

# Experience-induced Malleability in Neural Encoding of *Pitch*, *Timbre*, and *Timing*

## Implications for Language and Music

Nina Kraus,<sup>a,b</sup> Erika Skoe,<sup>a</sup> Alexandra Parbery-Clark,<sup>a</sup> and Richard Ashley<sup>c</sup>

<sup>a</sup>*Auditory Neuroscience Lab, Department of Communication Sciences, <sup>b</sup>Departments of Neurobiology and Physiology, Department of Otolaryngology, and <sup>c</sup>Bienen School of Music, Northwestern University, Evanston, Illinois, USA*

Speech and music are highly complex signals that have many shared acoustic features. *Pitch*, *Timbre*, and *Timing* can be used as overarching perceptual categories for describing these shared properties. The acoustic cues contributing to these percepts also have distinct subcortical representations which can be selectively enhanced or degraded in different populations. Musically trained subjects are found to have enhanced subcortical representations of *pitch*, *timbre*, and *timing*. The effects of musical experience on subcortical auditory processing are pervasive and extend beyond music to the domains of language and emotion. The sensory malleability of the neural encoding of *pitch*, *timbre*, and *timing* can be affected by lifelong experience and short-term training. This conceptual framework and supporting data can be applied to consider sensory learning of speech and music through a hearing aid or cochlear implant.

**Key words:** brain stem; subcortical; musical training; cochlear implant

### Introduction

From the cochlea to the auditory cortex, sound is encoded at multiple locations along the ascending auditory pathway, eventually leading to conscious perception. While there is no doubt that the cortex plays a major role in the perception of speech, music, and other meaningful auditory signals, recent studies suggest that subcortical encoding of sound is not merely a series of passive, bottom-up processes successively transforming the acoustic signal into a more complex neural code. Rather, subcortical sensory processes dynamically interact with cortical processes, such as memory, attention,

and multisensory integration, to shape the perceptual system's response to speech and music.

In the last two decades there has been a surge in research devoted to how musical experience affects brain structure, cortical activity, and auditory perception. These three lines of research have uncovered several interesting byproducts of musical training. Musicians have brain structural differences not only in the motor cortices—the parts of the brain controlling hand/finger movement and coordination—but also in the auditory cortices.<sup>1,2</sup> In addition to structural differences, musicians show different patterns of neural activation. For example, musicians show stronger responses to simple, artificial tones and heightened responses to the sound of their own instrument compared to other instruments.<sup>3–7</sup> Interestingly, such cortical differences can be seen as early as 1 year after the onset of musical training<sup>8</sup> and extend to speech signals.<sup>9,10</sup> Recently, this line of research has moved to subcortical

Address for correspondence: Nina Kraus, Ph.D., Auditory Neuroscience Lab, Department of Communication Sciences Departments of Neurobiology and Physiology, Department of Otolaryngology, Northwestern University, 2240 Campus Drive, Evanston IL, 60208. Voice: +847-491-3181; fax: +847-491-2523. <http://brainvolts.northwestern.edu/nkraus@northwestern.edu>

levels. This work, along with supporting data, will be presented here within the *pitch*, *timbre*, and *timing* conceptual framework. In the final section of this review, we will switch the focus to cochlear implants and apply this conceptual framework to consider sensory learning of speech and music through an implant.

### Conceptual Framework for Studying Subcortical Responses: *Pitch, Timbre, and Timing*

Work from our laboratory<sup>d</sup> points to *pitch*, *timbre*, and *timing* as having distinct subcortical representations which can be selectively enhanced or degraded in different populations.

*Pitch*, as defined by the Standard Acoustical Terminology of the Acoustical Society of America, is “that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high” S12.01, P.34.<sup>11</sup> For pure tones, the frequency, or cycles per second of the waveform, is the physical correlate of *pitch*; however when considering more complex sounds, *pitch* corresponds, in part, to the lowest resonant frequency, also known as the fundamental frequency ( $F_0$ ).<sup>e</sup> For speech,  $F_0$  is dictated by the rate of vocal fold vibration and for music it depends on the instrument. For example, the reed is the source of  $F_0$  vibration for the oboe and clarinet, whereas the string is the source for the violin and guitar. For the purposes of this review, we use the word *pitch* as shorthand for referring to the information carried by the  $F_0$ , and so in this context, *pitch* and  $F_0$  are synonymous.

*Timbre*, also referred to as “sound color,” enables us to differentiate two sounds with the same *pitch*. *Timbre* is a multidimensional prop-

erty resulting from the interaction of spectral and temporal changes associated with the harmonics of the fundamental along with the *timing* cues of the attack (onset) and decay (offset). Together this gives rise to the characteristic sound quality associated with a given instrument or voice. *Timbre* is also an important cue for distinguishing contrastive speech sounds (i.e., phonemes). As the vocal tract is shaped by the movement of the articulators during speech production, the resonance structure of the vocal tract changes and certain harmonics are attenuated while others are amplified. These amplified harmonics are known as speech-formants and they are important for distinguishing phonemes. Our focus here is on the harmonic aspects of *timbre* and the corresponding subcortical representation.

*Timing* refers to the major acoustic landmarks in the temporal envelope of speech and music signals. For speech, *timing* arises from the alternating opening and closing of the articulators and from the interplay between laryngeal and supralaryngeal gestures. *Timing* also includes spectrotemporal features of speech, such as time-varying formants. As such, *timing* arises from the interplay between the actions of the source (glottal pulse train) and filter (articulators). For music, *timing* can be considered in conjunction with the temporal information contributing to *timbre* perception. Likewise, on a more global scale, it refers to the duration of sounds and their subsequent perceptual groupings into rhythm. For the purposes of this review, we will focus on the neural representation of transient temporal features, such as onsets and offsets occurring as fast as fractions of milliseconds.

### The Auditory brain stem Response

The auditory brain stem, an ensemble of nuclei belonging to the efferent and afferent auditory systems, receives and processes the output of the cochlea en route to higher centers of auditory processing. The auditory brain stem response (ABR), a highly replicable far-field

<sup>d</sup>For more information about our laboratory and the work reviewed herein, please visit our website: <http://www.brainvolts.northwestern.edu/>

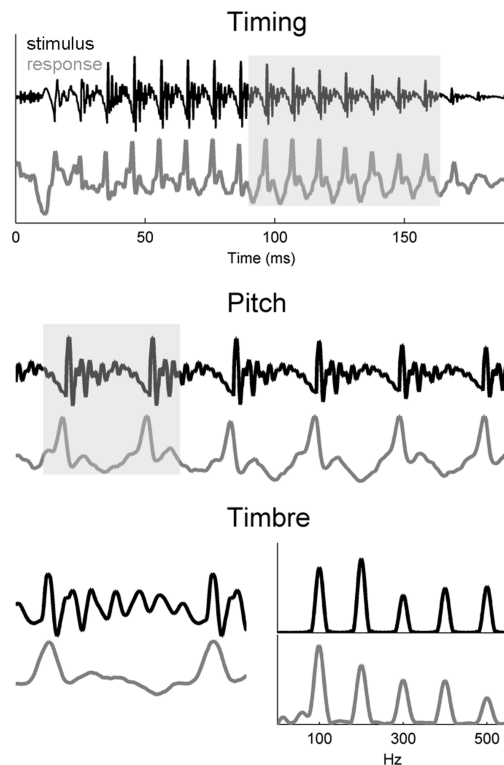
<sup>e</sup>It should be noted that  $F_0$  is one of several elements contributing to the perception of *pitch*. There is also the phenomenon of the *missing fundamental* in which the perceived sound is not present in the acoustic spectrum, but results from interaction of the harmonics.

potential recorded from surface electrodes placed on the scalp, reflects the acoustic properties of the sound stimulus with remarkable fidelity. In fact, when the electrical response is converted into an audio signal, the audio signal maintains a striking similarity to the eliciting stimulus.<sup>12</sup> Because of the transparency of this subcortical response, it is possible to compare the response *timing* and frequency composition to the corresponding features of the stimulus (Fig. 1). *Timing* features (including sound onsets, offsets, and format transitions) are represented in the brain stem response as large transient peaks, whereas *pitch* ( $F_0$ ) and *timbre* (harmonics up to about 1000 Hz) information is represented as interspike intervals that match the periodicity of the signal, a phenomenon known as phase locking.<sup>f</sup> By means of commonly employed digital signal processing tools, such as autocorrelation<sup>g</sup> and Fourier analysis,<sup>h</sup> features relating to stimulus *pitch* and *timbre* can be extracted from the response. As a consequence of being such a highly replicable measure, incredibly subtle differences in the *timing* and phase locking of the ABR are indicative of sensory processing malleability and abnormality.

### Subcortical Representation of Pitch

Musicians have extensive experience manipulating *pitch* within the context of music. Work by the Kraus Laboratory<sup>9,13,14</sup> shows that life-long musical training is associated with heightened subcortical representations of both *musical* and *linguistic pitch*, suggesting transfer effects from music to speech processing.

Musacchia *et al.*<sup>14</sup> employed an audiovisual (AV) paradigm to tap into the multisensory nature of music. Given that music performance



**Figure 1.** Schematic representation of *timing*, *pitch*, and *timbre* in the stimulus (black) and brain stem response (gray) waveforms. Top: The full view of the time-domain stimulus waveform “da.” The temporal features of the stimulus, including the sound offset and onset, are preserved in the response. The gray box demarcates six cycles of the fundamental frequency ( $F_0$ ); a blowup of this section is plotted in the middle panel. Middle: Major waveform peaks occur at an interval of 10 ms (i.e., the periodicity of a 100-Hz signal). This stimulus periodicity, which elicits the perception of *pitch*, is faithfully represented in the response. Bottom: The left panel shows a close-up of an  $F_0$  cycle. The harmonics of the stimulus are represented as small-amplitude fluctuations between the major  $F_0$  peaks in the stimulus and response. In the right panel, the stimulus and response are plotted in the spectral domain. Frequencies important for the perception of *pitch* (100 Hz) and *timbre* (frequencies at multiples of 100 Hz) are maintained in the brain stem response.

<sup>f</sup>The phase locking measured by the ABR likely reflects the activity from the lateral lemniscus and inferior colliculus.<sup>76–80</sup>

<sup>g</sup>Autocorrelation can be used to detect repeating patterns within a signal, such as the fundamental periodicity.

<sup>h</sup>Fourier analysis is method for decomposing complex signals into component sine waves. Fourier analysis of brain stem responses to speech and music shows concentrations of energy at frequencies important for *pitch* and *timbre* perception.

involves the integration of auditory, visual, and tactile information, we hypothesized that life-long musical practice would influence AV integration. Subcortical responses were compared in three conditions: AV, auditory alone (A), and

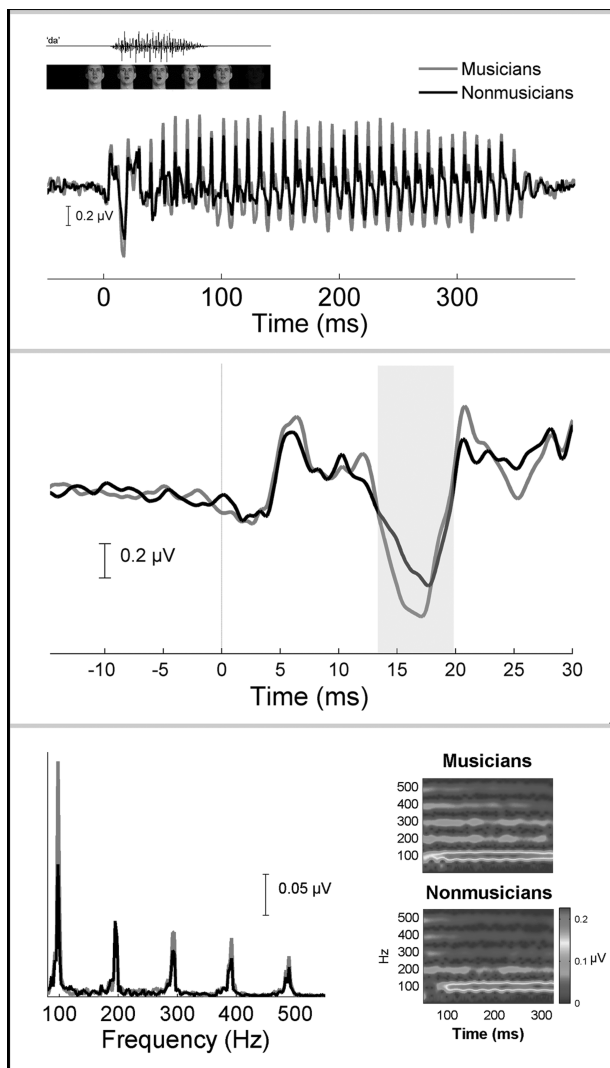
visual alone (V). In the AV condition, subjects watched and listened to a movie of a person playing the cello or saying “da.” In the A condition, no movie was displayed, and in the V condition, no sounds were presented. For both musicians and nonmusicians, the *pitch* responses to both speech and music were larger in the multimodal condition (AV) compared to unimodal A condition. However, musicians showed comparatively larger *pitch* response in both A and AV conditions (AV responses are plotted in Fig. 2), and more pronounced multimodal effects, that is, greater amplitude increase between A and AV conditions. In addition, *pitch* representation strongly correlated with years of musical practice, such that the longer a person had been playing, the larger the *pitch* response (Fig. 3, top). When the cortical responses to the AV condition were examined, this *pitch* representation was positively correlated with the steepness of the P1–N1 slope, such that the sharper (i.e., more synchronous) the cortical response, the larger the *pitch* representation.<sup>9</sup> Other aspects of these multisensory responses will be explored in the sections relating to subcortical representation of *timbre* and *timing*. Taken together these data indicate that multisensory training, such as is acquired with musical experience, has pervasive effects on subcortical and cortical sensory encoding mechanisms for both musical and speech stimuli and leads to training-induced malleability of sensory processing.

In music and language, *pitch* changes convey melodic and semantic or pragmatic information. Recently, a number of studies have looked at the representation of linguistic *pitch* contours (i.e., sounds which change in *pitch* over time) in the brain stem response. In Mandarin Chinese, unlike English, *pitch* changes signal lexical semantic changes. Compared to native English speakers, Mandarin Chinese speakers have stronger and more precise brain stem phase locking to Mandarin *pitch* contours, suggesting that the subcortical representation of *pitch* can be influenced by linguistic experience.<sup>15,16</sup> Using a similar paradigm, we ex-

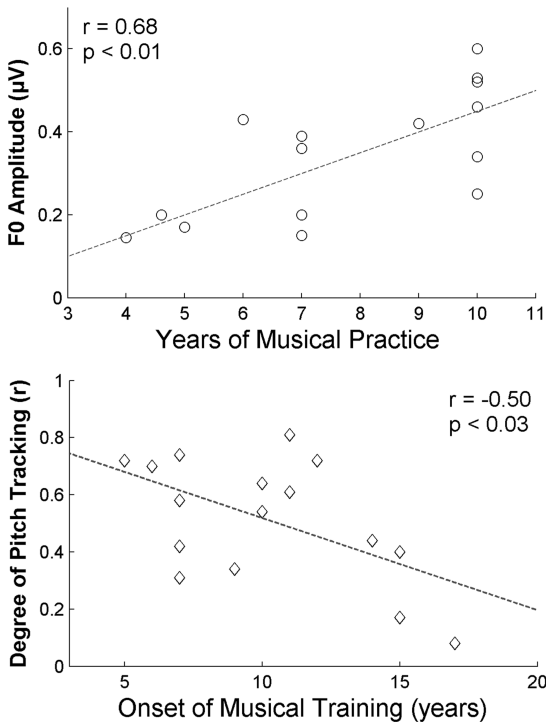
plored the idea that musical *pitch* experience can lead to enhanced linguistic *pitch* tracking.<sup>13</sup> ABRs were recorded to three Mandarin tone contours: tone 1 (level contour), tone 2 (rising contour), and tone 3 (dipping contour). Musically trained native English speakers, with no knowledge of Mandarin or other tone languages, were found to have more accurate tracking of tone 3 (Fig. 4), a complex contour not occurring at the lexical (word) level in English.<sup>17</sup> In addition, we found that the accuracy of *pitch* tracking was correlated with two factors: years of musical training and the age that musical training began (Fig. 3, bottom). The differences between musicians and nonmusicians were less pronounced for tone 2 and not evident for tone 1. In contrast to tone 3, which only occurs at the phrase level in English, tones 1 and 2 are found at the word and syllable level. Taken together with the finding that musicians exhibit distinctive responses to emotionally salient *pitch* cues<sup>18</sup> and enhanced *pitch* elements in musical chords<sup>23</sup> (reviewed below), we concluded that musical training alters subcortical sensory encoding of dynamic *pitch* contours, especially for complex and novel stimuli.

The studies reviewed above investigated the effects of lifelong auditory (linguistic and musical) experience on the subcortical representation of *pitch*. Recent work from Song *et al.*<sup>19</sup> suggests that lifelong experience may not be necessary for engendering changes in the subcortical representation of *pitch*. In fact, we found that as few as eight training sessions (30 mins each) can produce more accurate and more robust subcortical *pitch* tracking in native-English-speaking adults. Interestingly, improvement occurred only for the most complex and least familiar *pitch* contour (tone 3).

Unlike musicians who have heightened *pitch* perception,<sup>20,21</sup> some individuals with autism spectrum disorders (ASD) are known to have issues with *pitch* perception in the context of language. For example, these individuals often cannot take advantage of the prosodic aspects of language and have difficulty distinguishing a question (rising *pitch*) from a statement (level or



**Figure 2.** Grand average brain stem responses to the speech syllable “da” for both musician (red) and non-musician (black) groups in the audiovisual condition. Top: Amplitude differences between the groups are evident over the entire response waveform. These differences translate into enhanced *pitch* and *timbre* representation (see bottom panel). Auditory and visual components of the speech stimulus (man saying “da”) are plotted on top. Middle: Musicians exhibit faster (i.e., earlier) onset responses. The grand average brain stem responses in the top panel have been magnified here to highlight the onset response. The large response negativity (shaded region) occurs on average  $\sim 0.50$  ms earlier for musicians compared to nonmusicians. Bottom. Fourier analysis shows musicians to have more robust amplitudes of the  $F_0$  peak (100 Hz) and the peaks corresponding to the harmonics (200, 300, 400, 500 Hz) (left). To illustrate frequency tracking of *pitch* and *harmonics* over time, narrow-band spectrograms (right) were calculated to produce time–frequency plots (1-ms resolution) for the musician (right top) and non-musician groups (right bottom). Spectral amplitudes are plotted along a color continuum, with warmer colors corresponding to larger amplitudes and cooler colors representing smaller amplitudes. Musicians have more pronounced harmonic tracking over time. This is reflected in repeating parallel bands of color occurring at 100 Hz intervals. In contrast, the spectrogram for the nonmusician group is more diffuse, and the harmonics appear more faded (i.e., weaker) relative to the musician group. (Adapted from Musacchia *et al.*<sup>9,14</sup>) (In color in *Annals* online.)



**Figure 3.** Neural enhancement varies according to the extent (top) and onset (bottom) of musical practice. Top: The number of years (over the last 10 years) of consistent practice is correlated with the strength of subcortical *pitch* encoding. Thus, the longer an individual has been practicing music, the larger the  $F_0$  amplitude. (Adapted from Musacchia *et al.*<sup>14</sup>) Bottom: The precision of brain stem *pitch* tracking is associated with the age that musical training began. Subjects who started earlier show a higher degree of *pitch* tracking. [N.B.: “Perfect” *pitch* tracking (i.e., no deviation between the stimulus *pitch* trajectory and response *pitch* trajectory) would be plotted as a 1 along the y-axis.] (Adapted from Wong *et al.*<sup>13</sup>)

falling *pitch*). Russo *et al.*<sup>22</sup> explored whether this prosodic deficit was related to subcortical representation of *pitch*. We found that a subset of autistic children showed poor *pitch* tracking to syllables with linearly rising and falling *pitch* contours. Given that the subcortical representation of *pitch* can be enhanced with short-term linguistic *pitch* training and life-long musical experience, this suggests that some children with ASD might benefit from an auditory training paradigm that integrates musical

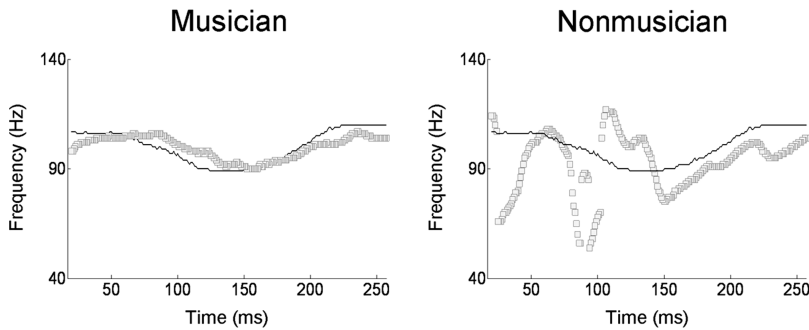
and linguistic training as a means of improving brain stem *pitch* tracking.

### Subcortical Representation of *Timbre*

A growing body of research is showing that musicians represent the harmonics of the stimulus more robustly than their nonmusician counterparts.<sup>9,18,23</sup> This is evident for a whole host of stimuli including speech and emotionally affective sounds as well as musical sounds. Lee *et al.*<sup>23</sup> recorded brain stem responses to harmonically rich musical intervals and found that musicians had heightened responses to the harmonics, as well as the combination tones<sup>i</sup> produced by the interaction of the two notes of the interval. In music, the melody is typically carried by the upper voice and the ability to parse out the melody from other voices is a fundamental musical skill. Consistent with previous behavioral and cortical studies,<sup>24–27</sup> we found that musicians demonstrated larger subcortical responses to the harmonics of the upper note relative to the lower note. In addition, an acoustic correlate of consonance perception (i.e., temporal envelope) was more precisely represented in the musician group. When two tones are played simultaneously, the two notes interact to create periodic amplitude modulations. These modulations generate the perception of “beats,” “smoothness,” and “roughness,” and contribute to the sensory consonance of the interval. Thus by actively attending to the upper note of a melody and the harmonic relation of concurrent tones, musicians may develop specialized sensory systems for processing behaviorally relevant aspects of musical signals. These specializations likely occur throughout the course of musical training—a viewpoint supported by a correlation between the length of musical training (years) and the extent of subcortical enhancements.

The link between behavior and subcortical enhancements is also directly supported by

<sup>i</sup>Combination tones are distortion products that result from the non-linear nature of the auditory system.



**Figure 4.** Pitch tracking plots from a musician (left) and nonmusician (right). The thin black line represents the *pitch* contour of the stimulus (Mandarin tone 3), and the thick gray line represents the extracted *pitch* trajectory of the brain stem response. The musician's brain response follows the *pitch* of the stimulus more precisely, a phenomenon known as pitch tracking. (Adapted from Wong *et al.*<sup>13</sup>)

Musacchia *et al.*,<sup>9</sup> who found that better performance on a *timbre* discrimination task was associated with larger subcortical representations of *timbre*. *Timbre* was also an important distinguishing factor for separating out musicians from nonmusicians. As a group, the musically trained subjects had heightened representation of the harmonics (Fig. 2, bottom). Furthermore, when the subjects were analyzed along a continuum according to the age musical training began, subjects who started at a younger age were found to have larger *timbre* representations compared to those who began later in life. In addition, a correlation was found between cortical response *timing* and subcortical *timbre* encoding, which may be indicative of cortical structures being active in the processing of more subtle stimulus features.

### Subcortical Representation of Timing

*Timing* measures provide insight into the accuracy with which the brain stem nuclei synchronously respond to acoustic stimuli. The hallmark of normal perception is an accurate representation of the temporal features of sound. In fact, disruptions on the order of fractions of milliseconds are clinically significant for the diagnosis of hearing loss, brain stem pathology, and certain learning disorders. Compared to normally hearing nonmusicians, musicians have more precise subcortical rep-

resentation of *timing*, resulting in earlier (i.e., faster) and larger onset peaks<sup>14,18</sup> (Fig. 2, middle). Furthermore, the results of these studies suggest an intricate relationship between years of musical practice and neural representation of *timing*. Taken together, the outcomes of our correlational analyses show that subcortical sensory malleability is dynamic and continues beyond the first few years of musical training.

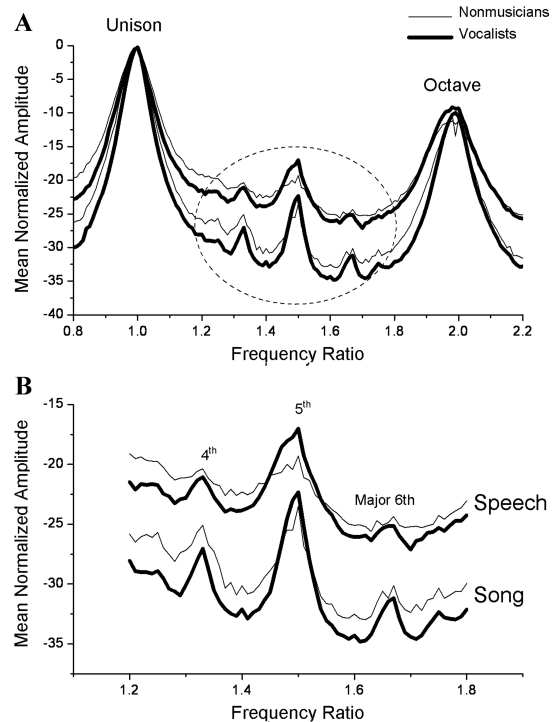
## Summary: Music Experience and Neural Plasticity

### Transfer Effects

By binding together multimodal information and actively engaging cognitive and attentional mechanisms, music is an effective vehicle for auditory training.<sup>29,30</sup> By showing that the effects of musical experience on the nervous system's response to sound are pervasive and extend beyond music,<sup>9,13,14,18,31</sup> work from our laboratory fits within the larger scientific body of evidence. We find transfer effects between the musical domain and the speech domain resulting in enhanced subcortical representation of linguistic stimuli.<sup>9,13,14</sup> However, these enhancements are not only specific to musical and linguistic stimuli, but also occur with non-linguistic emotionally rich stimuli as well. Strait *et al.*<sup>18</sup> (also appearing in this

volume<sup>31a</sup>) recorded ABRs to the sound of a baby's cry, an emotionally laden sound. Compared to the nonmusician cohort, musicians showed enhanced *pitch* and *timbre* amplitudes to the most spectrally complex section of the sound, and attenuated responses to the more periodic, less complex section. These results provide the first biological evidence for enhanced perception of emotion in musicians<sup>32,33</sup> and indicate the involvement of subcortical mechanisms in processing of vocally expressed emotion. Another compelling finding is that extensive auditory training can lead to both enhancement and efficiency (i.e., smaller amplitudes are indicative of allocation of fewer neural resources) of subcortical processing, with both enhancement and economy being evident in the subcortical response to a single acoustic stimulus. This finding reinforces the idea that subcortical responses to behaviorally relevant signals are not hardwired, but are malleable with auditory training.

The multisensory nature of music may also have an impact on vocal production by engaging auditory/vocal-motor mechanisms. Stegemöller and colleagues<sup>31</sup> recorded speech and song samples from musicians and non-musicians. Vocal productions were analyzed using a statistical analysis of frequency ratios.<sup>34</sup> The vocal productions (speech and music) of both groups showed energy concentrations at ratios corresponding to the 12-tone musical scale. However, musicians' samples were smoother and had fewer deviant (i.e., non 12-tone ratio) peaks (Fig. 5), showing that musicians had less harmonic jitter in their voices. This pattern was apparent even in the speech condition, where nonmusicians were found to differ from the vocally trained subjects in the musician group. This suggests that musical vocal training has an impact on vocal tract resonance during speech production. Also notable is that the musicians who did not undergo vocal training (instrumentalists) had smoother spectra for the song samples. Therefore, exposure to the 12-tone scale through instrumental training can be seen to influence vocal produc-



**Figure 5.** Normalized spectra of speech (top two traces) and song (bottom two traces) tokens for non-musicians and vocalists. Prominent peaks in the spectra correspond to the intervals of the 12-tone scale. Unison, Perfect 4th, Perfect 5th, Major 6th, and Octave are labeled and represent the most well-defined spectral peaks in the speech and song tokens. Compared to nonmusicians, vocalists and professional musicians (not plotted) have smoother normalized spectra which include fewer unexpected (non-12-tone interval) peaks. The encircled portion of (A) is magnified in (B) to show the decrease in the number of unexpected peaks from speech to song, and from no musical experience to trained vocal experience. (Adapted from Stegemöller *et al.*<sup>31</sup>)

tion, indicating a transfer from the auditory to the motor modalities.

### Subcortical Enhancements and the Interaction of Top-down Processes

At first blush, it would appear that musical training is akin to a volume knob, leading to musicians' processing sounds as if they were presented at a louder decibel level. While it is clear that musicians show subcortical enhancements for *pitch*, *timbre*, and *timing*, a



simple stimulus-independent gain effect cannot explain all of the results reviewed above. A better analogy is that musical training helps to focus auditory processing, much in the same way that glasses help to focus vision, and that this leads to clearer and more fine-grained subcortical representations. If only a gain effect was operative, we might expect all stimuli and all stimulus features to show more or less equivalent enhancements. However, available data do not support this stimulus-independent view. What we find instead is that only certain stimuli<sup>13</sup> or certain aspects of the stimuli are enhanced in musicians.<sup>14,18,23</sup> So while musical training might help focus auditory processing at a subcortical level, it does not do so blindly. Instead the behavioral relevance and complexity of the stimulus likely influences how the sensory system responds. This suggests that higher-level cognitive factors are at play. In order to obtain auditory acuity, musicians actively-engage top-down mechanisms, such as attention, memory, and context, and it is this binding of sensory acuity and cognitive demands that may in fact drive the subcortical enhancements we observe in musicians. Our findings suggest that higher-order processing levels (i.e., cortical) have efficient feedback pathways to lower-order (i.e., brain stem) processing levels. This top-down feedback is likely mediated by the corticofugal pathway, a vast track of efferent fibers that link together the cortex and lower structures.<sup>35–38</sup> While the corticofugal system has been extensively studied in animal models, the direct involvement of this efferent system in human auditory processing has also been demonstrated by Perrot and colleagues.<sup>39</sup> In the animal model, the corticofugal system works to fine-tune subcortical auditory processing of behaviorally relevant sounds by linking learned representations and the neural encoding of the physical acoustic features. This can lead to short-term plasticity and eventually long-term reorganization of subcortical sound encoding (for a review see Suga *et al.*<sup>35</sup>). Importantly, corticofugal modulation of specific auditory information is evident in the earliest stages

of auditory processing.<sup>6</sup> It is therefore our view that corticofugal mechanisms apply to human sensory processing, and can account, at least in part, for the pattern of results observed in musicians. Consistent with this corticofugal hypothesis and observations of experience-dependent sharpening of primary auditory cortex receptive fields,<sup>7,40</sup> we maintain that subcortical enhancements do not result simply from passive, repeated exposure to musical signals or pure genetic determinants. Instead, the refinement of auditory sensory encoding is driven by a *combination* of these factors and behaviorally relevant experiences, such as lifelong music making. This idea is reinforced by correlational analyses showing that subcortical enhancements vary as a function of musical experience<sup>9,13,14,18,23</sup> (Fig. 3).

### When Auditory Processing Goes Awry

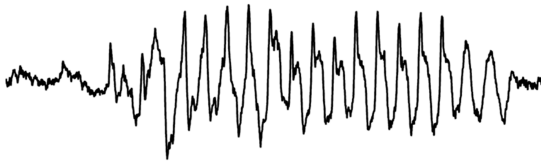
Impaired auditory processing is the hallmark of several clinical conditions, such as auditory-processing disorder (APD), a condition characterized by difficulty perceiving speech in noisy environments. Work from our laboratory has shown that a significant subset of children with language-based learning problems, such as dyslexia, where APD is common, show irregular subcortical representations of *timing* and *timbre* (harmonics), but not *pitch*.<sup>28,41</sup> This pattern is consistent with the phonological processing problems inherent in reading disorders. Our research into the subcortical representation of speech in the learning-impaired population has been translated into a clinical tool, BioMARK (Biological Marker of Auditory Processing; see Clinical Technologies at <http://www.brainvolts.northwestern.edu/>). This test provides a standardized metric of auditory encoding and can be used to disentangle roles of *pitch*, *timbre*, and *timing* in normal and disordered auditory processing.

For a significant number of children with reading disabilities, sound is atypically encoded

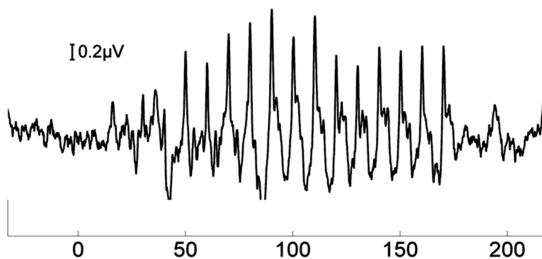
### Language Impaired Child



### Normal Adult



### Professional Musician



**Figure 6.** Brain stem responses from a child with reading difficulties (top), a young adult with typical hearing (middle) and a professional musician (bottom). Note the differences in waveform morphology, with the musician having larger and more defined (sharper) peaks.

at multiple levels of the auditory system—the auditory brain stem,<sup>28,41–44</sup> the auditory cortex<sup>45–47</sup> or both<sup>48–50</sup>—suggesting a complex interaction between subcortical and cortical levels. Thus, the deficits we find in language impairment, such as developmental dyslexia<sup>28,48</sup> (Fig. 6) and ASD,<sup>22</sup> might be the consequence of faulty or suboptimal corticofugal engagement of auditory activity.

Further evidence for the dynamic nature of subcortical auditory processing can be found by studying the effects of short-term training in children. After undergoing an 8-week commercially available auditory training program, children with language-based learning impair-

ments showed improved subcortical response *timing* for speech signals presented in background noise.<sup>51</sup> Because the auditory training was not specific to speech perception in noise, it raises the possibility that training-induced brain stem plasticity was mediated by top-down, cortically driven processes, a conclusion also supported by work from de Boer and Thornton.<sup>52</sup>

### Cochlear Implants and Music Perception

Cochlear implants (CIs) have proven to be enormously successful in engendering speech perception, especially in quiet settings, yet music perception is still below par. This is perhaps not surprising given that CI processing strategies are primarily designed to promote speech perception and thereby provide only a rough estimation of spectral shape, despite comparably fine-grained temporal resolution. While both speech and music have spectral and temporal elements, the weighting of these elements is not the same: speech perception requires more temporal precision whereas music perception requires more spectral precision.<sup>53</sup> The CI user's poor performance on musical tasks can be explained in large part by this underlying CI processing scheme and the acoustic differences between speech and music.

Real-world music listening requires the integration of multiple cues including *pitch*, *timing* (e.g., tempo and rhythm), and *timbre* (e.g., instrument identification). For research purposes, music can be analytically decomposed into perceptual tasks that tap into each individual element. The *pitch*, *timbre*, and *timing* model that we employ in our laboratory for studying brain stem responses is also a useful trichotomy for assessing CI performance on musical tasks. With respect to *timing* tasks, the general consensus in the CI literature is that CI users and normal-hearing listeners have nearly comparable performances, yet the CI users perform far below average on *timbre* and *pitch* tasks.<sup>54–60</sup> On

*timbre* tasks, CI wearers often have a difficult time telling two instruments apart.<sup>54–56,58,60</sup> However, despite this well-documented performance, Koelsch and colleagues<sup>61</sup> have demonstrated that *timbral* differences can elicit subliminal cortical responses. This suggests that even though many CI users cannot formally acknowledge differences in sound quality, these differences may in fact be registered in the brain.

When it comes to *pitch* perception, CI users could be described as having an extreme form of amusia (tone deafness). For example, whereas normally hearing adults can easily tell the difference between two adjacent keys on a piano (i.e., 1 semitone difference), for the average postlingually implanted CI wearer, the notes must be at least 7 keys apart.<sup>54</sup> However, even if implantation occurs later in life, recent work by Guiraud and colleagues,<sup>62</sup> indicates that CIs can help reverse the effects of sensory deprivation by reorganizing how spectral information is mapped in the cortex.

For CI users, rehabilitative therapy has traditionally focused on improving speech perception and production. Despite numerous anecdotal and case reports showing that music therapy is being integrated into the rehabilitative process, the effects of musical training after CI implantation have garnered little scientific attention. Nevertheless, two known published reports reinforce the idea that focused short-term training can improve *timbre* and *pitch* perception.<sup>54,63</sup>

While vocoded sounds—sounds that have been manipulated to simulate the input that CI users receive—cannot fully mimic the CI acoustic experience, they serve as a useful surrogate for studying how the nervous system deals with degraded sensory input before and after training. Studies are currently under way in our laboratory to explore how the normal hearing system encodes *pitch*, *timbre*, and *timing* features of speech and musical stimuli, and their vocoded counterparts. Special attention will be paid to the relationship between musical experience and how vocoded and more natural

conditions are differentially represented at subcortical and cortical levels.

Because of magnetic and electromagnetic interference from the CI transmitter, magnetoencephalography and magnetic resonance imaging cannot be performed while a person is wearing a CI. Although an electrical artifact can plague electrophysiological recordings from CI wearers, techniques have been developed to minimize these effects in cortical potentials.<sup>64,65</sup> ABRs to speech and music have the capacity to be a highly objective and revealing measure of auditory processing in normal subjects listening to vocoded sounds, and with technological advances speech- and music-evoked ABRs may eventually be recorded in CI users. This work would complement the existing literature that has documented the integrity and plasticity of the CI user's subcortical auditory pathways using simple click stimuli.<sup>66,67</sup>

Furthermore, in order to promote large scale and cross-laboratory/cross-clinic comparisons there is a need for standardized measurements of electrophysiology (equivalent to BioMARK) and music perception in this population (for three examples of music tests, see Nimmons *et al.*,<sup>68</sup> Cooper *et al.*,<sup>69</sup> and Spitzer *et al.*<sup>70</sup>). The benchmark of an effective test is one that can track changes before and after training, and is also sensitive enough to keep up with advancing CI technologies.

Speech and music perception are without question constrained by the current state of CI technology. However, technology alone cannot explain the highly variable performance across implantees, including the exceptional cases of children and adults who demonstrate near-normal *pitch* perception and production.<sup>71,72</sup> These “super-listeners” serve as beacons for where commonplace CI performance can aspire in the near future.

While most CI wearers have limited musical experience before implantation,<sup>73</sup> a growing number of trained musicians are receiving implants. These individuals seem to have an advantage when it comes to music perception through a CI, especially for *pitch* perception.

This underscores the important role that music experience plays in shaping sensory skills and lends further support for experience-dependent corticofugal (top-down) modulation of cortical and subcortical auditory pathway.<sup>13,35,39,74</sup> Through the use electrophysiology and standardized music tests, we will gain better insight into the biological processes underlying super-listeners and ordinary listeners, which will ultimately lead to more refined CI technology and improved music enjoyment among CI users.

## Conclusion and Future Outlook

Subcortical auditory processes are dynamic and not hardwired. As discussed here, auditory sensory processing interacts with other modalities (e.g., visual and motor influences) and is influenced by language and music experience. The role of subcortical auditory processes in perception and cognition is far from understood, but available data suggest a rich interplay between the sensory and cognitive processes involved in language and music, and a common subcortical pathway for these functions. It appears that in the normal system, music and language experience fundamentally shape auditory processing that occurs early in the sensory processing stream.<sup>13–16,18,19,23</sup> This top-down influence is likely mediated by the extensive corticofugal circuitry of descending efferent fibers that course from the cortex to the cochlea.<sup>75</sup> In order to facilitate sensory learning, the impaired system can capitalize on the shared biological resources underlying the neural processing of language and music, the impact music has on auditory processing and multisensory integration, and the apparent cognitive-sensory reciprocity.

## Acknowledgments

This work is supported the National Institutes of Health (Grant R01DC001510), the National Science Foundation (NSF 0544846), and the Hugh Knowles Center for Clinical and

Basic Science in Hearing and Its Disorders at Northwestern University, Evanston Illinois.

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

1. Pantev, C. *et al.* 1998. Increased auditory cortical representation in musicians. *Nature* **392**: 811–814.
2. Gaser, C. & G. Schlaug. 2003. Brain structure differs between musicians and non-musicians. *J. Neurosci.* **23**: 9240–9245.
3. Peretz, I. & R.J. Zatorre. 2005. Brain organization for music processing. *Annu. Rev. Psychol.* **56**: 89–114.
4. Pantev, C. *et al.* 2001. Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport* **12**: 169–174.
5. Margulis, E.H. *et al.* 2009. Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Hum. Brain Mapp.* **30**: 267–275. First published December 10th, 2007, DOI: 10.1002/hbm.20503.
6. Luo, F. *et al.* 2008. Corticofugal modulation of initial sound processing in the brain. *J. Neurosci.* **28**: 11615–11621.
7. Schreiner, C.E. & J.A. Winer. 2007. Auditory cortex mapmaking: principles, projections, and plasticity. *Neuron* **56**: 356–365.
8. Fujioka, T. *et al.* 2006. One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain* **129**: 2593–2608.
9. Musacchia, G., D. Strait & N. Kraus. 2008. Relationships between behavior, brain stem and cortical encoding of seen and heard speech in musicians. *Hearing Res.* **241**: 34–42.
10. Chandrasekaran, B., A. Krishnan & J.T. Gandour. 2009. Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain Lang.* **108**(1): 1–9.
11. American National Standards Institute. 1994. *American National Standard Acoustical Terminology*. Acoustical Society of America. New York, NY. *S12.01*. P. 34.
12. Galbraith, G.C. *et al.* 1995. Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* **6**: 2363–2367.
13. Wong, P.C. *et al.* 2007. Musical experience shapes human brain stem encoding of linguistic pitch patterns. *Nat. Neurosci.* **10**: 420–422.
14. Musacchia, G. *et al.* 2007. Musicians have enhanced subcortical auditory and audiovisual

- processing of speech and music. *Proc. Natl. Acad. Sci. USA* **104**: 15894–15898.
15. Krishnan, A. *et al.* 2005. Encoding of pitch in the human brain stem is sensitive to language experience. *Brain Res. Cogn. Brain Res.* **25**: 161–168.
  16. Xu, Y., A. Krishnan & J.T. Gandour. 2006. Specificity of experience-dependent pitch representation in the brain stem. *Neuroreport* **17**: 1601–1605.
  17. Pierrehumbert, J. 1979. The perception of fundamental frequency declination. *J. Acoust. Soc. Am.* **66**: 363–369.
  18. Strait, D. *et al.* 2009. Musical experience influences subcortical processing of emotionally-salient vocal sounds. *Eur. J. Neurosci.* **29**: 661–668.
  19. Song, J.H., E. Skoe, P.C. Wong & N. Kraus. 2008. Plasticity in the adult human auditory brain stem following short-term linguistic training. *J. Cogn. Neurosci.* **20**: 1892–1902.
  20. Kishon-Rabin, L. *et al.* 2001. Pitch discrimination: are professional musicians better than non-musicians? *J. Basic Clin. Physiol. Pharmacol.* **12**(2 Suppl): 125–143.
  21. Micheyl, C. *et al.* 2006. Influence of musical and psychoacoustical training on pitch discrimination. *Hear Res.* **219**: 36–47.
  22. Russo, N.M. *et al.* 2008. Deficient brain stem encoding of pitch in children with autism spectrum disorders. *Clin. Neurophysiol.* **119**: 1720–1731.
  23. Lee, K.M., *et al.* 2009. Selective subcortical enhancement of musical intervals in musicians. *J. Neurosci.* in press.
  24. Palmer, C. & S. Holleran. 1994. Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Percept. Psychophys.* **56**: 301–312.
  25. Crawley, E.J. *et al.* 2002. Change detection in multi-voice music: the role of musical structure, musical training, and task demands. *J. Exp. Psychol. Hum. Percept. Perform.* **28**: 367–378.
  26. Fujioka, T., L. Trainor & B. Ross. 2008. Simultaneous pitches are encoded separately in auditory cortex: an MMNm study. *Neuroreport* **19**: 361–368.
  27. Fujioka, T. *et al.* 2005. Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J. Cogn. Neurosci.* **17**: 1578–1592.
  28. Banai, K. *et al.* 2009. Reading and subcortical auditory function. *Cereb. Cortex* doi: 10.1093/cercor/bhp024.
  29. Zatorre, R.J. *et al.* 2002. Where is ‘where’ in the human auditory cortex? *Nat. Neurosci.* **5**: 905–909.
  30. Saunders, J. 1996. Real-time discrimination of broadcast speech/music. *IEEE Int. Conf. Acoust., Speech, Signal Process.: Proc. ICASSP* **2**: 993–996.
  31. Stegemöller, E.L. *et al.* 2008. Musical training and vocal production of speech and song. *Music Percept.* **25**: 419–428.
  - 31a. Strait, D. *et al.* 2009. Musical experience promotes subcortical efficiency in processing emotional vocal sounds. *Ann. N. Y. Acad. Sci. Neurosciences and Music III - Disorders and Plasticity.* **1169**: 209–213.
  32. Dmitrieva, E.S. *et al.* 2006. Ontogenetic features of the psychophysiological mechanisms of perception of the emotional component of speech in musically gifted children. *Neurosci. Behav. Physiol.* **36**: 53–62.
  33. Thompson, W.F., E.G. Schellenberg & G. Husain. 2004. Decoding speech prosody: do music lessons help? *Emotion* **4**: 46–64.
  34. Schwartz, D.A., C.Q. Howe & D. Purves. 2003. The statistical structure of human speech sounds predicts musical universals. *J. Neurosci.* **23**: 7160–7168.
  35. Suga, N. *et al.* 2002. Plasticity and corticofugal modulation for hearing in adult animals. *Neuron* **36**: 9–18.
  36. Winer, J.A. 2006. Decoding the auditory corticofugal systems. *Hear Res.* **212**: 1–8.
  37. Suga, N. 2008. Role of corticofugal feedback in hearing. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* **194**: 169–183.
  38. Suga, N. & X. Ma. 2003. Multiparametric corticofugal modulation and plasticity in the auditory system. *Nat. Rev. Neurosci.* **4**: 783–794.
  39. Perrot, X. *et al.* 2006. Evidence for corticofugal modulation of peripheral auditory activity in humans. *Cereb. Cortex* **16**: 941–948.
  40. Fritz, J.B., M. Elhilali & S.A. Shamma. 2007. Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J. Neurophysiol.* **98**: 2337–2346.
  41. Wible, B., T. Nicol & N. Kraus. 2004. Atypical brain stem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol. Psychol.* **67**: 299–317.
  42. Cunningham, J. *et al.* 2001. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin. Neurophysiol.* **112**: 758–767.
  43. Johnson, K.L. *et al.* 2007. Auditory brain stem correlates of perceptual timing deficits. *J. Cogn. Neurosci.* **19**: 376–385.
  44. King, C. *et al.* 2002. Deficits in auditory brain stem encoding of speech sounds in children with learning problems. *Neurosci. Lett.* **319**: 111–115.
  45. Leppanen, P.H. *et al.* 1999. Cortical responses of infants with and without a genetic risk for dyslexia: II. Group effects. *Neuroreport* **10**: 969–973.

46. Kraus, N. *et al.* 1996. Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science* **273**: 971–973.
47. Nagarajan, S. *et al.* 1999. Cortical auditory signal processing in poor readers. *Proc. Natl. Acad. Sci. USA* **96**: 6483–6488.
48. Banai, K. *et al.* 2005. Brain Stem timing: implications for cortical processing and literacy. *J. Neurosci.* **25**: 9850–9857.
49. Wible, B., T. Nicol & N. Kraus. 2005. Correlation between brain stem and cortical auditory processes in normal and language-impaired children. *Brain* **128**: 417–423.
50. Abrams, D. *et al.* 2006. Auditory brain stem timing predicts cerebral asymmetry for speech. *J. Neurosci.* **26**: 11131–11137.
51. Russo, N.M. *et al.* 2005. Auditory training improves neural timing in the human brain stem. *Behav. Brain Res.* **156**: 95–103.
52. de Boer, J. & R.D. Thornton. 2008. Neural correlates of perceptual learning in the auditory brain stem: efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. *J. Neurosci.* **28**: 4929–4937.
53. Shannon, R.V. 2005. Speech and music have different requirements for spectral resolution. *Int. Rev. Neurobiol.* **70**: 121–134.
54. Gfeller, K. *et al.* 2002. Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *J. Am. Acad. Audiol.* **13**: 132–145.
55. Gfeller, K. *et al.* 2002. Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Ann. Otol. Rhinol. Laryngol.* **111**: 349–356.
56. Leal, M.C. *et al.* 2003. Music perception in adult cochlear implant recipients. *Acta. Otolaryngol.* **123**: 826–835.
57. McDermott, H.J. 2004. Music perception with cochlear implants: a review. *Trends Amplifi.* **8**: 49–82.
58. Limb, C.J. 2006. Cochlear implant-mediated perception of music. *Curr. Opin. Otolaryngol. Head Neck Surg.* **14**: 337–340.
59. Looi, V. *et al.* 2008. The effect of cochlear implantation on music perception by adults with usable pre-operative acoustic hearing. *Int. J. Audiol.* **47**: 257–268.
60. Looi, V. *et al.* 2008. Music perception of cochlear implant users compared with that of hearing aid users. *Ear Hear.* **29**: 421–434.
61. Koelsch, S. *et al.* 2004. Music perception in cochlear implant users: an event-related potential study. *Clin. Neurophysiol.* **115**: 966–972.
62. Guiraud, J. *et al.* 2007. Evidence of a tonotopic organization of the auditory cortex in cochlear implant users. *J. Neurosci.* **27**: 7838–7846.
63. Galvin, J.J., Q.J. Fu & G. Nogaki. 2007. Melodic contour identification by cochlear implant listeners. *Ear Hear.* **28**: 302–319.
64. Debener, S. *et al.* 2008. Source localization of auditory evoked potentials after cochlear implantation. *Psychophysiology* **45**: 20–24.
65. Gilley, P.M. *et al.* 2006. Minimization of cochlear implant stimulus artifact in cortical auditory evoked potentials. *Clin. Neurophysiol.* **117**: 1772–1782.
66. Thai-Van, H. *et al.* 2007. The pattern of auditory brain stem response wave V maturation in cochlear-implanted children. *Clin. Neurophysiol.* **118**: 676–689.
67. Gordon, K.A., B.C. Papsin & R.V. Harrison. 2006. An evoked potential study of the developmental time course of the auditory nerve and brain stem in children using cochlear implants. *Audiol. Neurotol.* **11**: 7–23.
68. Nimmons, G.L. *et al.* 2008. Clinical assessment of music perception in cochlear implant listeners. *Otol. Neurotol.* **29**: 149–155.
69. Cooper, W., E. Tobey & P.C. Loizou. 2008. Music perception by cochlear implant and normal hearing listeners as measures by the Montreal Battery for Evaluation of Amusia. *Ear Hear.* **29**(4): 618–626.
70. Spitzer, J.B., D. Mancuso & M.Y. Cheng. 2008. Development of a clinical test of musical perception: appreciation of music in cochlear implantees (AMICI). *J. Am. Acad. Audiol.* **19**: 56–81.
71. Chorost, M. 2008. *Helping the deaf hear music: a new test measures music perception in cochlear-implant users.* MIT Technological Review. Available at <http://www.technologyreview.com/Infotech/20334/?a=f>.
72. Peng, S.C. *et al.* 2004. Perception and production of mandarin tones in prelingually deaf children with cochlear implants. *Ear Hear.* **25**: 251–264.
73. Lassaletta, L. *et al.* 2008. Changes in listening habits and quality of musical sound after cochlear implantation. *Otolaryngol. Head Neck Surg.* **138**: 363–367.
74. Banai, K., D. Abrams & N. Kraus. 2007. Sensory-based learning disability: insights from brain stem processing of speech sounds. *Int. J. Audiol.* **46**: 524–532.
75. Suga, N. *et al.* 2002. Plasticity and corticofugal modulation for hearing in adult animals. *Neuron* **36**: 9–18.
76. Hoormann, J. *et al.* 1992. The human frequency-following response (FFR): normal variability and relation to the click-evoked brain stem response. *Hearing Res.* **59**: 179–188.

77. Marsh, J., F. Worden & J. Smith. 1970. Auditory frequency-following response: neural or artifact? *Science* **169**: 1222–1223.
78. Smith, J.C., J.T. Marsh & W.S. Brown. 1975. Far-field recorded frequency-following responses: evidence for the locus of brain stem sources. *Electroencephalogr. Clin. Neurophysiol.* **39**: 465–472.
79. Worden, F. & J. Marsh. 1968. Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalogr. Clin. Neurophysiol.* **25**: 42–52.
80. Moushegian, G., A.L. Rupert & R.D. Stillman. 1973. Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalogr. Clin. Neurophysiol.* **35**: 665–667.