



Why would musical training benefit the neural encoding of speech? The OPERA hypothesis

Aniruddh D. Patel*

Department of Theoretical Neurobiology, The Neurosciences Institute, San Diego, CA, USA

Edited by:

Lutz Jäncke, University of Zurich, Switzerland

Reviewed by:

Kai Alter, Newcastle University, UK
Stefan Koelsch, Freie Universität Berlin, Germany

*Correspondence:

Aniruddh D. Patel, Department of Theoretical Neurobiology, The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121, USA.
e-mail: apatel@nsi.edu

Mounting evidence suggests that musical training benefits the neural encoding of speech. This paper offers a hypothesis specifying why such benefits occur. The “OPERA” hypothesis proposes that such benefits are driven by adaptive plasticity in speech-processing networks, and that this plasticity occurs when five conditions are met. These are: (1) **O**verlap: there is anatomical overlap in the brain networks that process an acoustic feature used in both music and speech (e.g., waveform periodicity, amplitude envelope), (2) **P**recision: music places higher demands on these shared networks than does speech, in terms of the precision of processing, (3) **E**motion: the musical activities that engage this network elicit strong positive emotion, (4) **R**epetition: the musical activities that engage this network are frequently repeated, and (5) **A**ttention: the musical activities that engage this network are associated with focused attention. According to the OPERA hypothesis, when these conditions are met neural plasticity drives the networks in question to function with higher precision than needed for ordinary speech communication. Yet since speech shares these networks with music, speech processing benefits. The OPERA hypothesis is used to account for the observed superior subcortical encoding of speech in musically trained individuals, and to suggest mechanisms by which musical training might improve linguistic reading abilities.

Keywords: music, speech, neural plasticity, neural encoding, hypothesis

INTRODUCTION

Recent EEG research on human auditory processing suggests that musical training benefits the neural encoding of speech. For example, across several studies Kraus and colleagues have shown that the neural encoding of spoken syllables in the auditory brainstem is superior in musically trained individuals (for a recent overview, see Kraus and Chandrasekaran, 2010; this is further discussed in *The Auditory Brainstem Response to Speech: Origins and Plasticity*). Kraus and colleagues have argued that experience-dependent neural plasticity in the brainstem is one cause of this enhancement, based on the repeated finding that the degree of enhancement correlates significantly with the amount of musical training (e.g., Musacchia et al., 2007, 2008; Wong et al., 2007; Strait et al., 2009). They suggest that plasticity in subcortical circuits could be driven by descending (“corticofugal”) neural projections from cortex onto these circuits. There are many such projections in the auditory system (exceeding the number of ascending fibers), providing a potential pathway for cortical signals to tune subcortical circuits (Winer, 2006; Kral and Eggermont, 2007; **Figure 1**).

The arguments of Kraus and colleagues have both theoretical and practical significance. From a theoretical standpoint, their proposal contradicts the view that auditory processing is strictly hierarchical, with hardwired subcortical circuits conveying neural signals to cortical regions in a purely feed-forward fashion. Rather, their ideas support a view of auditory processing involving rich two-way interactions between subcortical and cortical regions, with structural malleability at both levels (cf. the “reverse hierarchy theory” of auditory processing, Ahissar et al., 2009). From a practical standpoint, their proposal suggests that the neural encoding of speech

can be enhanced by non-linguistic auditory training (i.e., learning to play a musical instrument or sing). This has practical significance because the quality of brainstem speech encoding has been directly associated with important language skills such as hearing in noise and reading ability (Banai et al., 2009; Parbery-Clark et al., 2009). This suggests that musical training can influence the development of these skills in normal individuals (Moreno et al., 2009; cf. Strait et al., 2010; Parbery-Clark et al., 2011). Furthermore, as noted by Kraus and Chandrasekaran (2010), musical training appears to strengthen the same neural processes that are impaired in individuals with certain speech and language processing problems, such as developmental dyslexia or hearing in noise. This has clear clinical implications for the use of music as a tool in language remediation.

Kraus and colleagues have provided a clear hypothesis for *how* musical training might influence the neural encoding of speech (i.e., via plasticity driven by corticofugal projections). Yet, from a neurobiological perspective, *why* would musical training drive adaptive plasticity in speech processing networks in the first place? Kraus and Chandrasekaran (2010) point out that both music and speech use pitch, timing, and timbre to convey information, and suggest that years of processing these cues in a fine-grained way in music may enhance their processing in the context of speech. The current “OPERA” hypothesis builds on this idea and makes it more specific. It proposes that music-driven adaptive plasticity in speech processing networks occurs because five essential conditions are met. These are: (1) **O**verlap: there is overlap in the brain networks that process an acoustic feature used in both speech and music, (2) **P**recision: music places higher demands on these networks than does speech, in terms of the precision of processing,

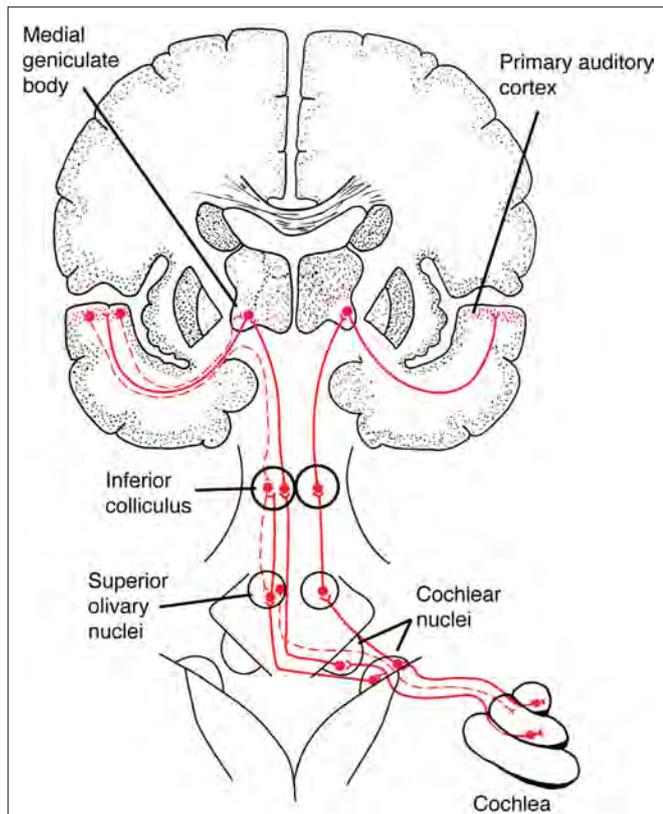


FIGURE 1 | A simplified schematic of the ascending auditory pathway between the cochlea and primary auditory cortex, showing a few of the subcortical structures involved in auditory processing, such as the cochlear nuclei in the brainstem and the inferior colliculus in the midbrain. Solid red lines show ascending auditory pathways, dashed lines show descending (“corticofugal”) auditory pathways. (In this diagram, corticofugal pathways are only shown on one side of the brain for simplicity. See **Figure 2** for a more detailed network diagram). From Patel and Iversen (2007), reproduced with permission.

(3) **Emotion**: the musical activities that engage this network elicit strong positive emotion, (4) **Repetition**: the musical activities that engage this network are frequently repeated, and (5) **Attention**: the musical activities that engage this network are associated with focused attention. According to the OPERA hypothesis, when these conditions are met neural plasticity drives the networks in question to function with higher precision than needed for ordinary speech communication. Yet, since speech shares these networks with music, speech processing benefits.

The primary goals of this paper are to explain the OPERA hypothesis in detail (section The Conditions of the OPERA Hypothesis), and to show how it can be used to generate predictions to drive new research (section Putting OPERA to Work: Musical Training and Linguistic Reading Skills). The example chosen to illustrate the prediction-generating aspect of OPERA concerns relations between musical training and linguistic reading skills (cf. Goswami, 2010). One motivation for this choice is to show that the OPERA hypothesis, while developed on the basis of research on subcortical auditory processing, can also be applied to the cortical processing of acoustic features of speech.

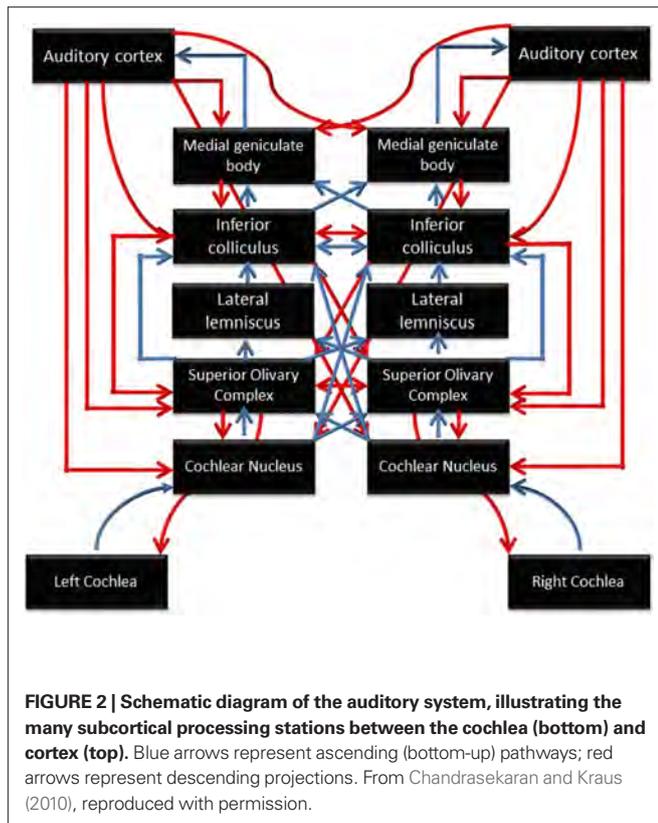
Prior to these core sections of the paper, section “The Auditory Brainstem Response to Speech: Origins and Plasticity” provides background on the auditory brainstem response to speech and evidence for neural plasticity in this response. Following the two core sections, section “Musical vs. Linguistic Training for Speech Sound Encoding” addresses the question of the relative merits of musical vs. linguistic training for improving the neural encoding of speech.

THE AUDITORY BRAINSTEM RESPONSE TO SPEECH: ORIGINS AND PLASTICITY

As noted by Chandrasekaran and Kraus (2010), “before speech can be perceived and integrated with long-term stored linguistic representations, relevant acoustic cues must be represented through a neural code and delivered to the auditory cortex with temporal and spectral precision by subcortical structures.” The auditory pathway is notable for complexity of its subcortical processing, which involves many distinct regions and a rich network of ascending and descending connections (**Figure 2**). Using scalp-recorded EEG, population-level neural responses to speech in subcortical regions can be recorded non-invasively from human listeners (for a tutorial, see Skoe and Kraus, 2010). The brainstem response to the spoken syllable /da/ is shown in **Figure 3**, and the details of this response near syllable onset are shown in **Figure 4**. (Note that the responses depicted in these figures were obtained by averaging together the neural responses to thousands of repetitions of the syllable /da/, in order to boost the signal-to-noise ratio in the scalp-recorded EEG).

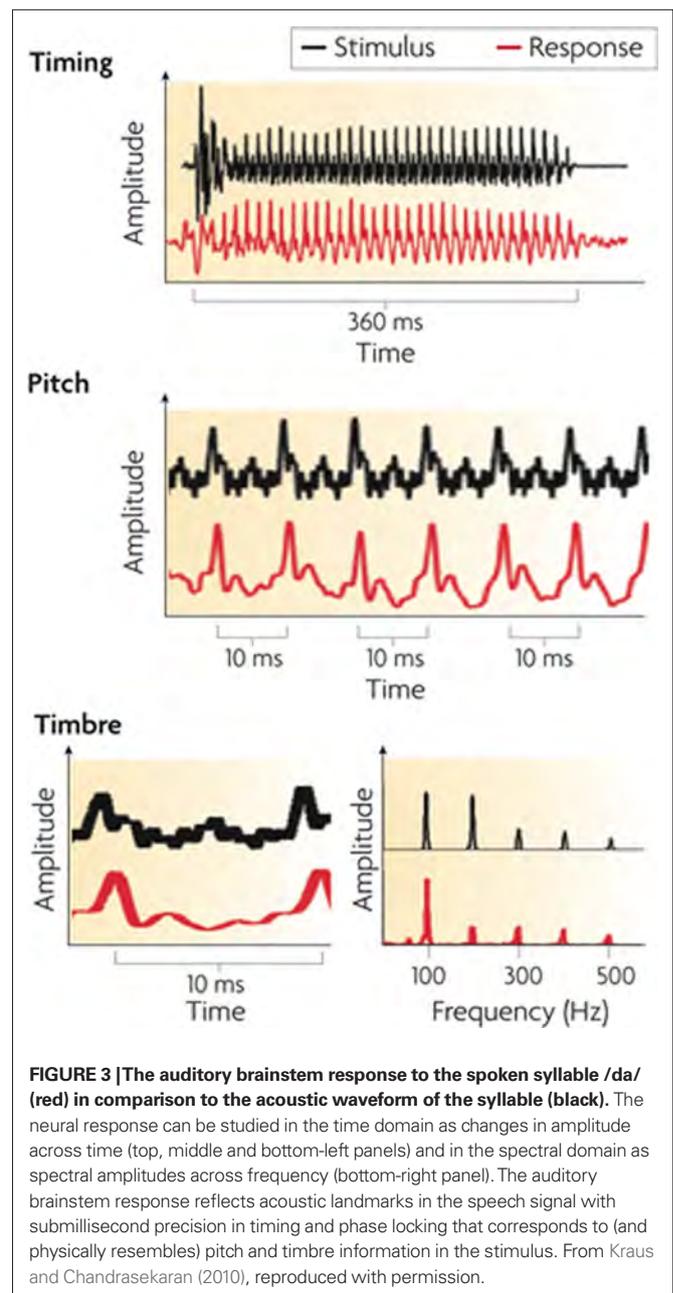
Broadly speaking, the response can be divided into two components: a transient onset response (commencing 5–10 ms after syllable onset, due to neural delays) and an ongoing component known as the frequency-following response (FFR). The transient onset response reflects the synchronized impulse response of neuronal populations in a number of structures, including the cochlea, brainstem, and midbrain. This response shows electrical peaks which are sensitive to the release burst of the /d/ and the formant transition into the vowel. The FFR is the summed responses of electrical currents from structures including cochlear nucleus, superior olivary complex, lateral lemniscus, and the inferior colliculus (Chandrasekaran and Kraus, 2010). The periodicities in the FFR reflect the synchronized component of this population response, arising from neural phase locking to speech waveform periodicities in the auditory nerve and propagating up the auditory pathway (cf. Cariani and Delgutte, 1996). The form of the FFR reflects the transition period between burst onset and vowel, and the vowel itself, via reflection of the fundamental frequency (F0) and some of the lower-frequency harmonics of the vowel.

The auditory brainstem response occurs with high reliability in all hearing subjects (Russo et al., 2004), and can be recorded in passive listening tasks (e.g., while the listener is watching a silent movie or even dozing). Hence it is thought to reflect sensory, rather than cognitive processing (though it may be shaped over longer periods of time via cognitive processing of speech or music, as discussed below). Since the response resembles the acoustic signal in several respects (e.g., in its temporal and spectral structure), correlation measures can be used to quantify the similarity of the neural response to the spoken sound, and hence to quantify the quality of speech sound encoding by the brain (Russo et al., 2004).



As noted previously, the quality of this encoding is correlated with important real-world language abilities, such as hearing in noise and reading. For example, Banai et al. (2009) found that the latency of specific electrical peaks in the auditory brainstem response to /da/ predicted word reading scores, when the effects of age, IQ, and the timing of click-evoked brainstem responses were controlled via partial correlation.

Several recent studies have found that the quality of subcortical speech sound encoding is significantly greater in musically trained individuals (e.g., Musacchia et al., 2007; Wong et al., 2007; Lee et al., 2009; Parbery-Clark et al., 2009; Strait et al., 2009). To take one example, Parbery-Clark et al. (2009) examined the auditory brainstem response in young adults to the syllable /da/ in quiet and in background babble noise. When listening to “da” in noise, musically trained listeners showed shorter-latency brainstem responses to syllable onset and formant transitions relative to their musically untrained counterparts, as well as enhanced representation of speech harmonics and less degraded response morphology. The authors suggest that these differences indicate that musicians had more synchronous neural responses to the syllable. To take another example, Wong et al. (2007) examined the FFR in musically trained and untrained individuals. Prior work had shown that oscillations in the FFR track the F0 and lower harmonics of the voice dynamically over the course of a single syllable (Krishnan et al., 2005). In the study of Wong et al. (2007), native English speakers unfamiliar with Mandarin listened passively to the Mandarin syllable /mi/ (pronounced “me”) with three different lexical tones. The salient finding was that the quality of F0 tracking was superior in musically trained individuals (Figure 5). It is worth noting that musician–non-musician differences in brainstem



encoding in this study, and the study of Parbery-Clark et al. (2009), were evident in the absence of attention to the speech sounds or any overt behavioral task.

One interpretation of the above findings is that they are entirely due to innate anatomical or physiological differences in the auditory systems of musicians and non-musicians. For example, the shorter latencies and stronger FFR responses to speech in musicians could be due to increased synchrony among neurons that respond to harmonically complex sounds, and this in turn could reflect genetically mediated differences in the number or spatial distribution of synaptic connections between neurons. Given the heritability of certain aspects of cortical structure (e.g., cortical thickness, which may in turn reflect the amount of arborization of dendrites, Panizzon et al., 2009), it is plausible that

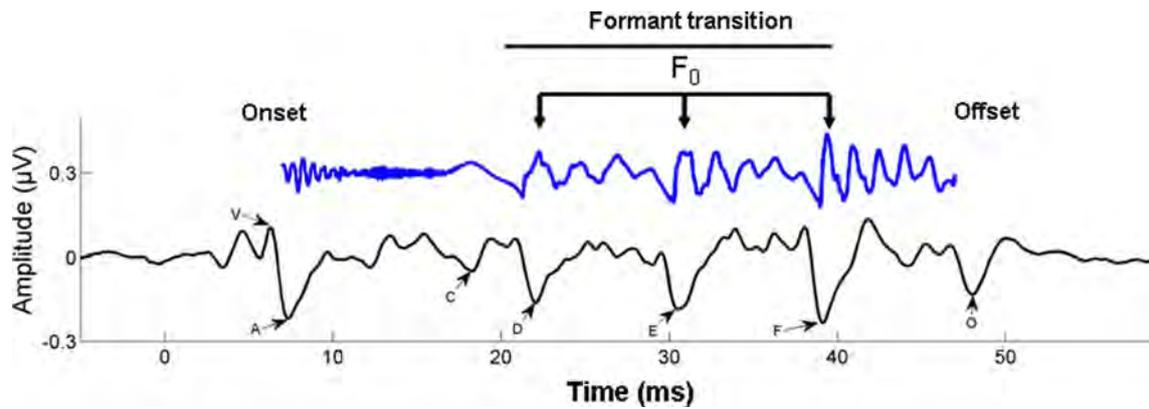


FIGURE 4 | Time–amplitude waveform of a 40-ms synthesized speech stimulus /da/ is shown in blue (time shifted by 6 ms to be comparable with the neural response). The first 10 ms of the syllable are characterized by the onset burst of the consonant /d/; the following 30 ms are the formant transition to the vowel /a/. The time–amplitude waveform of the time-locked brainstem response to the 40-ms /da/ is shown below the stimulus, in black. The onset response (V) begins 6–10 ms following the stimulus, reflecting the

time delay to the auditory brainstem. The start of the formant transition period is marked by wave C, marking the change from the burst to the periodic portion of the syllable, that is, the vowel. Waves D, E, and F represent the periodic portion of the syllable (frequency-following response) from which the fundamental frequency (F0) of the stimulus can be extracted. Finally, wave O marks stimulus offset. From Chandrasekaran and Kraus (2010), reproduced with permission.

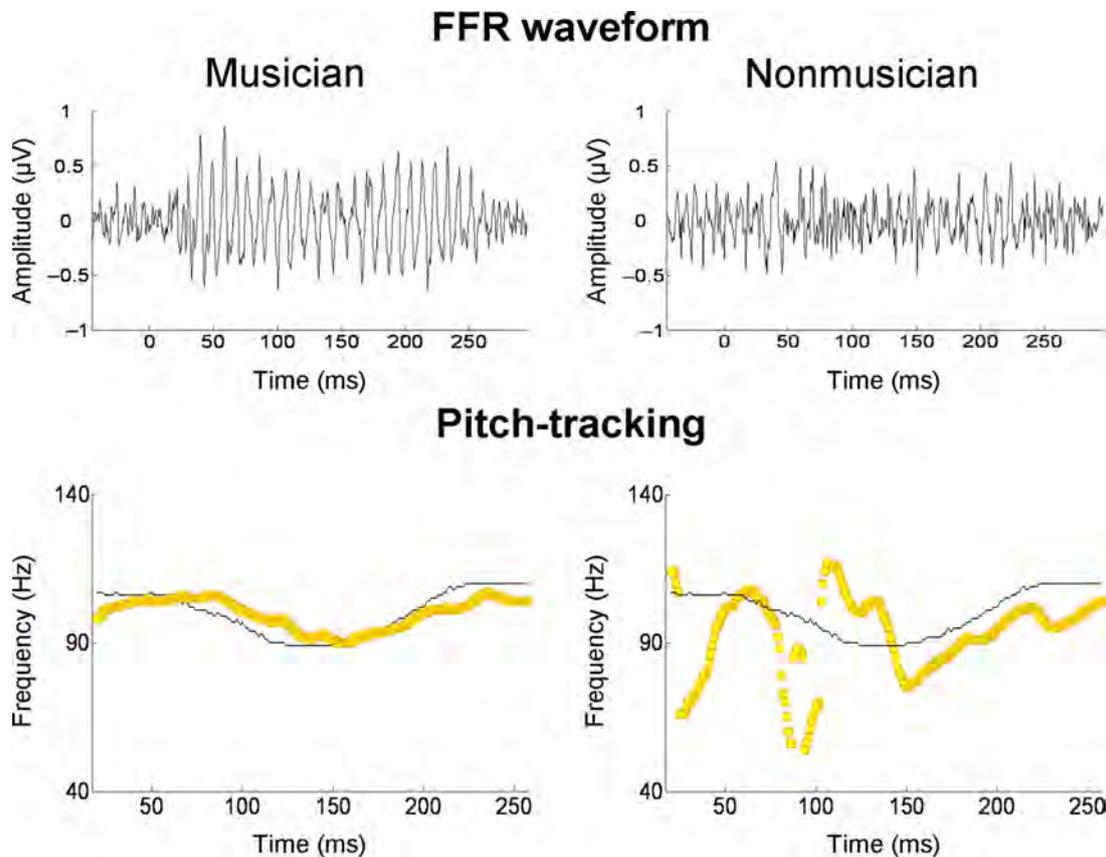


FIGURE 5 | Frequency-following responses to the spoken syllable /mi/ with a “dipping” (tone 3) pitch contour from Mandarin. The top row shows FFR waveforms of a musician and non-musician subject; the bottom row shows the fundamental frequency of the voice (thin black line) and the trajectories (yellow

lines) of the FFR’s primary periodicity, from the same two individuals. For the musician, the FFR waveform is more periodic and its periodicity tracks the time-varying F0 contour of the spoken syllable with greater accuracy. From Wong et al. (2007), reproduced with permission.

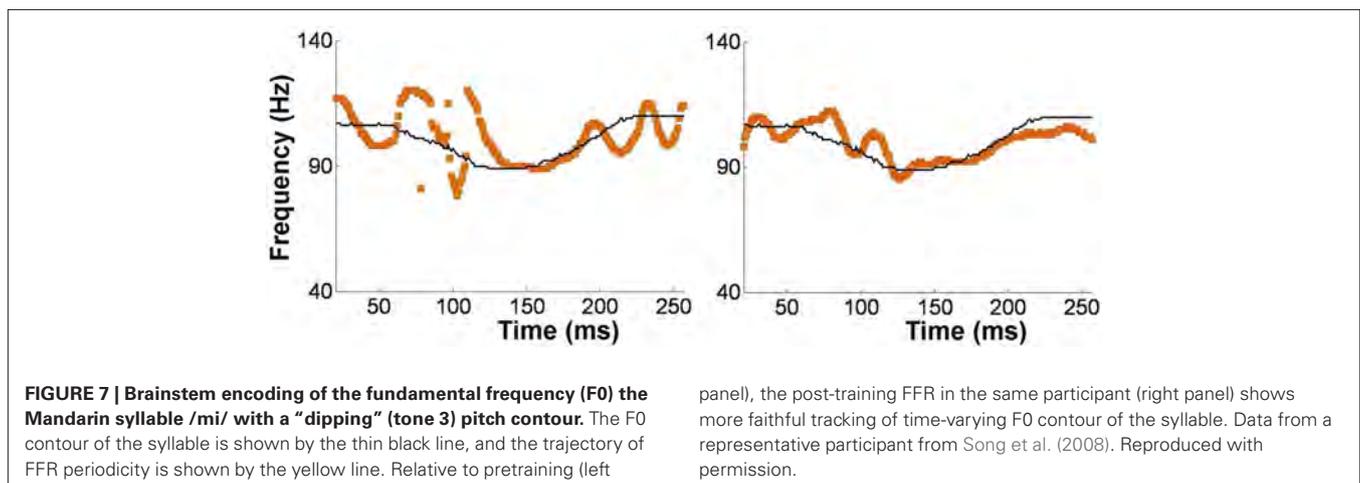
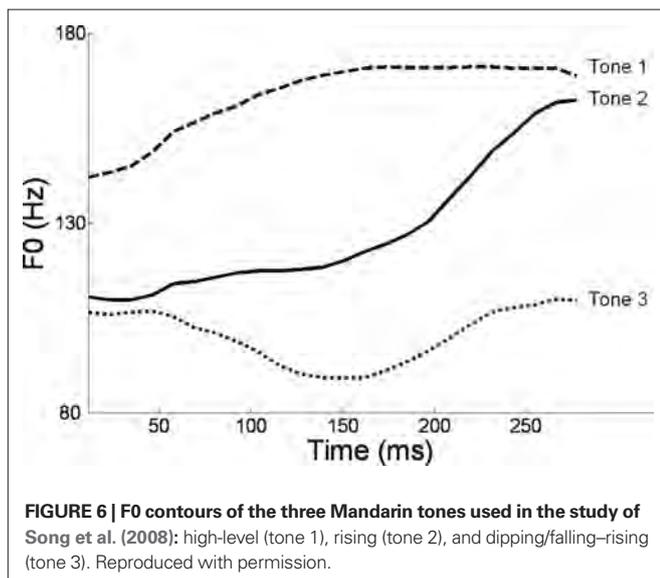
some proportion of the variance in brainstem responses to speech is due to genetic influence on subcortical neuroanatomy. In other words, it is plausible that some individuals are simply born with sharper hearing than others, which confers an advantage in processing any kind of complex sound. If sharper hearing makes an individual more likely to pursue musical training, this could explain the positive association between musical training and the superior neural encoding of speech sounds, in the absence of any experience-dependent brainstem plasticity.

Alternatively, musical training could be a *cause* of the enhanced neural encoding of speech sounds. That is, experience-dependent neural plasticity due to musical training might cause changes in brain structure and/or function that result in enhanced subcortical speech processing. There are several reasons to entertain this possibility. First, there is recent evidence that the brainstem encoding of speech sounds shows training-related neural plasticity, even in adult listeners. This was demonstrated by Song et al. (2008), who had native English speakers learn a vocabulary of six nonsense syllables, each paired with three lexical tones based on Mandarin tones 1–3 (Figure 6). (The participants had no prior knowledge

of tone languages.) For example, the syllable “pesh” meant glass, pencil, or table depending on whether it was spoken with a level, rising, or falling–rising tone. Participants underwent a training program in which each syllable, spoken with a lexical tone, was heard while viewing a picture of the word’s meaning. Quizzes were given periodically to measure word learning. The FFR to an untrained Mandarin syllable (/mi/) with tones 1–3 was measured before and after training. The salient finding was that the quality of FFR tracking of fundamental frequency (F0) of Mandarin tones improved with training. Compared to pretraining measures, participants showed increased energy at the fundamental frequency of the voice and fewer F0 tracking errors (Figure 7). Song et al. (2008) note that this enhancement likely reflects enhanced synchronization of neural firing to stimulus F0, which could in turn result from more neurons firing at the F0 rate, more synchronous firing, or some combination of these two. A striking aspect of these findings is the small amount of training involved in the study: just eight 30-min sessions spread across 2 weeks. These results show that the adult auditory brainstem response to speech is surprisingly malleable, exhibiting plasticity following relatively brief training.

The second reason to consider a role for musical training in enhanced brainstem responses to speech is that several studies which report superior brainstem encoding of speech in musicians also report that the degree of enhancement correlates significantly with the amount of musical training (e.g., Musacchia et al., 2007, 2008; Wong et al., 2007; Lee et al., 2009; Strait et al., 2009). Finally, the third reason is that longitudinal brain-imaging studies have shown that musical training causes changes in auditory cortical structure and function, which are correlated with increased auditory acuity (Hyde et al., 2009; Moreno et al., 2009). As noted previously there exist extensive corticofugal (top-down) projections from cortex to all subcortical auditory structures. Evidence from animal studies suggests that activity in such projections can tune the response patterns of subcortical circuits to sound (see Tzounopoulos and Kraus, 2009 for a review, cf. Kral and Eggermont, 2007; Suga, 2008; Bajo et al., 2010).

Hence the idea that musical training benefits the neural encoding of speech is neurobiologically plausible, though longitudinal randomized controlled studies are needed to establish this with



certainty. To date, randomized controlled studies of the influence of musical training on auditory processing have focused on cortical rather than subcortical signals (e.g., Moreno et al., 2009). Hopefully such work will soon be extended to include subcortical measures. Many key questions remain to be addressed, such as how much musical training is needed before one sees benefits in speech encoding, how large and long-lasting such benefits are, whether cortical changes precede (and cause) subcortical changes, and the relative effect of childhood vs. adult training. Even at this early stage, however, it is worth considering why musical training would benefit the neural encoding of speech.

THE CONDITIONS OF THE OPERA HYPOTHESIS

The OPERA hypothesis aims to explain why musical training would lead to adaptive plasticity in speech-processing networks. According to this hypothesis, such plasticity is engaged because five essential conditions are met by music processing. These are: overlap, precision, emotion, repetition, and attention (as detailed below). It is important to note that music processing does not automatically meet these conditions. Rather, the key point is that music processing has the potential to meet these conditions, and that by specifying these conditions, OPERA opens itself to empirical testing. Specifically, musical activities can be designed which fulfill these conditions, with the clear prediction that they will lead to enhanced neural encoding of speech. Conversely, OPERA predicts that musical activities not meeting the five conditions will not lead to such enhancements.

Note that OPERA is agnostic about the particular cellular mechanisms involved in adaptive neural plasticity. A variety of changes in subcortical circuits could enhance the neural encoding of speech, including changes in the number and spatial distribution of synaptic connections, and/or changes in synaptic efficacy, which in turn can be realized by a broad range of changes in synaptic physiology (Edelman and Gally, 2001; Schnupp et al., 2011, Ch. 7). OPERA makes no claims about precisely which changes are involved, or precisely how corticofugal projections are involved in such changes. These are important questions, but *how* adaptive plasticity is manifested in subcortical networks is a distinct question from *why* such plasticity is engaged in the first place. The OPERA hypothesis addresses the latter question, and is compatible with a range of specific physiological mechanisms for neural plasticity.

The remainder of this section lists the conditions that must be met for musical training to drive adaptive plasticity in speech processing networks, according to the OPERA hypothesis. For the sake of illustration, the focus on one particular acoustic feature shared by speech and music (periodicity), but the logic of OPERA applies to any acoustic feature important for both speech and music, including spectral structure, amplitude envelope, and the timing of successive events. (In section “Putting OPERA to Work: Musical Training and Linguistic Reading Skills,” where OPERA is used to make predictions about how musical training might benefit reading skills, the focus will be on amplitude envelope).

OVERLAP

For musical training to influence the neural encoding of speech, an acoustic feature important for both speech and music perception must be processed by overlapping brain networks. For example,

periodicity is an important feature of many spoken and musical sounds, and contributes to the perceptual attribute of pitch in both domains. At the subcortical level, the auditory system likely uses similar mechanisms and networks for encoding periodicity in speech and music, including patterns of action potential timing in neurons in the auditory pathway between cochlea and inferior colliculus (Cariani and Delgutte, 1996; cf. **Figure 2**). As noted in section “The Auditory Brainstem Response to Speech: Origins and Plasticity,” the synchronous aspect of such patterns may underlie the FFR.

Of course, once the pitch of a sound is determined (via mechanisms in subcortical structures and perhaps in early auditory cortical regions, cf. Bendor and Wang, 2005), then it may be processed in different ways depending on whether it occurs in a linguistic or musical context. For example, neuroimaging of human pitch perception reveals that when pitch makes lexical distinctions between words (as in tone languages), pitch processing shows a left hemisphere bias, in contrast to the typical right-hemisphere dominance for pitch processing (Zatorre and Gandour, 2008). The normal right-hemisphere dominance in the cortical analysis of pitch may reflect enhanced spectral resolution in right-hemisphere auditory circuits (Zatorre et al., 2002), whereas left hemisphere activations in lexical tone processing may reflect the need to interface pitch information with semantic and/or syntactic information in language.

The larger point is that *perceptual attributes* (such as pitch) should be conceptually distinguished from *acoustic features* (such as periodicity). The cognitive processing of a perceptual attribute (such as pitch) can be quite different in speech and music, reflecting the different patterns and functions the attribute has in the two domains. Hence some divergence in the cortical processing of that attribute is to be expected, based on whether it is embedded in speech or music (e.g., Peretz and Coltheart, 2003). On the other hand, the basic encoding of acoustic features underlying that attribute (e.g., periodicity, in the case of pitch) may involve largely overlapping subcortical circuits. After all, compared to auditory cortex, with its vast numbers of neurons and many functionally specialized subregions (Kaas and Hackett, 2000), subcortical structures have far fewer neurons, areas, and connections and hence less opportunity to partition speech and music processing into different circuits. Hence the idea that the sensory encoding of periodicity (or other basic acoustic features shared by speech and music) uses overlapping subcortical networks is neurobiologically plausible.

PRECISION

Let us assume that OPERA condition 1 is satisfied, that is, that a shared acoustic feature in speech and music is processed by overlapping brain networks. OPERA holds that for musical training to influence the neural encoding of speech, music must place higher demands on the nervous system than speech does, in terms of the precision with the feature must be encoded for adequate communication to occur. This statement immediately raises the question of what constitutes “adequate communication” in speech and music. For the purposes of this paper, adequate communication for speech is defined as conveying the semantic and propositional content of spoken utterances, while for music,

it is defined as conveying the structure of musical sequences. Given these definitions, how can one decide whether music places higher demands on the nervous system than speech, in terms of the precision of encoding of a given acoustic feature? One way to address this question is to ask to what extent a perceiver *requires* detailed information about the patterning of that feature in order for adequate communication to occur. Consider the feature of periodicity, which contributes to the perceptual attribute of pitch in both speech and music. In musical melodies, the notes of melodies tend to be separated by small intervals (e.g., one or two semitones, i.e., approximately 6 or 12% changes in pitch, cf. Vos and Troost, 1989), and a pitch movement of just one semitone (~6%) can be structurally very important, for example, when it leads to a salient, out-of-key note that increases the complexity of the melody (e.g., a C# note in the key of C, cf. Eerola et al., 2006). Hence for music perception, detailed information about pitch is important. Indeed, genetically based deficits in fine-grained pitch processing are thought to be one of the important underlying causes of musical tone deafness or “congenital amusia” (Peretz et al., 2007; Liu et al., 2010).

How crucial is detailed information about spoken pitch patterns to the perception of speech? In speech, pitch has a variety of linguistic functions, including marking emphasis and phrase boundaries, and in tone languages, making lexical distinctions between words. Hence there is no doubt that pitch conveys significant information in language. Yet the crucial question is: to what extent does a perceiver *require* detailed information about pitch patterns for adequate communication to occur? One way to address this question is to manipulate the pitch contour of natural sentences and see how this impacts a listener’s ability to understand the semantic and propositional content of the sentences. Recent research on this topic has shown that spoken language comprehension is strikingly robust to manipulations of pitch contour. Patel et al. (2010) measured the intelligibility of natural Mandarin Chinese sentences with intact vs. flattened (monotone) pitch contours (the latter created via speech resynthesis, with pitch fixed at the mean fundamental frequency of the sentence). Native speakers of Mandarin found the monotone sentences just as intelligible as the natural sentences when heard in a quiet background. Hence despite the complete removal of all details of pitch variation, the sentences were fully intelligible. How is this possible? Presumably listeners used the remaining phonetic information and their knowledge of Mandarin to guide their perception in a way that allowed them to infer which words were being said. The larger point is that spoken language comprehension is remarkably robust to lack of detail in pitch variation, and this presumably relaxes the demands placed on high-precision encoding of periodicity patterns.

It seems likely that this sort of robustness is not just limited to the acoustic feature of periodicity. One reason that spoken language comprehension is so robust is that it involves integrating multiple cues, some of which provide redundant sources of information regarding a sound’s phonological category. For example, in judging whether a sound within a word is a /b/ or a /p/, multiple acoustic cues are relevant, including voice onset time (VOT), vowel length, fundamental frequency (F0; i.e., “microintonational” perturbations of F0, which behave differently after voiced vs. voiceless stop consonants), and first and second formant patterns (i.e.,

their frequencies and rates of change; Stevens, 1998). While any individual cue may be ambiguous, when cues are integrated and interpreted in light of the current phonetic context, they provide a strong pointer to the intended phonological category being communicated (Toscano and McMurray, 2010; Toscano et al., 2010). Furthermore, when words are heard in sentence context, listeners benefit from multiple knowledge sources (including semantics, syntax, and pragmatics) which provide mutually interacting constraints that help a listener identify the words in the speech stream (Mattys et al., 2005).

Hence one can hypothesize that the use of context-based cue integration and multiple knowledge sources in word recognition helps relax the need for a high degree of precision in acoustic analysis of the speech signal, at least in terms of what is needed for adequate communication (cf. Ferreira and Patson, 2007). Of course, this is not to say that fine phonetic details are not relevant for linguistic communication. There is ample evidence that listeners are sensitive to such details when they are available (e.g., McMurray et al., 2008; Holt and Idemaru, 2011). Furthermore, these details help convey rich “indexical” information about speaker identity, attitude, emotional state, and so forth. Nevertheless, the question at hand is what demands are placed on the nervous system of a perceiver in terms of the precision of acoustic encoding for basic semantic/propositional communication to occur (i.e., “adequate” linguistic communication). How do these demands compare to the demands placed on a perceiver for adequate musical communication?

As defined above, adequate musical communication involves conveying the structure of musical sequences. Conveying the structure of a musical sequence involves playing the intended notes with appropriate timing. Of course, this is not to say that a few wrong notes will ruin musical communication (even professionals make occasional mistakes in complex passages), but by and large the notes and rhythms of a performance need to adhere fairly closely to a specified model (such as a musical score, or to a model provided by a teacher) for a performance to be deemed adequate. Even in improvisatory music such as jazz or the classical music of North India, a performer must learn to produce the sequence of notes they intend to play, and not other, unintended notes. Crucially, this requirement places fairly strict demands on the regulation of pitch and timing, because intended notes (whether from a musical score, a teacher’s model, or a sequence created “on the fly” in improvisation) are never far in pitch or duration from unintended notes. For example, as mentioned above, a note just one semitone away from an intended note can make a perceptually salient change in a melody (e.g., when the note departs from the prevailing musical key). Similarly, a note that is just a few hundred milliseconds late compared to its intended onset time can make a salient change in the rhythmic feel of a passage (e.g., the difference between an “on beat” note and a “syncopated” note). In short, conveying musical structure involves a high degree of precision, due to the fact that listeners use fine acoustic details in judging the structure of the music they are hearing. Furthermore, listeners use very fine acoustic details in judging the expressive qualities of a performance. Empirical research has shown that expressive (vs. deadpan) performances involve subtle, systematic modifications to the duration, intensity, and (in certain instruments) pitch of notes relative to

their nominal score-based values (Repp, 1992; Clynes, 1995; Palmer, 1997). Listeners are quite sensitive to these inflections and use them in judging the emotional force of musical sequences (Bhatara et al., 2011). Indeed, neuroimaging research has shown that expressive performances containing such inflections are more likely (than deadpan performances) to activate limbic and paralimbic brain areas associated with emotion processing (Chapin et al., 2010). This helps explain why musical training involves learning to pay attention to the fine acoustic details of sound sequences and to control them with high-precision, since these details matter for the aesthetic and emotional qualities of musical sequences.

Returning to our example of pitch variation in speech vs. music, the above paragraph suggests that the adequate communication of melodic music *requires* the detailed regulation and perception of pitch patterns. According to OPERA, this puts higher demands on the sensory encoding of periodicity than does speech, and helps drive experience-dependent plasticity in subcortical networks that encode periodicity. Yet since speech and music share such networks, speech processing benefits (cf. the study of Wong et al., 2007, described in section The Auditory Brainstem Response to Speech: Origins and Plasticity).

Yet would this benefit in the neural encoding of voice periodicity have any consequences for real-world language skills? According to the study of Patel et al. (2010) described above, the details of spoken pitch patterns are not essential for adequate spoken language understanding, even in the tone language Mandarin. So why would superior encoding of voice pitch patterns be helpful to a listener? One answer to this question is suggested by another experimental manipulation in the study of Patel et al. (2010). In this manipulation, natural vs. monotone Mandarin sentences were embedded in background babble noise. In this case, the monotone sentences were significantly less intelligible to native listeners (in the case where the noise was as loud as the target sentence, monotone sentences were 20% less intelligible than natural sentences). The precise reasons why natural pitch modulation enhances sentence intelligibility in noise remain to be determined. Nevertheless, based on these results it seems plausible that a nervous system which has high-precision encoding of vocal periodicity would have an advantage in understanding speech in a noisy background, and recent empirical data support this conjecture (Song et al., 2010). Combining this finding with the observation that musicians show superior brainstem encoding of voice F0 (Wong et al., 2007) leads to the idea that musically trained listeners should show better speech intelligibility in noise.

In fact, it has recently been reported that musically trained individuals show superior intelligibility for speech in noise, using standard tests (Parbery-Clark et al., 2009). This study also examined brainstem responses to speech, but rather than focusing on the encoding of periodicity in syllables with different pitch contour, it examined the brainstem response to the syllable /da/ in terms of latency, representation of speech harmonics, and overall response morphology. When /da/ was heard in noise, all of these parameters were enhanced in musically trained individuals compared to their untrained counterparts. This suggests that musical training influences the encoding of the temporal onsets and spectral details of speech, in addition to the encoding of voice periodicity, all of which may contribute to enhanced speech perception in noise (cf. Chandrasekaran et al., 2009).

This section has argued that music perception is often likely to place higher demands on the encoding of certain acoustic features than does speech perception, at least in terms of what is needed for adequate communication in the two domains. However, OPERA makes no *a priori* assumption that influences between musical and linguistic neural encoding are unidirectional. Hence an important question for future work is whether certain types of linguistic experience with heightened demands in terms of auditory processing (e.g., multilingualism, or learning a tone language) can impact the neural encoding of music (cf. Bidelman et al., 2011). This is an interesting question, but is not explored in the current paper. Instead, the focus is on explaining why musical training would benefit the neural encoding of speech, given the growing body of evidence for such benefits and the practical importance of such findings.

EMOTION, REPETITION, AND ATTENTION

For musical training to enhance the neural encoding of speech, the musical activities that engage speech-processing networks must elicit strong positive emotion, be frequently repeated, and be associated with focused attention. These factors work in concert to promote adaptive plasticity, and are discussed together here.

As noted in the previous section, music can place higher demands on the nervous system than speech does, in terms of the precision of encoding of a particular acoustic feature. Yet this alone is not enough to drive experience-dependent plasticity to enhance the encoding of that feature. There must be some (internal or external) motivation to enhance the encoding of that feature, and there must be sufficient opportunity to improve encoding over time. Musical training has the potential to fulfill these conditions. Accurate music performance relies on accurate perception of the details of sound, which is presumably associated in turn with high-precision encoding of sound features. Within music, accurate performance is typically associated with positive emotion and reward, for example, via internal satisfaction, praise from others, and from the pleasure of listening to well-performed music, in this case music produced by oneself or the group one is in. (Note that according to this view, the particular emotions expressed by the music one plays, e.g., joy, sadness, tranquility, etc., are not crucial. Rather, the key issue is whether the musical experience as a whole is emotionally rewarding.) Furthermore, accurate performance is typically acquired via extensive practice, including frequent repetitions of particular pieces. The neurobiological association between accurate performance and emotional rewards, and the opportunity to improve with time (e.g., via extensive practice) create favorable conditions for promoting plasticity in the networks that encode acoustic features.

Another factor likely to promote plasticity is focused attention on the details of sound during musical training. Animal studies of auditory training have shown that training-related cortical plasticity is facilitated when sounds are actively attended vs. experienced passively (e.g., Fritz et al., 2005; Polley et al., 2006). As noted by Jagadeesh (2006), attention “marks a particular set of inputs for special treatment in the brain,” perhaps by activating cholinergic systems or increasing the synchrony of neural firing, and this modulation plays an important role in inducing neural plasticity (cf. Kilgard and Merzenich, 1998; Thiel, 2007; Weinberger, 2007). While animal studies of attention and plasticity have focused on cortical circuits, it is worth recalling that the auditory system has rich

cortico-subcortical (corticofugal) connections, so that changes at the cortical level have the potential to influence subcortical circuits, and hence, the basic encoding of sound features (Schofield, 2010).

Focused attention on sound may also help resolve an apparent “bootstrapping” problem raised by OPERA: if musical training is to improve the precision of auditory encoding, how does this process get started? For example, focusing on pitch, how does the auditory system adjust itself to register finer pitch distinctions if it cannot detect them in the first place? One answer might be that attention activates more neurons in the frequency (and perhaps periodicity) channels that encode pitch, such that more neurons are recruited to deal with pitch, and hence acuity improves (P. Carians, pers. comm.; cf. Recanzone et al., 1993; Tervaniemi et al., 2009).

It is worth noting that the emotion, repetition, and attention criteria of OPERA are falsifiable. Imagine a child who is given weekly music lessons but who dislikes the music he or she is taught, who does not play any music outside of lessons, and who is not very attentive to the music during lessons. In such circumstances, OPERA predicts that musical training will not result in enhanced neural encoding of speech.

PUTTING OPERA TO WORK: MUSICAL TRAINING AND LINGUISTIC READING SKILLS

There is growing interest in links between musical training, auditory processing, and linguistic reading skills. This stems from two lines of research. The first line has shown a relationship between musical abilities and reading skills in normal children, for example, via correlational and experimental studies (e.g., Anvari et al., 2002; Moreno et al., 2009). For example, Moreno et al. (2009) assigned normal 8-year olds to 6 months of music vs. painting lessons, and found that after musical (but not painting) training, children showed enhanced reading abilities and improved auditory discrimination in speech, with the latter shown by both behavioral and neural measures (scalp-recorded cortical EEG).

The second line of research has demonstrated that a substantial portion of children with reading problems have auditory processing deficits, leading researchers to wonder whether musical training might be helpful for such children (e.g., Overy, 2003; Tallal and Gaab, 2006; Goswami, 2010). For example, Goswami has drawn attention to dyslexics’ impairments in discriminating the rate of change of amplitude envelope at a sound’s onset (its “rise time”; **Figure 8**), and have shown that such deficits predict reading abilities even after controlling for the effects of age and IQ. Of course, rise-time deficits are not the only auditory processing deficits that have been associated with dyslexia (Tallal and Gaab, 2006; Vandermosten et al., 2010), but they are of interest to speech–music studies because they suggest a problem with encoding amplitude envelope patterns, and amplitude envelope is an important feature in both speech and music perception.

In speech, the amplitude envelope is the relatively slow (syllable-level) modulation of overall energy within which rapid spectral changes take place. Amplitude envelopes play an important role as cues to speech rhythm (e.g., stress) and syllable boundaries, which in turn help listeners segment words from the flow of speech (Cutler, 1994). Envelope fluctuations in speech have most of their energy in the 3–20 Hz range, with a peak around 5 Hz (Greenberg, 2006), and experimental manipulations have shown that envelope modulations

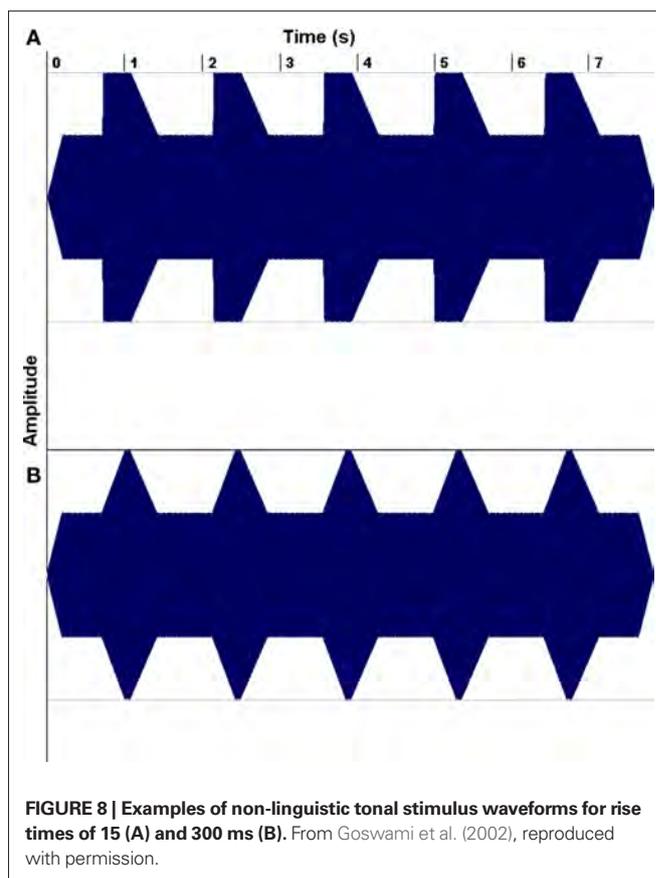


FIGURE 8 | Examples of non-linguistic tonal stimulus waveforms for rise times of 15 (A) and 300 ms (B). From Goswami et al. (2002), reproduced with permission.

in this range are necessary for speech intelligibility (Drullman et al., 1994; cf. Ghitza and Greenberg, 2009). In terms of connections to reading, Goswami has suggested that problems in envelope perception during language development could result in less robust phonological representations at the syllable-level, which would then undermine the ability to consciously segment syllables into individual speech sounds (phonemes). Phonological deficits are a core feature of dyslexia (if the dyslexia is not primarily due to visual problems), likely because reading requires the ability to segment words into individual speech sounds in order to map the sounds onto visual symbols. In support of Goswami’s ideas about relations between envelope processing and reading, a recent meta-analysis of studies measuring dyslexics’ performance on non-speech auditory tasks and on reading tasks reported that amplitude modulation and rise time discrimination were linked to developmental dyslexia in 100% of the studies that they reviewed (Hämäläinen et al., in press).

Amplitude envelope is also an important acoustic feature of musical sounds. Extensive psychophysical research has shown that envelope is one of the major contributors to a sound’s musical timbre (e.g., Caclin et al., 2005). For example, one of the important cues that allows a listener to distinguish between the sounds of a flute and a French horn is the amplitude envelope of their musical notes (Strong and Clark, 1967). Furthermore, the slope of the amplitude envelope at tone onset is an important cue to the perceptual attack time of musical notes, and thus to the perception of rhythm and timing in sequences of musical events (Gordon, 1987). Given that envelope is an important acoustic feature in both speech and music,

the OPERA hypothesis leads to the prediction that musical training which relies on high-precision envelope processing could benefit the neural processing of speech envelopes, via mechanisms of adaptive neural plasticity, *if the five conditions of OPERA are met*. Hence the rest of this section is devoted to examining these conditions in terms of envelope processing in speech and music.

The first condition is that envelope processing in speech and music relies on overlapping neural circuitry. What is known about amplitude envelope processing in the brain? Research on this question in humans and other primates has focused on cortical responses to sound. It is very likely, however, that subcortical circuits are involved in envelope encoding, so that measures of envelope processing taken at the cortex reflect in part the encoding capacities of subcortical circuits. Yet since extant primate data come largely from cortical studies, those will be the focus here.

Primate neurophysiology research suggests that the temporal envelope of complex sound is represented by population-level activity of neurons in auditory cortex (Nagarajan et al., 2002). In human auditory research, two theories suggest that slower temporal modulations in sounds (such as amplitude envelope) are preferentially processed by right-hemisphere cortex (Zatorre et al., 2002; Poeppel, 2003). Prompted by these theories, Abrams et al. (2008) recently examined cortical encoding of envelope using scalp EEG. They collected neural responses to an English sentence (“the young boy left home”) presented to healthy 9- to 13-year-old children. (The children heard hundreds of repetitions of the sentence in one ear, while watching a movie and hearing the soundtrack quietly in the other ear.) The EEG response to the sentence was averaged across trials and low-pass filtered at 40 Hz to focus on cortical responses and how they might reflect the amplitude envelope of the sentence. Abrams et al. discovered that the EEG signal at temporal electrodes over both hemispheres tracked the amplitude envelope of the spoken sentence. (The authors argue that the neural EEG signals they studied are likely to arise from activity in secondary auditory cortex.) Notably, the quality of tracking, as measured by cross-correlating the EEG waveform with the speech envelope, was far superior in right-hemisphere electrodes (approx. 100% superior), in contrast to the usual left hemisphere dominance for spoken language processing (Hickok and Poeppel, 2007; **Figure 9**). (In

light of Goswami’s ideas, described above, it is worth noting that in a subsequent study Abrams et al. [2009] repeated their study with good vs. poor readers, aged 9–15 years. The good readers showed a strong right-hemisphere advantage for envelope tracking in multiple measures [e.g., higher cross-correlation, shorter-latency between neural response and speech input], while poor readers did not. Furthermore, empirical measures of neural envelope tracking predicted reading scores, even after controlling for IQ differences between good and poor readers).

Thus it appears that right-hemisphere auditory cortical regions (and very likely, subcortical inputs to these regions) are involved in envelope processing in speech. Are the same circuits involved in envelope processing in music? This remains to be tested directly, and could be studied using an individual differences approach. For example, one could use the design of Abrams et al. (2008) to examine cortical responses to the amplitude envelope of spoken sentences and musical melodies in the same listeners, with the melodies played using instruments that have sustained (vs. percussive) notes, and hence envelopes reminiscent of speech syllables. If envelope processing mechanisms are shared between speech and music, then one would predict that individuals would show similar envelope-tracking quality across the two domains. While such data are lacking, there are indirect indications of a relationship between envelope processing in music and speech. In music, one set of abilities that should depend on envelope processing are rhythmic abilities, because such abilities depend on sensitivity to the timing of musical notes, and envelope is an important cue for the perceptual onset and duration of event onsets in music (Gordon, 1987). Hence if musical and linguistic amplitude envelopes are processed by overlapping brain circuits, one would expect (somewhat counterintuitively) a relationship between musical rhythmic abilities and reading abilities. Is there any evidence for such a relationship? In fact, Goswami and colleagues have reported that dyslexics have problems with musical rhythmic tasks (e.g., Thompson and Goswami, 2008; Huss et al., 2011), and have shown that their performance on these tasks correlates with their reading skills, after controlling for age and IQ. Furthermore, normal 8-year-old children show positive correlations between performance on rhythm discrimination tasks (but not pitch discrimination tasks) and reading tasks, even after factoring out effects of age, parental education, and the number of hours children spend reading per week (Corrigall and Trainor, 2010).

Based on such findings, let us assume for the sake of argument that the first condition of OPERA is satisfied, that is, that envelope processing in speech and music utilizes overlapping brain networks. It is not required that such networks are entirely overlapping, simply that they overlap to a significant degree at some level of processing (e.g., subcortical, cortical, or both). With such an assumption, one can then turn to the next component of OPERA and consider what sorts of musical tasks would promote high-precision envelope processing. In ordinary musical circumstances, the amplitude envelope of sounds is an acoustic feature relevant to timbre, and while timbre is an important attribute of musical sound, it (unlike pitch) is rarely a primary structural parameter for musical sequences, at least in Western melodic music (Patel, 2008). Hence in order to encourage high-precision envelope processing in music, it may be necessary

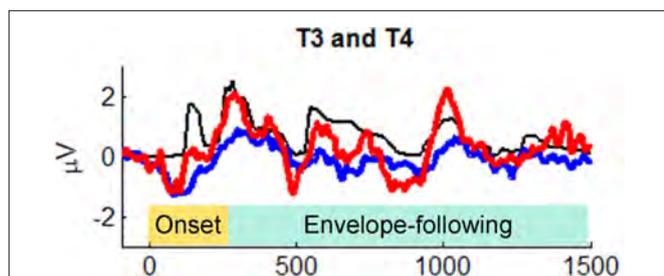


FIGURE 9 | Grand average cortical responses from temporal electrodes T3 and T4 (red: right hemisphere, blue: left hemisphere) and broadband speech envelope (black) for the sentence “the young boy left home.” Ninety-five milliseconds of the prestimulus period is plotted. The speech envelope was shifted forward in time 85 ms to enable comparison to cortical responses. From Abrams et al. (2008), reproduced with permission.

to devise novel musical activities, rather than relying on standard musical training. For example, using modern digital technology in which keyboards can produce any type of synthesized sounds, one could create synthesized sounds with similar spectral content but slightly different amplitude envelope patterns, and create musical activities (e.g., composing, playing, listening) which rely on the ability to distinguish such sounds. Note that such sounds need not be musically unnatural: acoustic research on orchestral wind-instrument tones has shown, for example, that the spectrum of the flute, bassoon, trombone, and French horn are similar, and that listeners rely on envelope cues in distinguishing between these instruments (Strong and Clark, 1967). The critical point is that success at the musical activity should *require* high-precision processing of envelope patterns.

For example, if working with young children, one could create computer-animated visual characters who “sing” with different “voices” (i.e., their “songs” are non-verbal melodies in which all tones have a particular envelope shape). After learning to associate different visual characters and their voices, one could then do games where novel melodies are heard without visual cues and the child has to guess which character is singing. Such “envelope training” games could be done adaptively, so that at the start of training, the envelope shapes of the different characters are quite different, and then once successful discrimination is achieved, new characters are introduced whose voices are more similar in terms of envelope cues.

According to OPERA, if this sort of training is to have any effect on the precision of envelope encoding by the brain, the musical tasks should be associated with strong positive emotion, extensive repetition, and focused attention. Fortunately, music processing is known to have a strong relationship to the brain’s emotion systems (Koelsch, 2010) and it seems plausible that the musical tasks could be made pleasurable (e.g., via the use of attractive musical sounds and melodies, and stimulating rewards for good performance). Furthermore, if they are sufficiently challenging, then participants will likely want to engage in them repeatedly and with focused attention.

To summarize, the OPERA hypothesis predicts that musical training which requires high-precision amplitude envelope processing will benefit the neural encoding of amplitude envelopes in speech, via mechanisms of neural plasticity, if the five conditions of OPERA are met. Based on research showing relationships between envelope processing and reading abilities (e.g., Goswami, 2010), this in turn may benefit linguistic reading skills.

MUSICAL VS. LINGUISTIC TRAINING FOR SPEECH SOUND ENCODING

As described in the previous section, the OPERA hypothesis leads to predictions for how musical training might benefit specific linguistic abilities via enhancements of the neural encoding of speech. Yet this leads to an obvious and important question. Why attempt to improve the neural encoding of speech by training in other domains? If one wants to improve speech processing, would it not be more effective to do acoustic training in the context of speech? Indeed, speech-based training programs that aim to improve sensitivity to acoustic features of speech are now widely used (e.g., the Fast ForWord program, cf. Tallal and Gaab, 2006). It

is important to note that the OPERA hypothesis is not proscriptive: it says nothing about the relative merits of musical vs. linguistic training for speech sound encoding. Instead, it tries to account for why musical training would benefit speech sound encoding in the first place, given the growing empirical evidence that this is indeed the case. Hence the relative efficacy of musical vs. linguistic training for speech sound encoding is an empirical question which can only be resolved by direct comparison in future studies. A strong motivation for conducting such comparisons is the demonstration that non-linguistic auditory perceptual training generalizes to linguistic discrimination tasks that rely on acoustic cues similar to those trained in the non-linguistic context (Lakshminarayanan and Tallal, 2007).

While direct comparisons of musical vs. linguistic training have yet to be conducted, it is worth considering some of the potential merits of music-based training. First, musical activities are often very enjoyable, reflecting the rich connections between music processing and the emotion systems of the brain (Koelsch, 2010; Salimpoor et al., 2011). Hence it may be easier to have individuals (especially children) participate repeatedly in training tasks, particularly in home-based tasks which require voluntary participation. Second, if an individual is experiencing a language problem, then a musical activity may not carry any of the negative associations that have developed around the language deficits and language-based tasks. This increases the chances of associating auditory training with strong positive emotions, which in turn may facilitate neural plasticity (as noted in section Emotion, Repetition, and Attention). Third, speech is acoustically complex, with many acoustic features varying at the same time (e.g., amplitude envelope, harmonic spectrum), and its perception automatically engages semantic processes that attempt to extract conceptual meaning from the signal, which draws attention away from the acoustic details of the signal. Musical sounds, in contrast, can be made acoustically relatively simple, with variation primarily occurring along specific dimensions that are the focus of training. For example, if the focus is on training sensitivity to amplitude envelope (cf. section Putting OPERA to work: musical training and linguistic reading skills), sounds can be created in which the spectral content is simple and stable, and in which the primary differences between tones are in envelope structure. Furthermore, musical sounds do not engage semantic processing, leaving the perceptual system free to focus more attention on the details of sound. This ability of musical sounds to isolate particular features for attentive processing could lower the overall complexity of auditory training tasks, and hence make it easier for individual to make rapid progress in increasing their sensitivity to such features, via mechanisms of neural plasticity.

A final merit of musical training is that it typically involves building strong sensorimotor links between auditory and motor skills (i.e., the sounds one produces are listened to attentively, in order to adjust performance to meet a desired model). Neuroscientific research suggests that sensorimotor musical training is a stronger driver of neural plasticity in auditory cortex than purely auditory musical training (Lappe et al., 2008). Hence musical training provides an easy, ecologically natural route to harness the power of sensorimotor processing to drive adaptive neural plasticity in the auditory system.

CONCLUSION

The OPERA hypothesis suggests that five essential conditions must be met in order for musical training to drive adaptive plasticity in speech processing networks. This hypothesis generates specific predictions, for example, regarding the kinds of musical training that could benefit reading skills. It also carries an implication regarding the notion of “musical training.” Musical training can involve different skills depending on what instrument and what aural abilities are being trained. (For example, learning to play the drums places very different demands on the nervous system than learning to play the violin.) The OPERA hypothesis states that the benefits of musical training depend on the particular acoustic features emphasized in training, the demands that music places on those features in terms of the precision of processing, and the degree of emotional reward, repetition and attention associated with musical activities. According to this hypothesis, simply giving an individual music lessons may not result in any benefits for speech processing. Indeed, depending on the acoustic feature being trained (e.g., amplitude envelope), learning a standard musical instrument may not be an effective way to enhance neural processing of that feature. Instead, novel

digital instruments may be necessary, which allow controlled manipulation of sound features. Thus OPERA raises the idea that in the future, musical activities aimed at benefiting speech processing should be purposely shaped to optimize the effects of musical training.

From a broader perspective, OPERA contributes to the growing body of research aimed at understanding the relationship between musical and linguistic processing in the human brain (Patel, 2008). Understanding this relationship will likely have significant implications for how we study and treat a variety of language disorders, ranging from sensory to syntactic processing (Jentschke et al., 2008; Patel et al., 2008).

ACKNOWLEDGMENTS

Supported by Neurosciences Research Foundation as part of its research program on music and the brain at The Neurosciences Institute, where ADP is the Esther J. Burnham Senior Fellow. I thank Dan Abrams, Peter Cariani, Bharath Chandrasekaran, Jeff Elman, Lori Holt, John Iversen, Nina Kraus, Stephen McAdams, Stefanie Shattuck-Hufnagel, and the reviewers of this paper for insightful comments.

REFERENCES

- Abrams, D. A., Nicol, T., Zecker, S., and Kraus, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J. Neurosci.* 28, 3958–3965.
- Abrams, D. A., Nicol, T., Zecker, S., and Kraus, N. (2009). Abnormal cortical processing of the syllable rate of speech in poor readers. *J. Neurosci.* 29, 7686–7693.
- Ahissar, M., Nahum, M., Nelken, I., and Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 285–299.
- Anvari, S., Trainor, L. J., Woodside, J., and Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *J. Exp. Child Psychol.* 83, 111–130.
- Bajo, V. M., Nodal, F. R., Moore, D. R., and King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13, 252–260.
- Banai, K., Hornik, J., Skoe, E., Nicol, T., Zecker, S., and Kraus, N. (2009). Reading and subcortical auditory function. *Cereb. Cortex* 19, 2699–2707.
- Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature* 436, 1161–1165.
- Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., and Levitin, D. J. (2011). Perception of emotional expression in musical performance. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 921–934.
- Bidelman, G. M., Gandour, J. T., and Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* 23, 425–434.
- Caclin, A., McAdams, S., Smith, B. K., and Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *J. Acoust. Soc. Am.* 118, 471–482.
- Cariani, P. A., and Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* 76, 1698–1716.
- Chandrasekaran, B., Hornik, J., Skoe, E., Nicol, T., and Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. *Neuron* 64, 311–319.
- Chandrasekaran, B., and Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47, 236–246.
- Chapin, H., Jantzen, K., Kelso, S. J. A., Steinberg, F., and Large, E. (2010). Dynamic emotional and neural responses to music depend on performance expression and listener experience. *PLoS ONE* 5: e13812. doi: 10.1371/journal.pone.0013812
- Clynes, M. (1995). Microstructural musical linguistics: composers’ pulses are liked most by the best musicians. *Cognition* 55, 269–310.
- Corrigall, K., and Trainor, L. (2010). “The predictive relationship between length of musical training and cognitive skills in children,” in *Paper presented at the 11th Intl. Conf. on Music Perception & Cognition (ICMPC11)*, August 2010, Seattle, WA.
- Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua* 92, 81–104.
- Drullman, R., Festen, J. M., and Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 95, 1053–1064.
- Edelman, G. M., and Gally, J. (2001). Degeneracy and complexity in biological systems. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13763–13768.
- Eerola, T., Himberg, T., Toiviainen, P., and Louhivuori, J. (2006). Perceived complexity of western and African folk melodies by western and African listeners. *Psychol. Music* 34, 337–371.
- Ferreira, F., and Patson, N. D. (2007). The ‘good enough’ approach to language comprehension. *Lang. Linguist. Compass* 1, 71–83.
- Fritz, J., Elhilali, M., and Shamma, S. (2005). Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear. Res.* 206, 159–176.
- Ghitza, O., and Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66, 113–126.
- Gordon, J. W. (1987). The perceptual attack time of musical tones. *J. Acoust. Soc. Am.* 82, 88–105.
- Goswami, U., Thompson, J., Richardson, U., Stainthorpe, R., Hughes, D., Rosen, S., and Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: a new hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10911–10916.
- Goswami, U. (2010). A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* 15, 3–10.
- Greenberg, S. (2006). “A multi-tier framework for understanding spoken language,” in *Listening to Speech: An Auditory Perspective*, eds S. Greenberg and W. A. Ainsworth (Mahwah, NJ: Erlbaum), 411–433.
- Hämäläinen, J. A., Salminen, H. K., and Leppänen, P. H. T. (in press). Basic auditory processing deficits in dyslexia: Review of the behavioural, event-related potential and magnetoencephalographic evidence. *J. Learn. Disabil.*
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Holt, L. L., and Idemaru, K. (2011). “Generalization of dimension-based statistical learning of speech,” in *Proc. 17th Intl. Cong. Phonetic Sci. (ICPhS XVII)*, Hong Kong, China.
- Huss, M., Verney, J. P., Fosker, T., Mead, N., and Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex* 47, 674–689.
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. E., and Schlaug, G. (2009). Musical training shapes structural brain development. *J. Neurosci.* 29, 3019–3025.
- Jagadeesh, B. (2006). “Attentional modulation of cortical plasticity,” in *Textbook of Neural Repair and Rehabilitation:*

- Neural Repair and Plasticity*, Vol. 1. eds M. Selzer, S. E. Clarke, L. G. Cohen, P. W. Duncan, and F. H. Gage (Cambridge: Cambridge University Press), 194–205.
- Jentschke, S., Koeslch, S., Sallat, S., and Friederici, A. (2008). Children with specific language impairment also show impairment of music-syntactic processing. *J. Cogn. Neurosci.* 20, 1940–1951.
- Kaas, J., and Hackett, T. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11793–11799.
- Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends Cogn. Sci.* 14, 131–137.
- Kilgard, M., and Merzenich, M. (1998). Cortical reorganization enabled by nucleus basalis activity. *Science* 279, 1714–1718.
- Kral, A., and Eggermont, 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.* 56, 259–269.
- Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605.
- Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res. Cogn. Brain Res.* 25, 161–168.
- Lakshminarayanan, K., and Tallal, P. (2007). Generalization of non-linguistic auditory perceptual training to syllable discrimination. *Restor. Neurol. Neurosci.* 25, 263–272.
- Lappe, C., Herholz, S. C., Trainor, L. J., and Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *J. Neurosci.* 8, 9632–9639.
- Lee, K. M., Skoe, E., Kraus, N., and Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *J. Neurosci.* 29, 5832–5840.
- Liu, F., Patel, A. D., Fourcin, A., and Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification, and imitation. *Brain* 133, 1682–1693.
- Mattys, S. L., White, L., and Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *J. Exp. Psychol. General* 134, 477–500.
- McMurray, B., Aslin, R., Tanenhaus, M., Spivey, M., and Subik, D. (2008). Gradient sensitivity to within-category variation in speech: implications for categorical perception. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1609–1631.
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* 19, 712–723.
- Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894–15898.
- Musacchia, G., Strait, D. L., and Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and nonmusicians. *Hear. Res.* 241, 34–42.
- Nagarajan, S. S., Cheung, S. W., Bedenbaugh, P., Beitel, R. E., Schreiner, C. E., and Merzenich, M. M. (2002). Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. *J. Neurophysiol.* 87, 1723–1737.
- Overy, K. (2003). Dyslexia and music: from timing deficits to musical intervention. *Ann. N. Y. Acad. Sci.* 999, 497–505.
- Palmer, C. (1997). Music performance. *Ann. Rev. Psychol.* 48, 115–138.
- Panizzon, M. W., Fennema-Notestine, C., Eyer, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., Jacobson, K., Lyons, M. J., Grant, M. D., Franz, C. E., Xian, H., Tsuang, M., Tsuang, M., Fischl, B., Seidman, L., Dale, A., and Kremen, W. S. (2009). Distinct genetic influences on cortical surface area and cortical thickness. *Cereb. Cortex* 19, 2728–2735.
- Parbery-Clark, A., Skoe, E., and Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100–14107.
- Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., and Kraus, N. (2011). Musical training and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS ONE* 6: e18082. doi: 10.1371/journal.pone.0018082
- Patel, A. D. (2008). *Music, Language, and the Brain*. New York: Oxford University Press.
- Patel, A. D., and Iversen, J. R. (2007). The linguistic benefits of musical abilities. *Trends Cogn. Sci.* 11, 369–372.
- Patel, A. D., Iversen, J. R., Wassenaar, M., and Hagoort, P. (2008). Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology* 22, 776–789.
- Patel, A. D., Xu, Y., and Wang, B. (2010). “The role of F0 variation in the intelligibility of Mandarin sentences,” in *Proc. Speech Prosody 2010*, May 11–14, 2010, Chicago, IL, USA.
- Peretz, I., and Coltheart, M. (2003). Modularity of music processing. *Nat. Neurosci.* 6, 688–691.
- Peretz, I., Cummings, S., and Dubé, M-P. (2007). The genetics of congenital amusia (or tone-deafness): a family aggregation study. *Am. J. Hum. Genet.* 81, 582–588.
- Poeppl, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun.* 41, 245–255.
- Polley, D. B., Steinberg, E. E., and Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *J. Neurosci.* 26, 4970–4982.
- Recanzone, G. H., Schreiner, C. C., and Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* 13, 87–103.
- Repp, B. H. (1992). Diversity and commonality in music performance: an analysis of timing microstructure in Schumann's “Träumerei.” *J. Acoust. Soc. Am.* 92, 2546–2568.
- Russo, N., Nicol, T., Musacchia, G., and Kraus, N. (2004). Brainstem response to speech syllables. *Clin. Neurophysiol.* 115, 2021–2030.
- Salimpoor, V., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat. Neurosci.* 14, 257–262.
- Schnupp, J., Nelken, I., and King, A. (2011). *Auditory Neuroscience: Making Sense of Sound*. Cambridge, MA: MIT Press.
- Schofield, B. R. (2010). Projections from auditory cortex to midbrain cholinergic neurons that project to the inferior colliculus. *Neuroscience* 166, 231–240.
- Skoe, E., and Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear Hear.* 31, 302–324.
- Song, J., Skoe, E., Banai, K., and Kraus, N. (2010). Perception of speech in noise: neural correlates. *J. Cogn. Neurosci.* 23, 2268–2279.
- Song, J. H., Skoe, E., Wong, P. C., and Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J. Cogn. Neurosci.* 10, 1892–1902.
- Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.
- Strait, D. L., Kraus, N., Parbery-Clark, A., and Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear. Res.* 261, 22–29.
- Strait, D. L., Skoe, E., Kraus, N., and Ashley, R. (2009). Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *Eur. J. Neurosci.* 29, 661–668.
- Strong, W., and Clark, M. C. (1967). Perturbations of synthetic orchestral wind-instrument tones. *J. Acoust. Soc. Am.* 41, 277–285.
- Suga, N. (2008). Role of corticofugal feedback in hearing. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 194, 169–183.
- Tallal, P., and Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends Neurosci.* 29, 382–390.
- Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., and Friederici, A. D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *Eur. J. Neurosci.* 30, 1636–1642.
- Thiel, C. M. (2007). Pharmacological modulation of learning-induced plasticity in human auditory cortex. *Restor. Neurol. Neurosci.* 25, 435–443.
- Thompson, J. M., and Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *J. Physiol. Paris* 102, 120–129.
- Toscano, J. C., and McMurray, B. (2010). Cue integration with categories: weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cogn. Sci.* 34, 434–464.
- Toscano, J. C., McMurray, B., Dennhardt, J., and Luck, S. J. (2010). Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychol. Sci.* 21, 1532–1540.
- Tzounopoulos, T., and Kraus, N. (2009). Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 62, 463–469.
- Vandermosten, M., Boets, B., Luts, H., Poelmans, H., Golestani, N., Wouters, J., and Ghesquière, P. (2010). Adults with dyslexia are impaired in categorizing speech and nonspeech sounds on the basis of temporal cues. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10389–10394.
- Vos, P. G., and Troost, J. M. (1989). Ascending and descending melodic

- intervals: statistical findings and their perceptual relevance. *Music Percept.* 6, 383–396.
- Weinberger, N. (2007). Auditory associative memory and representational plasticity in the primary auditory cortex. *Hear. Res.* 229, 54–68.
- Winer, J. (2006). Decoding the auditory corticofugal systems. *Hear. Res.* 212, 1–8.
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.
- Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46.
- Zatorre, R. J., and Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Phil. Trans. R. Soc. B* 363, 1087–1104.
- Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 16 March 2011; paper pending published: 05 April 2011; accepted: 12 June 2011; published online: 29 June 2011.
- Citation: Patel AD (2011) Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front. Psychology* 2:142. doi: 10.3389/fpsyg.2011.00142
- This article was submitted to *Frontiers in Auditory Cognitive Neuroscience*, a specialty of *Frontiers in Psychology*. Copyright © 2011 Patel. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.