking agents on outer hair cell (OHC) membrane capacitance and physiology. Northwestern University, under the direction of Peter Dallos, Ph.D. and Enrique Navarrete, Ph.D.
1\02 - present	The main work of my doctoral study, under the direction of Dr. Nina Kraus, has been to investigate the neurophysiological bases of real-world (unimodal and multisensory) speech and music perception in normal and learning-disabled populations. Currently, I am investigating non-auditory (e.g., visual, attentional)

influences on auditory brainstem and cortical responses to speech and music. I use electroencephalographic (EEG) recordings, psychophysical and behavioral measures to test prominent theories of multisensory perception and plasticity such as the Feature Integration Theory (Treisman, 1993), the Fuzzy Logic Model of Perception (Massaro, 1980) and the Reverse Hierarchy Theory (Ahissar, 2004). In Dr. Kraus' lab I have developed skills in scientific writing, experimental design, EEG recording, database management, statistical analysis, teaching and mentorship.

TEACHING EXPERIENCE

2004 Teaching Assistant and Laboratory Instructor. Conducted lab exercises and assisted with test compilation and correction for Central Auditory Neurophysiology of Evoked Potentials (CSD 425) under the direction of Nina Kraus, Ph.D.

2004 Guest lecturer. Music Acoustics lecture for Biological Foundations of Speech and Music (CSD 310):

2005 Teaching Assistant and Laboratory Instructor. Conducted lab exercises and assisted with test compilation and correction for Biological Foundations of Speech and Music and Central Auditory Neurophysiology of Evoked Potentials (CSD 310 and 425) under the direction of Nina Kraus, Ph.D.

2005-7 Guest lecturers. Music Acoustics and Music Encoding lectures for Biological Foundations of Speech and Music (CSD 310):

2006-7 Student Mentor. Train and supervise undergraduate, Au.D. and new Ph.D. student projects.

RESEARCH INTERESTS

Dissertation: Visual influence on auditory processing of speech and music.
Adviser: Nina Kraus, Ph.D.
Laboratory: Auditory Neuroscience Laboratory, Northwestern University
Defense Date: September 2007

Because our senses do not act as independent channels of perception, my research interests are focused on how non-auditory factors (e.g. multimodal, training, and attentional) influence auditory encoding. Current theories posit that lower levels (i.e. brainstem) of the auditory pathway are immune to non-auditory influence. However, my previous and current studies are part of a growing body of neurophysiological data that shows tantalizing evidence of dynamic peripheral sensory mechanisms. At present, my emphasis is on investigating how visual stimuli influence the auditory brainstem response to biologically relevant sounds, such as speech and music. By using music, I can also ask whether lifelong experience with this stimulus, as with professional musicians, engenders plasticity. It is through this and future work, that I hope to better understand the neural basis of our dynamic sensory system and audition at various representational levels.

Dissertation Abstract:

The current proposal aims to determine whether low-level, early interaction effects as those previously observed with speech can be elicited by non-speech stimuli. The central hypothesis is that audiovisual encoding mechanisms abide by similar principles for biologically relevant sounds. Although interaction effects have been observed with speech and non-speech stimuli, the magnitude of these effects has not been directly compared. Additionally, when speech and non-speech have been compared, the biological significance of the types of stimuli has been unequal. Here, I intend to compare visual influence on auditory brainstem responses using speech and music tokens. Music was chosen because music is also biologically significant to humans, especially for professional players. Specifically, we aim:

1) To determine whether concordant visual stimuli influence auditory brainstem encoding mechanisms for speech and music stimuli.

2) To determine whether brainstem mechanisms of auditory and audiovisual processing are plastic, or malleable with complex cognitive training, such as is undertaken during musical training.

Results from data collected thus far show that seeing and hearing enhances the magnitude of the brainstem response to speech and music in all subjects. In addition, musicians show enhanced auditory responses to speech, relative to their non-musician controls as well as larger degree of visual influence in both speech and music responses. These results, as well as the continuation of the proposed dissertation will give insight into the underlying mechanisms of human communication and elucidate effects of musical training on the peripheral and central processing stream.

PUBLICATIONS

Russo N, Nicol T, **Musacchia G**, Kraus N. Encoding of Speech Sounds in Quiet and Background Noise in the Brainstem: Normal and Learning Impaired Children. Association for Research in Otolaryngology MidWinter Meeting, February 2004. St. Petersburg, USA.

Russo N., Nicol T, **Musacchia G**, Kraus N. Brainstem responses to speech syllables. Clin Neurophysiol. 2004 Sep;115(9):2021-30.

Musacchia G, Sams M, Nicol T, Kraus N. Seeing speech affects acoustic information processing in the human brainstem. Exp Brain Res. 2006 Jan;168(1-2):1-10. Epub 2005 Oct 11.

Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced auditory and audiovisual subcortical sensory processing to speech and music. Proc Natl Acad Sci. 2007 Oct 2;104(40):15894-8.

Musacchia G, Strait D, Nicol T, Kraus N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians. Hearing Research. 2008. *In Press*.

Arum L, **Musacchia G**, Garstecki D, Kraus N. Audiovisual deficits in older adults with hearing impairment. Ear Hear. May 2007. *Submitted*.

ABSTRACTS

Musacchia G, Sams M, Nicol T, Kraus N. Seeing speech affects acoustic information processing in the human brainstem. Association for Research in Otolaryngology MidWinter Meeting, February 2004. Daytona, USA.

Musacchia G, Sams M, Nicol T, Kraus N. Speech-evoked audiovisual interaction in the human brainstem, thalamus and cortex. Society for Neuroscience Annual Meeting, October 2004. San Diego, USA.

Cantarero G, **Musacchia G**, Kraus N. Brainstem responses and their relationship to cortical encoding of speech sounds. Undergraduate Honors Symposium, June 2006. Northwestern University, Evanston, USA

Musacchia G, Sams M, Kraus N. Multi-sensory interaction and plasticity in the human auditory brainstem. International Multisensory Research Forum, July 2007. Sydney, Australia.

Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced auditory and audiovisual brainstem responses to speech and music. Music and Language as Cognitive Systems, May 2007. Cambridge, United Kingdom.

OTHER SCIENTIFIC WRITING

4\00 Senior Thesis, “Essential Elements of an Auditory Gestalt”, U.C.S.C., Adviser: Dominic Massaro, Ph.D.

REFERENCES

Nina Kraus, Ph.D, Northwestern University
Email n-kraus@northwestern.edu
Mikko Sams, Ph.D., Helsinki University of Technology
Email Mikko.Sams@hut.fi
Scott Lipscomb, Ph.D., University of Minnesota
Email lipscomb@umn.edu
Leonard Trejo, Ph.D., N.A.S.A. Ames Research Center
Email ltrejo@mail.nasa.arc.gov