Musical Training Enhances Automatic Encoding of Melodic Contour and Interval Structure Takako Fujioka1,4, Laurel J. Trainor3 , Bernhard Ross1,2, Ryusuke Kakigi4 , and Christo Pantev1,2 Abstract & In music, melodic information is thought to be encoded in two forms, a contour code (up/down pattern of pitch changes) and an interval code (pitch distances between successive notes). A recent study recording the mismatch negativity (MMN) evoked by pitch contour and interval deviations in simple melodies demonstrated that people with no formal music education process both contour and interval information in the auditory cortex automatically. However, it is still unclear whether musical experience enhances both strategies of melodic encoding. We designed stimuli to examine contour and interval information separately. In the contour condition there were eight different standard melodies (presented on 80% of trials), each consisting of five notes all ascending in pitch, and the corresponding deviant melodies (20%) were altered to descending on their final note. The interval condition used one five-note standard melody transposed to eight keys from trial to trial, and on deviant trials the last note was raised by one whole tone without changing the pitch contour. There was also a control condition, in which a standard tone (990.7 Hz) and a deviant tone (1111.0 Hz) were presented. The magnetic counterpart of the MMN (MMNm) from musicians and nonmusicians was obtained as the difference between the dipole moment in response to the standard and deviant trials recorded by magnetoencephalography. Significantly larger MMNm was present in musicians in both contour and interval conditions than in nonmusicians, whereas MMNm in the control condition was similar for both groups. The interval MMNm was larger than the contour MMNm in musicians. No hemispheric difference was found in either group. The results suggest that musical training enhances the ability to automatically register abstract changes in the relative pitch structure of melodies. & INTRODUCTION Musical ability is an essential component of human nature and there is no known human society past or present without music (Wallin, Merker, & Brown, 2000). Across different societies, musical structure has two aspects, involving time (rhythm) and pitch (Deutsch, 1999), although the particular instantiations differ from one musical system to another. From a perceptual point of view, sequences of tones, or melodies, have two aspects, a contour and an interval code (Dowling, 1978, 1982). The contour representation consists of information about the up and down pattern of pitch changes, regardless of their exact size, and is common to both speech prosody and musical melody (Patel, Peretz, Tramo, & Labreque, 1998). The interval representation consists of the exact ratio of pitch between successive tones and is specific to music, forming the basis from which scales and harmony can emerge. Behavioral studies have provided evidence that contour is more fundamental than interval in that both infants and musically untrained adults are able to process contour information but have difficulties encoding the intervals of unfamiliar melodies (Trehub, Trainor, & Unyk, 1993; Bartlett & Dowling, 1980; Dowling, 1978; Cuddy & Cohen, 1976). The pitch intervals of unfamiliar melodies are better perceived by trained than untrained musicians (Trainor, Desjardins, & Rockel, 1999; Peretz & Babaı¨, 1992). How is melodic information processed, stored, and retrieved in the brain, and what is the effect of musical training? Several behavioral studies support the idea that plastic changes in the brain caused by years of training enable superior musical perception and performance. For example, it has been reported that musicians show a left hemispheric dominance in recognizing melodic sequences while nonmusicians show a right hemispheric dominance (Bever & Chiarello, 1974). As well, musicians demonstrate better performance in the left hemisphere than the right hemisphere when they are asked to recognize a part of a melodic sequence (Peretz & Babaı¨, 1992). Recent neuroimaging investigations have begun to identify brain structures relevant for music processing (e.g., Tillmann, Janata, & Bharucha, 2003; Janata et al., 1 The Rotman Research Institute, Baycrest Centre for Geriatric Care, 2 University of Mu¨nster, 3 McMaster University, 4 National Institute for Physiological Sciences

Institute of Technology Journal of Cognitive Neuroscience 16:6, pp. 1010–1021 2002; Koelsch et al., 2002; Ohnishi et al., 2001; Satoh, Takeda, Nagata, Hatazawa, & Kuzuhara, 2001; Halpern & Zatorre, 1999; Platel et al., 1997; Zatorre, Evans, & Meyer, 1994). For example, a functional magnetic resonance imaging (fMRI) study has shown that listening to music produces an enhanced activation in the planum temporale and the left posterior dorsolateral prefrontal cortex in musicians, in comparison to pure tones (Ohnishi et al., 2001). Long-term musical training has both anatomical and functional consequences. For example, anatomical asymmetries in the auditory cortices of musicians were reported using high-resolution MRI. The left planum temporale was larger than the right in skilled musicians, especially in those with absolute pitch (the ability to recognize and name the pitch of a musical tone without reference to a comparison tone) (Schlaug, Ja¨ncke, Huang, & Steinmetz, 1995). Furthermore, the specific anterior subregion of the corpus callosum was found to be larger in musicians who commenced musical training before the age of seven (Schlaug, Ja¨ncke, Huang, Staiger, & Steinmetz, 1995). Functional differences in musicians have also been demonstrated, mainly in the sensorimotor and in the auditory systems. Violinists practice rapid independent movement of the fingers of their left hand on the fingerboard of their instrument every day for many years. Somatosensory-evoked magnetic fields reveal the results of this experience in larger cortical representations of the left-hand fingers of musicians in comparison to either the right-hand fingers of musicians or the fingers of control subjects who never played a violin (Elbert, Pantev, Wienbruch, Rockstroh, & Taub, 1995). In contrast, behavioral and functional MRI studies on pianists (Ja¨ncke, Shah, & Peters, 2000; Ja¨ncke, Schlaug, & Steinmetz, 1997) have found less asymmetrical activity in the motor cortices of musicians than in those of control subjects, since keyboard performance requires similar motor control skills for both hands. Furthermore, transcranial magnetic stimulation (TMS) data demonstrated that bimanual motor activities in keyboard musicians were less inhibited than in normal subjects (Ridding, Brouwer, & Nordstrom, 2000), as is predicted from the observation of an enlarged corpus callosum in musicians (Schlaug, Ja¨ncke, Huang, Staiger, et al., 1995). Musicians and nonmusicians also differ in their evoked responses from auditory cortex. A magnetoencephalography (MEG) study reveals that musicians show larger amplitude auditory evoked N1m responses (peaking at about 100 msec after stimulus onset) for piano sounds (Pantev et al., 1998) and an electroencephalography (EEG) study reveals that they show larger P2 and N1c responses (Shahin, Bosnyak, Trainor, & Roberts, 2003). Furthermore, the effect of N1m enhancement is specific to the instrument of training (Pantev, Roberts, Schulz, Engelien, & Ross, 2001). Event related potential (ERP) studies have also demonstrated differences between musicians and nonmusicians in later processing that is likely outside primary or secondary sensory areas. Besson, Faı¨ta, & Requin (1994) and Trainor et al. (1999) report enhanced late positive waves between 300 and 600 msec. Relations between functional and anatomical measures are also emerging. Schneider et al. (2002) combined structural information from MRI and functional information from MEG to reveal a correlation between Heschl's gyrus enlargement and enhanced evoked magnetic field response in the latency of 19–30 msec in musicians. In summary, the comparison of brain responses between skilled musicians and naive subjects can give us new insights into brain plasticity associated with musical training. The ability to encode an acoustical context is reflected electrophysiologically in the mismatch negativity (MMN) component of ERPs. The MMN and its magnetic counterpart, the MMNm, are elicited within about 100 to 200 msec after any discriminable change that occurs infrequently in a repeatedly presented auditory stimuli, even when the stimulus is not attended (Picton et al., 2000; Na¨a¨ta¨nen & Picton, 1987; Na¨a¨ta¨nen, 1992). The MMN reflects processing in neural memory traces by which the auditory cortex handles representations of the recent acoustic past and its repetitive aspects. The sources of the MMN have been located mainly in

the supratemporal plane (Alho, 1995) by dipole modeling (Scherg, Vajsar, & Picton, 1989) and scalp current density (SCD) maps obtained from EEG (Giard, Perrin, Pernier, & Bouchet, 1990) and MEG (Leva¨nen, Ahonen, Hari, McEvoy, & Sams, 1996). The contribution of frontal generators of MMN has been also suggested not only by SCD (Giard et al., 1990) but also by lesion studies (Alain, Woods, & Knight, 1998; Alho, Woods, Algazi, Knight, & Na¨a¨ta¨nen, 1994). MMN is elicited not only for changes in single acoustic features, but also for more complex and abstract features. Recent studies show that MMN can be obtained even by changes in an auditory pattern (Alain, Cortese, & Picton, 1999; Alain, Woods, & Ogawa, 1994) and a change from ascending to descending tone pairs (Saarinen, Paavilainen, Schro¨ger, Tervaniemi, & Na¨a¨ta¨nen, 1992). Furthermore, MMN is affected by experience such as phoneme categorization in a particular language (Phillips et al., 2000; Cheour et al., 1998; Na¨a¨ta¨nen et al., 1997). Even after a short time of intense listening training, the increase in MMN amplitude parallels the increased discrimination performance with tone frequency discrimination (Menning, Roberts, & Pantev, 2000) or with foreign phoneme categories (Menning, Imaizumi, Zwitserlood, & Pantev, 2002). These studies indicate that the conscious process of discrimination and the unconscious process of extracting changes in memory traces interact and that these two processes are affected by training. The goal of the present study is to investigate the relationship between long-term musical training and automatic melodic processing. Changes in both pitch contour (Trainor, McDonald, & Alain, 2002; Tervaniemi, Rytkonen, Schro¨ger, Ilmoniemi, & Na¨a¨ta¨nen, 2001; PaaFujioka et al. 1011 vilainen, Jaramillo, & Na¨a¨ta¨nen, 1998; Tervaniemi, Maury, & Na¨a¨ta¨nen, 1994; Saarinen et al., 1992) and pitch interval (Trainor et al., 2002) can evoke MMN responses from nonmusicians even when the pitch level are changed from trial to trial. However, no study to date has assessed musical training effects separately in contour and interval encoding. We investigated the MMNm responses to contour and interval changes in both musicians and nonmusicians. The stimuli were designed to clearly separate contour and the interval encoding. Factors in the musical context other than contour and interval, such as out-of-key changes, familiarity of melody, and the range of pitch leaps, were carefully controlled. A control condition examining frequency deviations to single tones was also included. RESULTS Clear auditory evoked magnetic fields (AEFs) were obtained from both musicians and nonmusicians in all stimulus conditions. The grand-averaged dipole moment waveforms for both groups are shown in Figure 1. P1m– N1m–P2m responses from the first to the fourth note were observed in both melodic conditions and in both groups. Those waveforms showed similar slow baseline shifts over the duration of the stimulus in both groups. Despite the smaller signal- to-noise ratio in the deviants (smaller number of trials) than standards, highly reproducible response patterns were obtained in both melodic conditions. After the onset of the fifth note, musicians showed clear MMNm responses in both hemispheres for both contour and interval conditions. In contrast, nonmusicians showed unclear responses in both contour and interval conditions. In the single tone control condition, on the other hand, both groups showed clear MMNm responses to the frequency change of the stimulus. The magnified MMNm waveforms after onset of the deviation are shown in Figure 2 with the mean of 95% confidence limits calculated using bootstrap resampling from every data point (shown as horizontal lines around zero from 50 to 250 msec after onset of deviation). This allows us to identify the MMNm response significantly different from zero as the parts lying outside the limits. According to the analysis, musicians showed highly significant MMNm responses for the deviation of melodies at 100 to 200 msec after onset. In contrast, the MMNm responses in nonmusicians reached significance only in the right hemisphere for the interval condition. For both groups, the MMN evoked by single tone frequency deviation was highly significant between 95 and 200 msec after stimulus onset. The magnitude of MMNm for the two melodic conditions in the musician group was compared using the same analysis

method. As shown in Figure 3, musicians showed significantly larger MMNm in the interval than in the contour condition. The data in nonmusicians was not compared, since the contour MMNm was not present significantly above baseline levels. The MMNm peaks were slightly later in the interval condition compared to the contour condition, as displayed in Figure 2. However, individual response varied widely in morphology (i.e., single or double peak). This prevented an unambiguous identification of the MMNm peak latency in single subjects, which is required for statistics on the latency differences between the stimulus conditions. The amplitudes of MMNm calculated around the peak latency of the grand-averaged waveforms are shown in Table 1. The MMNm was significantly larger in musicians than nonmusicians according to the analysis of variance (ANOVA), $F(1,22) = 12.787$, $p < .01$. MMNm was also significantly different across conditions (contour, interval, control), $F(2,44) = 16.552$, $p < .0001$, and there was a significant interaction between group and condition, $F(2,44) = 4.02$, $p < .05$. In musicians, the MMNm was significantly larger in the control and the interval than in the contour condition ($p < .05$). The larger amplitude in the interval condition compared to the contour condition can also be seen in Figure 3. In nonmusicians, the MMNm amplitude was larger in control than both contour ($p < .0001$) and interval conditions ($p < .01$) as also clearly seen in the waveforms depicted in Figure 2. The factor hemisphere was not significant in any conditions for either group, nor did hemisphere interact with any other factors. The results of the behavioral tests are reported in Table 2. The performance of musicians was significantly better than that of nonmusicians according to the ANOVA, $F(1,22) = 23.865$, $p < .0001$. Musicians exhibited good performance in both contour (96.50%) and interval (95.83%) conditions and they did not show significant differences between tasks. Nonmusicians performed at 63.0% in the interval task, which was above the chance levels of 50%, $t(11) = 3.564$, $p < .01$. However, their performance of 86.17% in the contour task was significantly better, $t(11) = 5.946$, $p < .0001$, than in the interval task. DISCUSSION In the present study, musicians showed significantly larger MMNm responses than nonmusicians to deviations in melodic contour and interval structure, whereas both groups showed similar MMNm responses to frequency deviation in a single pure tone. As well, both groups tended to show larger MMNm responses to interval than to contour changes. The results strongly support the hypothesis that musical experience leads to specific changes in the neural mechanisms for processing of abstract, but not sensory, melodic information. The contour and interval processing must be performed at the level of melodic patterns because the overall pitch levels shifts from trial to trial. Because MMNm for simple 1012 Journal of Cognitive Neuroscience Volume 16, Number 6 frequency deviation did not differ between the two groups, whereas MMNm to contour and interval changes did, we can conclude that musical training mainly affects the pitch contour and interval relations between tones rather than the encoding of single tones. These results extend a number of recent studies showing that MMN reflects not only the encoding of simple sensory features, but also the encoding of abstract rules and patterns in an auditory context (Trainor et al., 2002; Alain, Achim, & Woods, 1999; Alain, Cortese, et al., 1999; Paavilainen et al., 1998; Alain et al., 1994; Saarinen et al., 1992), by showing differential effects of training on MMNm responses to different types of MMN. The results are also in line with observations of more pronounced MMN responses in musicians to the deviation of musicrelated stimuli, such as harmonic chord progressions Figure 1. Grand averages of the source space waveforms from both musicians and nonmusicians for each condition (contour, interval, and frequency—single tone) in both left and right hemispheres. The y-axis shows the dipole moment (positive upward), and the x-axis shows the time related to the stimulus onset. The standard averaged data, the deviant averaged data, and the difference between the deviant and the standard (MMNm response) are shown. The deviation of stimuli of both contour and interval conditions occur at the fifth note. Closed arrows indicated the clearly identifiable

MMNm responses, whereas open arrows indicated the vague responses. Fujioka et al. 1013 (Koelsch, Schro¨ger, & Tervaniemi, 1999), pitch sequences within a tonal structure (Brattico, Na¨a¨ta¨nen, & Tervaniemi, 2002), and complex temporal patterns of tones (Tervaniemi, Ilvonen, Karma, Alho, & Na¨a¨ta¨nen, 1997). The results also extend those of Tervaniemi et al. (2001) by showing that the enhancement in musicians is not restricted to contour processing, but also applies to interval processing without contour change. On the other hand, the almost absent MMNm in nonmusicians in the present study is different from the previous observations of Trainor et al. (2002) of clear MMN to both contour and interval changes. There are likely two factors contributing to this difference. First the behavioral performance of nonmusicians in our study was much lower than that of nonmusicians in Trainor et al.'s study, as we modified their contour and interval stimuli. Specifically, we matched the size of pitch deviation across conditions, with the result that the contour melodies in our study had a smaller pitch range and therefore a smaller size of deviation for contour changes. The present interval melody is also more complicated than that of Trainor et al. in two important musical features. First, their standard melody, consisting of the first five notes of an ascending diatonic scale, was highly familiar and had a simple pitch contour, whereas Figure 3. Comparison between interval and contour condition in musician's MMNm response. The difference waveform was obtained by subtraction of response for contour condition from interval condition combined across both hemispheres. Figure 2. The source space waveforms of MMNm. For both contour and interval condition the time scale on the x-axis refers to the onset of the fifth note. The thick line represents the MMNm response and thin lines above and below the zero line show the upper and lower limit of 95% confidence interval. The mean of confidence interval in the whole time series was subtracted from the response to adjust the baseline of MMNm waveform to the zero line. 1014 Journal of Cognitive Neuroscience Volume 16, Number 6 the melody of the present study was unfamiliar and contained a more complex melodic contour. As well, the interval changes in Trainor et al.'s study contained outof-key notes, whereas ours did not. Nai¨ve subjects more easily detect changes in a familiar melody than in an unfamiliar one, especially for nondiatonic changes (Besson & Fai¨ta, 1995; Besson et al., 1994). Thus, the task difficulty was certainly a contributing factor to the small amplitude MMN seen in the present study. A second major difference between studies was that we used MEG and analyzed data as represented in a negative oriented dipole source, whereas Trainor et al. used EEG. MEG is less sensitive than EEG in detecting radial oriented source current signals, which in the present case could be generated from multiple sources of MMN including frontal activation (Opitz, Rinne, Mecklinger, Von Cramon, & Schro¨ger, 2002; Rinne, Alho, Ilmoniemi, Virtanen, & Na¨a¨ta¨nen, 2000; Alho, 1995; Giard et al., 1990). Interestingly, MMNm responses were larger in the interval condition compared to the contour condition, despite the fact that we controlled for interval size and range across the conditions. Thus, we have no evidence that contour is a more fundamental process, or that it is neurally privileged, even the musically untrained. There are two possible explanations for the larger MMN to interval changes. The first concerns the relation between contour and interval information. Note that the melodic interval between successive notes essentially includes the up-and-down contour information as well. In other words, contour processing could be a part of the interval processing, even though it is hard to tell whether one of the processes comes first or both run in parallel. In the interval task, the same melody is presented transposed to different pitch levels on each trial. However, in the contour task, different intervals are present on each trial. If the auditory cortex automatically extracts interval information, and this information is irrelevant in the contour task, the presence of the constantly varying intervals could obscure the automatic response to the change in contour. In this case, the regularities of the standard melodies in the interval task could be extracted more easily than those of the contour melodies for which the various interval sizes need to be

ignored. If correct, this suggests that contour information is extracted after or concurrently with interval information, but not before it. A second explanation concerns the fact that the changes in the present interval stimuli are accompanied by a change in tonality. That is, the terminal note of interval standard melody is the fifth note of the scale and functions strongly as the dominant in the key. Thus, it is possible that the MMNm was larger for interval than contour changes because not only was the interval processed, but also a difference in tonality was detected. An explanation of the larger MMNm for interval than for contour changes must also account for the mismatch between the MMNm elicited and behavioral performance. The behavioral performance of musicians was almost perfect in both conditions. On the other hand, behavioral performance in nonmusicians was much better for contour than for interval changes, although MMNm was only significantly present for interval changes, and then only in the right hemisphere. Generally, MMN amplitude corresponds to behavioral accuracy (Winkler et al., 1999; Kraus, McGee, Carrell, & Sharma, Table 1. Mean of Dipole Moment (± SEM) as the MMNm Amplitude of Each Hemisphere in Control, Contour and Interval Conditions Dipole Moment (nAm) Subject Group n Control Contour Interval Left hemisphere Musicians 12 9.737 ± 1.543* 7.572 ± 1.202 9.644 ± 1.197* Nonmusicians 12 9.742 ± 2.418***,** 0.395 ± 1.458 3.061 ± 1.086 Right hemisphere Musicians 12 10.709 ± 1.808* 5.773 ± 1.078 10.344 ± 2.147* Nonmusicians 12 8.123 ± 1.947***,** 0.792 ± 1.262 3.732 ± 1.052 *p < .05 (musicians: control > interval, interval > contour). **p < .01 (nonmusicians: control > interval). ***p < .0001 (nonmusicians: control > contour). Table 2. Mean % Correct Performance (± SEM) in the Behavioral Discrimination Task of Contour and Interval Conditions Subject Group n Contour Interval Musicians 12 96.50 ± 2.55 95.83 ± 1.90 Nonmusicians 12 86.17 ± 5.04* 63.00 ± 3.65 *p < .0001. Fujioka et al. 1015 1995; Tiitinen, May, Reinikainen, & Na¨a¨ta¨nen, 1994; Na¨a¨ta¨nen, Jiang, Lavikainen, Reinikainen, & Paavilainen, 1993). However, some recent studies demonstrated that MMN responses emerge even before subjects consciously achieve the discrimination tasks in sound categorization (Allen, Kraus, & Bradlow, 2000; Dalebout & Stack, 1999; Tremblay, Kraus, & McGee, 1998). The nonmusicians may have difficulty in performing the interval discrimination task behaviorally if it depends on higher cognitive and attentive processes such as categorization, memorization, and decision making, which utilize the output of the automatic MMN processes (Na¨a¨ta¨nen & Alho, 1997; Na¨a¨ta¨nen, 1992). No statistically significant laterality effect in the MMNm response was found in the present study in either group, although the MMNm was only significant for nonmusicians in the right hemisphere. Previous behavioral studies have shown evidence of discrete lateralization for contour and interval processing as measured by psychophysical performance in unilateral lesion patients and normal listeners to monaural sound (Lie´geois-Chauvel, Peretz, Babaı¨, Laguitton, & Chauvel, 1998; Peretz & Morais, 1987; Peretz, Morais, & Bertelson, 1987; Zatorre, 1985). It is possible that either the effect is too subtle to detect by our technique, or it occurs at a processing stage after the automatic processes reflected in the MMN. It should also be noted that our stimulation was binaural, as was that of Trainor et al. (2002), who also did not show clear laterality effects. Laterality effects might be seen more clearly with monaural stimulation, and future studies should address this question. The clear MMNm differences between musicians and nonmusicians observed in our study contribute to the growing literature suggesting that musical training affects a whole network of brain areas, from those involved in stimulus encoding and deviance detection to those involved in conscious evaluation of the music. For example, during passive listening, in addition to the MMN, another preattentive negative response, early right anterior negativity (ERAN, peaking at about 200– 250 msec to the violation of musical-harmony syntax), has been demonstrated (Koelsch et al., 2001; Maess, Koelsch, Gunter, & Friederici, 2001), which is also more pronounced in musicians than in nonmusicians (Koelsch, Schmidt, & Kansok, 2002). During

active discrimination tasks, tonality violation elicits larger late event-related responses in musicians than in nonmusicians, such as the P3 (Trainor et al., 1999; Janata, 1995; Cohen, Granot, Pratt, & Barneah, 1993) and a long-latency positive component (LPC) around 600 msec (Regnault, Bigand, & Besson, 2001; Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Besson & Faı¨ta, 1995; Besson et al., 1994; Levett & Martin, 1992; Besson & Macar, 1987). All these indicate that there must exist multiple parallel processing modules related to various aspects of musical structures. Our results indicate that during the early automatic stages of processing, musical training particularly affects the detection of changes at the abstract level of pitch contour and interval patterns, but has less effect on the detection of simple pitch changes.

The role of the auditory brainstem in processing musically relevant pitch

Gavin M. Bidelman1,2*
1Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
2School of Communication Sciences and Disorders, University of Memphis, Memphis, TN, USA

Neuroimaging work has shed light on the cerebral architecture involved in processing the melodic and harmonic aspects of music. Here, recent evidence is reviewed illustrating that subcortical auditory structures contribute to the early formation and processing of musically relevant pitch. Electrophysiological recordings from the human brainstem and population responses from the auditory nerve reveal that nascent features of tonal music (e.g., consonance/dissonance, pitch salience, harmonic sonority) are evident at early, subcortical levels of the auditory pathway. The salience and harmonicity of brainstem activity is strongly correlated with listeners' perceptual preferences and perceived consonance for the tonal relationships of music. Moreover, the hierarchical ordering of pitch intervals/chords described by the Western music practice and their perceptual consonance is well-predicted by the salience with which pitch combinations are encoded in subcortical auditory structures. While the neural correlates of consonance can be tuned and exaggerated with musical training, they persist even in the absence of musicianship or long-term enculturation. As such, it is posited that the structural foundations of musical pitch might result from innate processing performed by the central auditory system. A neurobiological predisposition for consonant, pleasant sounding pitch relationships may be one reason why these pitch combinations have been favored by composers and listeners for centuries. It is suggested that important perceptual dimensions of music emerge well before the auditory signal reaches cerebral cortex and prior to attentional engagement. While cortical mechanisms are no doubt critical to the perception, production, and enjoyment of music, the contribution of subcortical structures implicates a more integrated, hierarchically organized network underlying music processing within the brain.

In Western tonal music, the octave is divided into 12 equally spaced pitch classes (i.e., semitones). These elements can be further arranged into seven tone subsets to construct the diatonic major/minor scales that define tonality and musical key. Music theory and composition stipulate that the pitch combinations (i.e., intervals) formed by these scale-tones carry different weight, or importance, within a musical framework (Aldwell and Schachter, 2003). That is, pitch intervals follow a hierarchical organization in accordance with their functional role in musical composition (Krumhansl, 1990). Intervals associated with stability and finality are

regarded as *consonant* while those associated with instability (i.e., requiring resolution) are regarded as *dissonant*. Given their anchor-like function in musical contexts, it is perhaps unsurprising that consonant pitch relationships occur more frequently in tonal music than dissonant relationships (Budge, 1943; Vos and Troost, 1989). Ultimately, it is the ebb and flow between consonance and dissonance which conveys musical tension and establishes the structural foundations of melody and harmony, the fundamental building blocks of Western tonal music (Rameau, 1722/1971; Krumhansl, 1990).

The Perception of Musical Pitch: Sensory Consonance and Dissonance

The music cognition literature distinguishes the aforementioned *musical* definitions from those used to describe the *psychological* attributes of musical pitch. The term *tonal-* or *sensory-consonance-dissonance* refers to the perceptual quality of two or more simultaneous tones presented in isolation (Krumhansl, 1990) and is distinct from consonance arising from contextual or cognitive influences (see Dowling and Harwood, 1986, for a discussion of non-sensory factors). Perceptually, consonant pitch relationships are described as sounding more pleasant, euphonious, and beautiful than dissonant combinations which sound unpleasant, discordant, or rough (Plomp and Levelt, 1965). Consonance is often described parsimoniously as the absence of dissonance. A myriad of empirical studies have quantified the perceptual qualities of musical pitch relationships. In such behavioral experiments, listeners are typically played various two-tone pitch combinations (dyads) constructed from the musical scale and asked to rate their degree of consonance (i.e., "pleasantness"). Examples of such ratings, as reported in the seminal studies of Kameoka and Kuriyagawa (1969a,b), are shown in Figure 1A. The rank order of intervals according to their perceived consonance is shown in Figure 1B. Two trends emerge from the pattern of ratings across a number of studies: (i) listeners routinely prefer consonant pitch relationships (e.g., octave, fifth, fourth, etc.) to their dissonant counterparts (e.g., major/minor second, sevenths) and (ii) intervals are not heard in a strict binary manner (i.e., consonant vs. dissonant) but rather, are processed differentially based on their degree of perceptual consonance (e.g., Kameoka and Kuriyagawa, 1969a,b; Krumhansl, 1990). These behavioral studies demonstrate that musical pitch relationships are perceived *hierarchically* and in an arrangement that parallels their relative use and importance in music composition (Krumhansl, 1990; Schwartz et al., 2003).

figure 1

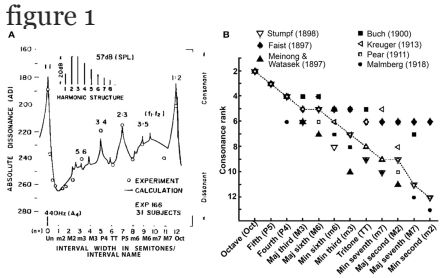

FIGURE 1. CONSONANCE RANKINGS FOR CHROMATIC SCALE TONE COMBINATIONS OF WESTERN MUSIC PRACTICE. (A) Consonance (i.e., "pleasantness") ratings reported by Kameoka and Kuriyagawa (1969b) for two-tone

intervals (dyads). Stimuli were composed of two simultaneously sounding complex tones (inset). The spacing between fundamental frequencies (*f*1, *f*2) was varied to form the various chromatic intervals within the range of an octave; the lower tone (*f*1) was always fixed at 440 Hz and the upper tone (*f*2) varied from 440 to 880 Hz in semitone spacing. Note the higher behavioral ratings for the consonant pitch relationships [e.g., 0 (Un), 7 (P5), 12 (Oct) semitones] relative to dissonant relationships [e.g., 2 (m2), 6 (TT), 11 (M7) semitones] as well as the hierarchical arrangement of intervals (Un > Oct > P5 > P4 > M6, etc). (B) Rank order of musical interval consonance ratings reported across seven psychophysical studies (Faist, 1897; Meinong and Witasek, 1897; Buch, 1900; Pear, 1911; Kreuger, 1913; Malmberg, 1918; Stumpf, 1989). Open circles represent the median consonance rank assigned to each of the 12 chromatic dyads. Figures adapted from Kameoka and Kuriyagawa (1969b) and Schwartz et al. (2003) with permission from The Acoustical Society of America and Society for Neuroscience, respectively.

Interestingly, the preference for consonance and the hierarchical nature of musical pitch perception is reported even for non-musician listeners (Van De Geer et al., 1962; Tufts et al., 2005; Bidelman and Krishnan, 2009). Thus, while the perceptual nuances of music might be augmented with experience (McDermott et al., 2010; Bidelman et al., 2011c) – or degraded with impairments (e.g., amusia: Cousineau et al., 2012) – a perceptual bias for consonant pitch combinations persists even in the absence of musical training. Indeed, this bias for consonance emerges early in life, well before an infant is exposed to the stylistic norms of culturally specific music (Trehub and Hannon, 2006). Evidence from animal studies indicates that even non-human species (e.g., sparrows and Japanese monkeys) discriminate consonant from dissonant pitch relationships (Izumi, 2000; Watanabe et al., 2005; Brooks and Cook, 2010) and some even show musical preferences similar to human listeners (e.g., Bach > Schönberg) (Sugimoto et al., 2010). These data provide convincing evidence that certain aspects of music perception might be innate, a byproduct of basic properties of the auditory system.

The current review aims to provide a comprehensive overview of recent work examining the psychophysiological bases of consonance, dissonance, and the hierarchical foundations of musical pitch. Discussions of these musical phenomena have enjoyed a rich history of arguments developed over many centuries. As such, treatments of early explanations are first provided based on mathematical, acoustic, and psychophysical accounts implicating peripheral auditory mechanisms (e.g., cochlear mechanics) in musical pitch listening. Counterexamples are then provided which suggest that strict acoustic and cochlear theories are inadequate to account for the findings of recent studies examining human consonance judgments. Lastly, recent neuroimaging evidence is highlighted which supports the notion that the perceptual attributes of musical pitch are rooted in *neurophysiological* processing performed by the central nervous system. Particular attention is paid to recent studies examining the neural encoding of musical pitch using scalp-recorded brainstem responses elicited from human listeners. Brainstem evoked potentials demonstrate that the perceptual correlates of musical consonance and pitch hierarchy are well represented in subcortical auditory structures, suggesting that attributes important to music listening emerge well before the auditory signal reaches cerebral cortex. The contribution of subcortical mechanisms implies that

music engages a more integrated, hierarchically organized network tapping both sensory (pre-attentive) and cognitive levels of brain processing.

Historical Theories and Explanations for Musical Consonance and Dissonance

The Acoustics of Musical Consonance

Early explanations of consonance and dissonance focused on the underlying acoustic properties of musical intervals. It was recognized as early as the ancient Greeks, and later by Galilei (1638/1963), that pleasant sounding (i.e., consonant) musical intervals were formed when two vibrating entities were combined whose frequencies formed simple integer ratios (e.g., 3:2 = perfect fifth, 2:1 = octave). In contrast, "harsh" or "discordant" (i.e., dissonant) intervals were created by combining tones with complex ratios (e.g., 16:15 = minor second). By these purely mathematical standards, consonant intervals were regarded as divine acoustic relationships superior to their dissonant counterparts and, as a result, were heavily exploited by early composers (for a historic account, see Tenney, 1988). Indeed, the most important pitch relationships in music, including the major chord, can be derived directly from the first few components of the harmonic series (Gill and Purves, 2009). Yet, while attractive *prima facie*, the long held theory that the ear prefers simple ratios is no longer tenable when dealing with contemporary musical tuning systems. For example, the ratio of the consonant perfect fifth under modern equal temperament (442:295) is hardly a small integer relationship. Though intimately linked, explanations of consonance-dissonance based purely on these physical constructs (e.g., frequency ratios) are, in and of themselves, insufficient in describing all of the cognitive aspects of musical pitch (Cook and Fujisawa, 2006; Bidelman and Krishnan, 2009). Indeed, it is possible for an interval to be esthetically dissonant while mathematically consonant, or vice versa (Cazden, 1958, p. 205). For example, tones combined at simple ratios (traditionally considered consonant), can be judged to be dissonant when their frequency components are stretched (i.e., made inharmonic) from their usual position in the harmonic series (Slaymaker, 1970) or when occurring in an unexpected musical context (Dowling and Harwood, 1986). These experimental paradigms cleverly disentangle stimulus acoustics (e.g., frequency ratios) from behavioral consonance judgments and, in doing so, indicate that pure acoustic explanations are largely inadequate as a sole basis of musical consonance.

Psychophysiology of Musical Consonance

Psychophysical roughness/beating and the cochlear critical band

Helmholtz (1877/1954) offered some of the earliest psychophysical explanations for sensory consonance-dissonance. He observed that when adjacent harmonics in complex tones interfere they create the perception of "roughness" or "beating," percepts closely related to the perceived dissonance of tones (Terhardt, 1974). Consonance, on the other hand, occurs in the absence of beating, when low-order harmonics are spaced sufficiently far apart so as not to interact. Empirical studies suggest this phenomenon is related to cochlear mechanics and the critical-band

hypothesis (Plomp and Levelt, 1965). This theory postulates that the overall consonance-dissonance of a musical interval depends on the total interaction of frequency components within single auditory filters. Pitches of consonant dyads have fewer partials which pass through the same critical bands and therefore, yield more pleasant percepts; in contrast, the partials of dissonant intervals compete within individual channels and as such, yield discordant percepts.

Unfortunately, roughness/beating is often difficult to isolate from consonance percepts given that both covary with the spacing between frequency components in the acoustic waveform, and are thus, intrinsically coupled. While within-channel interactions may produce some amount of dissonance, modern empirical evidence indicates that beating/roughness plays only a minor role in its perception. Indeed, at least three pieces of evidence support the notion that consonance may not be mediated by roughness/beating, *per se*. First, psychoacoustic findings indicate that roughness percepts are dominated by lower modulation rates (~30–150 Hz) (Terhardt, 1974; McKinney et al., 2001, p. 2). Yet, highly dissonant intervals are heard for tones spaced well beyond this range (Bidelman and Krishnan, 2009; McDermott et al., 2010). Second, dichotic listening tasks can been used to eliminate the monaural interactions necessary for roughness and beating. In these experiments, the constituent notes of a musical interval are separated between the ears. Dichotic listening ensures that roughness/beating along the cochlear partition is eliminated, as each ear processes a perfectly periodic, singular tone. Nevertheless, dichotic presentation does not alter human consonance judgments (Houtsma and Goldstein, 1972; Bidelman and Krishnan, 2009; McDermott et al., 2010), indicating that cochlear interactions (and the critical band) are insufficient explanations for explaining consonance/dissonance percepts. Lastly, lesion studies indicate a dissociation between roughness and the perception of dissonance as one percept can be selectively impaired independently of the other (Tramo et al., 2001). Taken together, converging evidence suggests that roughness/beating may not be as important a factor in sensory consonance-dissonance as conventionally thought (e.g., Helmholtz, 1877/1954; Plomp and Levelt, 1965; Terhardt, 1974).

Tonal fusion and harmonicity

Alternate theories have suggested musical consonance is determined by the sense of "fusion" or "tonal affinity" between simultaneously sounding pitches (Stumpf, 1890). Pitch fusion describes the degree to which multiple pitches are heard as a single, unitary tone (DeWitt and Crowder, 1987). Fusion is closely related to harmonicity, which describes how well a sound's acoustic spectrum agrees with a single harmonic series (Gill and Purves, 2009; McDermott et al., 2010; Bidelman and Heinz, 2011). Pitch relationships with more coinciding partials have spectra that are more harmonic (e.g., octave, perfect fifth). As a result, they are heard as being fused which consequently creates the sensation of consonance. In contrast, pitch relationships which are more inharmonic (e.g., minor second, tritone) have spectra which diverge from a single harmonic series, are less fused perceptually, and create the quality of dissonance. Under this hypothesis then, the auditory system formulates consonance based on the harmonicity of sound. Support for the fusion/harmonicity premise stems from experiments examining inharmonic tone complexes, which show that consonance is obtained when tones share coincident partials, even when other factors known to influence consonance are varied, e.g., the ratio of note fundamental frequencies or roughness/beating (Slaymaker, 1970; Bidelman and Krishnan,

2009; McDermott et al., 2010; Bidelman and Heinz, 2011). For example, even a complex ratio (typically associated with dissonance) can be heard as consonant if it fits into the template of a single complex tone. Recent behavioral work supports the dominance of harmonicity in musical pitch percepts: consonance preferences are strongly correlated with a preference for harmonicity but not, for example, a preference for lack of roughness (McDermott et al., 2010).

Neurophysiology of musical consonance

The fact that these perceptual factors do not depend on long-term enculturation or musical training and have been reported even in non-human species (Izumi, 2000; Watanabe et al., 2005; Brooks and Cook, 2010; Sugimoto et al., 2010) suggests that the basis of musical consonance and pitch hierarchy might be rooted in the fundamental processing and/or constraints of the auditory system (Trehub and Hannon, 2006). In particular, the similarity in percepts under dichotic listening indicates that consonance must be computed centrally by deriving information from the combined signals relayed from both cochleae (Houtsma and Goldstein, 1972; Bidelman and Krishnan, 2009). Indeed, converging evidence suggests that these properties of musical pitch may be reflected in intrinsic, temporal firing patterns, and synchronization of auditory neurons (Boomsliter and Creel, 1961; Ebeling, 2008). Having ruled out pure mathematical, acoustical, and cochlear explanations, neurophysiological studies will now be examined which suggest a neural basis of musical consonance, dissonance, and tonal hierarchy.

Neural Correlates of Consonance, Dissonance, and Musical Pitch Hierarchy

Neuroimaging methods have offered a window into the cerebral architecture underlying the perceptual attributes of musical pitch. Functional magnetic resonance imaging (fMRI), for example, has shown differential and enhanced activation across cortical regions (e.g., inferior/middle frontal gyri, premototor cortices, interior parietal lobule) when processing consonant vs. dissonant tonal relationships (Foss et al., 2007; Minati et al., 2009; Fujisawa and Cook, 2011). Scalp-recorded event-related brain potentials (ERPs) have proved to be a particularly useful technique to non-invasively probe the neural correlates of musical pitch. ERPs represent the time-locked neuroelectric activity of the brain generated by the activation of neuronal ensembles within cerebral cortex. The auditory cortical ERP consists of a series of voltage deflections (i.e., "waves") within the first ~250 ms after the onset of sound. Each deflection represents the subsequent activation in a series of early auditory cortical structures including thalamus and primary/secondary auditory cortex (Näätänen and Picton, 1987; Scherg et al., 1989; Picton et al., 1999). The millisecond temporal resolution of ERPs provides an ideal means to investigate the time-course of music processing within the brain not afforded by other, more sluggish neuroimaging methodologies (e.g., fMRI).

Cortical Correlates of Musical Consonance

Using far-field recorded ERPs, neural correlates of consonance, dissonance, and musical scale pitch hierarchy have been identified at a cortical level of processing (Brattico et al., 2006; Krohn et al., 2007; Itoh et al., 2010). Cortical evoked responses

elicited by musical intervals, as reported by (Itoh et al., 2010), are shown in Figure 2. In this experiment, listeners were played a random sequence of dyadic intervals (0–13 semitones) in a passive listening task while ERPs were recorded at the scalp. The use of pure tones ensured minimal roughness at the auditory periphery. Modulations in cortical activity were observed in the prominent waves of the ERP but were especially apparent in the later endogenous P2-N2 complex at a latency of ~200–300 ms (Figure 2A). Indeed, N2 magnitude varied with the dyad's degree of consonance; intervals established in previous studies as dissonant – those which are unpleasant to the ear – elicited larger N2 responses than the more pleasant sounding, consonant pitch intervals (Figure 2B). Importantly, these effects were observed even when the interval's separation exceeded the critical bandwidth (~3 semitones) suggesting that consonance, and its neural underpinnings, were computed based on properties other than roughness. Further examination revealed that N2 magnitude also corresponded with a measure of the intervals' "ratio simplicity" (Schellenberg and Trehub, 1994), defined as $1/\log(X + Y)$ for the ratio $X:Y$ (Figure 2C). These results demonstrate that (i) cortical activity distinguishes pitch relationships according to their consonance and in a manner consistent with standard musical practice and (ii) the central auditory system exploits the harmonicity of sound to code the perceptual pleasantness of music. These studies clearly demonstrate that *cortical activity* is especially sensitive to the pitch relationships found in music. Yet, a natural question that emerges is whether these neural correlates emerge prior to the auditory cortices, e.g., at *subcortical* stages of auditory processing.

figure 2

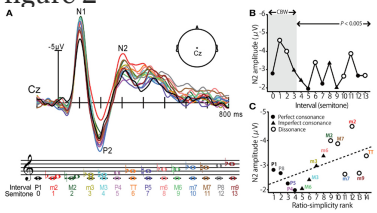

FIGURE 2. CORTICAL EVENT-RELATED POTENTIALS (ERPS) ELICITED BY MUSICAL DYADS. (A) Cortical ERP waveforms recorded at the vertex of the scalp (Cz lead) in response to chromatic musical intervals. Response trace color corresponds to the evoking stimulus denoted in music notation. Interval stimuli were composed of two simultaneously sounding pure tones. (B) Cortical N2 response magnitude is modulated by the degree of consonance; dissonant pitch relationships evoke larger N2 magnitude than consonant intervals. The shaded region demarcates the critical bandwidth (CBW); perceived dissonance created by intervals larger than the CBW cannot be attributed to cochlear interactions (e.g., beating between frequency components). Perfect consonant intervals (filled circles); imperfect consonant intervals (filled triangles); dissonant intervals (open circles) (C) Response magnitude is correlated with the degree of simplicity of musical pitch intervals; simpler, more consonant pitch relationships (e.g., P1, P8, P5) elicit smaller N2 than more complex, dissonant pitch relationships (e.g., M2, TT, M7). Figure adapted from Itoh et al. (2010) with permission from The Acoustical Society of America.

Brainstem Correlates of Musical Consonance and Scale Pitch Hierarchy

To assess human subcortical auditory processing, electrophysiological studies have utilized the frequency-following responses (FFRs). The FFR is a sustained evoked

potential characterized by a periodic waveform which follows the individual cycles of the stimulus (for review, see Krishnan, 2007; Chandrasekaran and Kraus, 2010; Skoe and Kraus, 2010). Based on its latency (Smith et al., 1975), lesion data (Smith et al., 1975; Sohmer et al., 1977), and known extent of phase-locking in the brainstem (Wallace et al., 2000; Aiken and Picton, 2008; Alkhoun et al., 2008), a number of studies recognize the inferior colliculus (IC) of the midbrain as the primary generator of the FFR. Employing this response, recent work from our lab has explored the neural encoding of musical pitch-relevant information at the level of the brainstem.

In a recent study (Bidelman and Krishnan, 2009) recorded FFRs elicited by nine musical dyads that varied in their degree of consonance and dissonance. Dichotic stimulus presentation ensured that peripheral roughness/beating was minimized and that consonance percepts were computed centrally after binaural integration (Houtsma and Goldstein, 1972). In addition, only non-musicians were recruited to ensure participants had no explicit exposure to the rules of musical theory, a potential bias, or knowledge of learned labels for musical pitch relationships. Exemplar FFRs and response spectra evoked by a subset of the dyads are shown in Figure 3. From brainstem responses, a measure of "neural pitch salience" was computed using a harmonic sieve analysis (Cedolin and Delgutte, 2005) to quantify the harmonicity of the neural activity (see Bidelman and Krishnan, 2009 for details). Essentially, this algorithm is a time-domain analog of the classic pattern recognition model of pitch whereby a "central pitch processor" matches harmonic information contained in the response to an internal template in order to compute the heard pitch (Goldstein, 1973; Terhardt et al., 1982). Results showed that brainstem responses to consonant intervals were more robust and yielded stronger neural pitch salience than those to dissonant intervals. In addition, the ordering of neural salience across musical intervals followed the hierarchical arrangement of pitch stipulated by Western music theory (Rameau, 1722/1971; Krumhansl, 1990). Lastly, neural pitch salience was well-correlated with listeners' behavioral consonance ratings (Figure 3C). That is, musical preferences could be predicted based on an individual's underlying subcortical response activity. Subsequent studies showed that brainstem encoding could similarly predict the sonority ratings of more complex musical pitch relationships including the four most common triadic chords in music (Bidelman and Krishnan, 2011). Together, results suggest that in addition to cortical processing (e.g., Itoh et al., 2010), *subcortical* neural mechanisms (i) show preferential encoding of consonant musical relationships and (ii) preserve and predict the hierarchical arrangement of pitch as described in music practice and in psychophysical studies.

figure 3

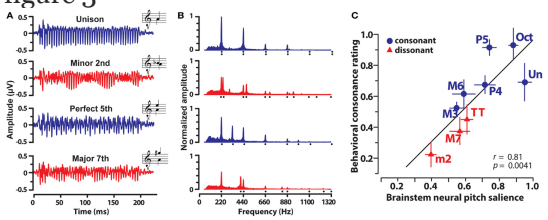

FIGURE 3. HUMAN BRAINSTEM FREQUENCY-FOLLOWING RESPONSES (FFRS) ELICITED BY MUSICAL DYADS. Grand average FFR waveforms (A) and their corresponding frequency spectra (B) evoked by the dichotic presentation of four representative musical intervals. Consonant intervals, blue; dissonant intervals, red. (A) Clearer, more robust periodicity is observed for consonant relative to dissonant intervals. (B) Frequency spectra reveal that FFRs faithfully preserve the

harmonic constituents of *both* musical notes of the interval (compare response spectrum, filled area, to stimulus spectrum, harmonic locations denoted by dots). Consonant intervals evoked more robust spectral magnitudes across harmonics than dissonant intervals. Amplitudes are normalized relative to the unison. (C) Correspondence between FFR pitch salience computed from brainstem responses and behavior consonance ratings. Neural responses well predict human preferences for musical intervals. Note the systematic clustering of consonant and dissonant intervals and the maximal separation of the unison (most consonant interval) from the minor second (most dissonant interval) in the neural-behavioral space. Data from Bidelman and Krishnan (2009).

Importantly, these strong brain-behavior relationships have been observed in non-musician listeners and under conditions of passive listening (most subjects fell asleep during EEG testing). These factors imply that basic perceptual aspects of music might be rooted in intrinsic sensory processing. Unfortunately, these brainstem studies employed adult human listeners. As such, they could not rule out the possibility that non-musicians' brain responses might have been preferentially tuned via long-term enculturation and/or implicit exposure to the norms of Western music practice.

Auditory Nerve Correlates of Musical Consonance

To circumvent confounds of musical experience, enculturation, memory, and other top-down factors which influence the neural code for music, Bidelman and Heinz (2011) investigated whether the correlates of consonance were present at very initial stages of the auditory pathway. Auditory nerve (AN) fiber responses were simulated using a computational model of the auditory periphery (Zilany et al., 2009). This model – originally used to describe AN response properties in the cat – incorporates many of the most important properties observed in the peripheral auditory system including, cochlear filtering, level-dependent gain (i.e., compression) and bandwidth control, as well as two-tone suppression. Details of this phenomenological model are beyond the scope of the present review. Essentially, the model accepts a sound input (e.g., musical interval) and outputs a realistic train of action potentials (i.e., spikes) that accurately simulates the discharge pattern of single AN neurons as recorded in animal studies (Zilany and Bruce, 2006). Actual neurophysiological experiments are often plagued by limited recording time, stimuli, and small sample sizes so their conclusions are often restricted. Modeling thus allowed for the examination of (i) possible differential AN encoding across a large continuum (i.e., 100s) of musical and non-musical pitch intervals and (ii) activation across an array of nerve fibers spanning the entire cochlear partition.

Auditory nerve population responses were obtained by pooling single-unit responses from 70 fibers with characteristic frequencies spanning the range of human hearing. Spike trains were recorded in response to 220 dyads within the range of an octave where $f1/f2$ separation varied from the unison (i.e., $f2 = f1$) to the octave (i.e., $f2 = 2f1$). First-order interspike interval histograms computed from raw spike times allowed for the quantification of periodicity information contained in the aggregate AN response (Figure 4A). Adopting techniques of (Bidelman and Krishnan, 2009), harmonic sieve analysis was used to extract the salience of pitch-related information

encoded in the entire AN ensemble. Neural pitch salience profiles elicited by exemplar consonant (P5) and dissonant (m2) musical dyads are shown in Figure 4B. The maximum of each profile provided a singular estimate of the neural salience for each dyad stimulus. Interestingly, rank order of the chromatic intervals according to this salience magnitude followed a predictable pattern; consonant intervals – those judged more pleasant sounding by listeners – yielded higher neural rankings than dissonant intervals (e.g., M7, TT, m2) (Figure 4C). Additionally, although neural rank ordering was derived from responses at the level of AN, they showed close agreement to rankings stipulated by Western music theory as well as those obtained from human listeners in psychophysical studies (e.g., Figure 1). As with human brainstem FFRs, AN responses were well-correlated with perceptual judgments of consonance (Figure 4D). That is, the hierarchical perception and perceived pleasantness of musical stimuli could be well-predicted based on neural responses at the level of AN. Our earlier findings from human brainstem ERPs suggested that such preferences might emerge based on subcortical neurocomputations well before cerebral cortex. Our AN modeling studies extend these results, and further suggest they might even be rooted in the most peripheral sites of the auditory brain.

figure 4

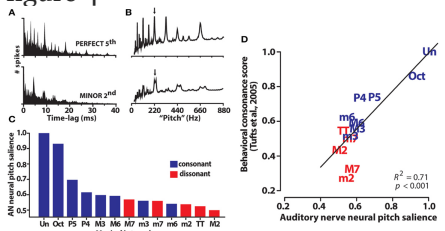

FIGURE 4. AUDITORY NERVE (AN) RESPONSES TO MUSICAL DYADS. (A) Population level interspike interval histograms (ISIHs) for a representative consonant (perfect fifth: 220 + 330 Hz) and dissonant (minor second: 220 + 233 Hz) musical interval. ISIHs quantify the periodicity of spike discharges from a population of 70 AN fibers driven by a single two-tone musical interval. (B) Neural pitch salience profiles computed from ISIHs via harmonic sieve analyses quantify the salience of all possible pitches contained in AN responses based on harmonicity of the spike distribution. Their peak magnitude (arrows) represents a singular measure of neural pitch salience for the eliciting musical interval. (C) AN pitch salience across the chromatic intervals is more robust for consonant than dissonant intervals. Rank order of the intervals according to their neural pitch salience parallels the hierarchical arrangement of pitches according to Western music theory (i.e., Un > Oct > P5, >P4, etc.). (D) AN pitch representations predict the hierarchical order of behavioral consonance judgments of human listeners (behavioral data from normal-hearing listeners of Tufts et al., 2005). AN data reproduced from Bidelman and Heinz (2011).

In follow-up analyses, it was shown that neither acoustic nor traditional psychophysical explanations (e.g., periodicity, roughness/beating) could fully account for human consonance ratings (Bidelman and Heinz, 2011). Of the number of explanatory factors examined, neural harmonicity was the most successful predictor of human percepts (cf. Bidelman and Krishnan, 2009). Recent psychoacoustical evidence corroborates these findings and confirms that the perception of consonance-dissonance is governed primarily by the harmonicity of a

musical interval/chord and not its roughness or beating (McDermott et al., 2010; Cousineau et al., 2012). That is, converging evidence indicates that consonance is largely computed based on the degree to which a stimulus sounds like a single harmonic series.

The Hierarchical Nature and Basis of Subcortical Pitch Processing

To date, overwhelming evidence suggests that *cortical* integrity is necessary to support the cognitive aspects of musical pitch (Johnsrude et al., 2000; Ayotte et al., 2002; Janata et al., 2002; Peretz et al., 2009; Itoh et al., 2010). Yet, aggregating our findings from AN, human brainstem responses, and behavior provides a coherent picture of the emergence and time-course of musical pitch percepts in the ascending auditory pathway (Figure 5). Collectively, our findings demonstrate that the perceptual sonority and behavioral preference for both musical intervals and chords (triads) is well-predicted from early subcortical brain activity. Most notably, they also suggest that nascent neural representations relevant to the perception and appreciation of music are emergent well before cortical involvement at pre-attentive stages of audition.

figure 5

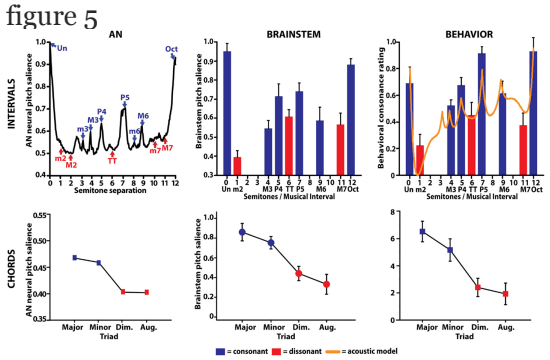

FIGURE 5. COMPARISON BETWEEN AUDITORY NERVE, HUMAN BRAINSTEM EVOKED POTENTIALS, AND BEHAVIORAL RESPONSES TO MUSICAL INTERVALS. (Top left) AN responses correctly predict perceptual attributes of consonance, dissonance, and the hierarchical ordering of musical dyads. AN neural pitch salience is shown as a function of the number of semitones separating the interval's lower and higher pitch over the span of an octave (i.e., 12 semitones). Consonant musical intervals (blue) tend to fall on or near peaks in neural pitch salience whereas dissonant intervals (red) tend to fall within trough regions, indicating more robust encoding for the former. Among intervals common to a single class (e.g., all consonant intervals), AN responses show differential encoding resulting in the hierarchical arrangement of pitch typically described by Western music theory (i.e., Un > Oct > P5, >P4, etc.). (Top middle) neural correlates of musical consonance observed in human brainstem responses. As in the AN, brainstem responses reveal stronger encoding of consonant relative to dissonant pitch relationships. (Top right) behavioral consonance ratings reported by human listeners. Dyads considered consonant according to music theory are preferred over those considered dissonant [minor second (m2), tritone (TT), major seventh (M7)]. For comparison, the solid line shows predictions from a mathematical model of consonance and dissonance (Sethares, 1993) where local maxima denote higher

degrees of consonance than minima, which denote dissonance. (Bottom row) auditory nerve (left) and brainstem (middle) responses similarly predict behavioral chordal sonority ratings (right) for the four most common triads in Western music. Chords considered consonant according to music theory (i.e., major, minor) elicit more robust subcortical responses and show an ordering expected by music practice (i.e., major > minor » diminished > augmented). AN data from Bidelman and Heinz (2011); interval data from Bidelman and Krishnan (2009); chord data from Bidelman and Krishnan (2011).

As in language (Hickok and Poeppel, 2004), brain networks engaged during music likely involve a series of computations applied to the neural representation at different stages of processing. It is likely that higher-level abstract representations of musical pitch structure are first initiated in acoustics (Gill and Purves, 2009; McDermott et al., 2010). Physical periodicity is then transformed to musically relevant neural periodicity very early along the auditory pathway (AN; Tramo et al., 2001; Bidelman and Heinz, 2011), transmitted, and further processed (or at least maintained) in subsequently higher levels in the auditory brainstem (McKinney et al., 2001; Bidelman and Krishnan, 2009, 2011; Lee et al., 2009). Eventually, this information ultimately feeds the complex cortical architecture responsible for generating (Fishman et al., 2001) and controlling (Dowling and Harwood, 1986) musical percepts.

Importantly, it seems that even the non-musician brain is especially sensitive to the pitch relationships found in music and is enhanced when processing consonant relative to dissonant chords/intervals. The preferential encoding of consonance might be attributable to the fact that it generates more robust and synchronous phase-locking than dissonant pitch intervals. A higher neural synchrony for the former is consistent with previous neuronal recordings in AN (Tramo et al., 2001), midbrain (McKinney et al., 2001), and cortex (Fishman et al., 2001) of animal models which show more robust temporal responses for consonant musical units. For these pitch relationships, neuronal firing occurs at precise, harmonically related pitch periods; dissonant relations on the other hand produce multiple, more irregular neural periodicities. Pitch encoding mechanisms likely exploit simple periodic (cf. consonant) information more effectively than aperiodic (cf. dissonant) information (Rhode, 1995; Langner, 1997; Ebeling, 2008), as the former is likely to be more compatible with pitch extraction templates and provides a more robust, unambiguous cue for pitch (McDermott and Oxenham, 2008). In a sense, dissonance may challenge the auditory system in ways that simple consonance does not. It is conceivable that consonant music relationships may ultimately reduce computational load and/or require fewer brain resources to process than their dissonant counterparts due to the more coherent, synchronous neural activity they evoke (Burns, 1999, p. 243).

One important issue concerning the aforementioned FFR studies is the degree to which responses reflect the output of a *subcortical*, brainstem "pitch processer" or rather, a reflection of the representations propagated from more peripheral sites (e.g., AN). Indeed, IC architecture [orthogonal frequency-periodicity maps (Langner, 2004; Baumann et al., 2011), frequency lamina (Braun, 1999)] and response properties (critical bands, spectral integration) make it ideally suited for the

extraction of pitch-relevant information (Langner, 1997). Yet, stark similarity between correlates observed in the AN (Bidelman and Heinz, 2011) and human brainstem FFRs (Bidelman and Krishnan, 2009, 2011) implies that the neurophysiological underpinnings of consonance and dissonance which may be established initially in the periphery, are no more than mirrored in brainstem responses observed upstream. Moreover, recent work also suggests that while brainstem responses may reflect pitch bearing-information, they themselves may not contain an adequate code to support all the intricacies of complex pitch perception (Gockel et al., 2011; but see Greenberg et al., 1987). Gockel et al. (2011), for instance, measured FFRs to complex tones where harmonics 2 and 4 were presented to one ear and harmonic 3 to the other (dichotic condition). Results showed that the FFR magnitude spectra under the dichotic listening condition were qualitatively similar to the sum of the response spectra for each ear when presented monaurally and furthermore, an absence of energy at F0 in the dichotic condition. These results imply that the FFR may preserve monaural pitch cues but may not reflect any additional "pitch" processing over and above what is contained in the combined representations from the periphery (i.e., AN). On the contrary, other studies have observed binaural interactions[1] (Hink et al., 1980; Krishnan and McDaniel, 1998) and neural correlates for complex pitch attributes, e.g., "missing fundamental" (Galbraith, 1994), in the human FFR which are not observed in far-field responses generated from more peripheral auditory structures. These discrepancies highlight the need for further work to disentangle the potential differential (or similar) roles of brainstem and peripheral auditory structures in the neurocomputations supporting pitch. One avenue of investigation which may offer insight to these questions is to examine the degree to which neural plasticity – induced via training or experience – might differentially tune the neural encoding of pitch across various levels of the auditory pathway. Differential plasticity across levels might indicate different functional roles at different stages of auditory processing.

Subcortical Plasticity in Musical Pitch Processing

The aforementioned studies demonstrate a critical link between sensory coding and the perceptual qualities of musical pitch which are independent of musical training and long-term enculturation. Electrophysiological studies thus largely converge with behavioral work, demonstrating that both musicians and non-musicians show both a similar bias for consonance and a hierarchical hearing of the pitch combinations in music (Roberts, 1986; McDermott et al., 2010). Yet, realizing the profound impact of musical experience on the auditory brain, recent studies have begun to examine how musicianship might impact the processing and perceptual organization of consonance, dissonance, and scale pitch hierarchy. Examining training-induced effects also provides a means to examine the roles of nature and nurture on the encoding of musical pitch as well as the influence of auditory experience on music processing.

Neuroplastic Effects on Pitch Processing Resulting from Musical Training

Comparisons between musicians and non-musicians reveal enhanced brainstem encoding of pitch-relevant information in trained individuals (Figure 6) (Musacchia et al., 2007; Bidelman and Krishnan, 2010; Bidelman et al., 2011a,d). Additionally,

as indicated by shorter, less "jittered" response latencies, musicians' neural activity is also more temporally precise than that of non-musicians. Musical training therefore not only magnifies the "gain" of subcortical brain activity (Figure 6D) but also refines it by increasing the temporal precision of the brain's response to complex pitch (Figure 6C) (Bidelman et al., 2011d). Interestingly, these neural indices are correlated with an individual's degree of musical training/experience (Musacchia et al., 2007; Wong et al., 2007) as well as their perceptual abilities (Bidelman et al., 2011b, 2013). Together, these enhancements observed in musicians' brainstem FFRs indicate that experience-dependent plasticity, well-established at cortical levels of processing, also extends to *subcortical* levels of the human brain. A natural question which then arises is the degree to which musical training might modulate the inherent (subcortical) auditory processing subserving musical consonance-dissonance reviewed earlier.

figure 6

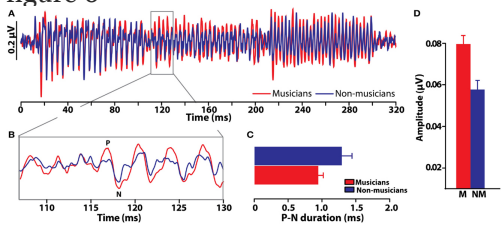

FIGURE 6. EXPERIENCE-DEPENDENT ENHANCEMENT OF BRAINSTEM RESPONSES RESULTING FROM MUSICAL TRAINING. (A) Brainstem FFR time-waveforms elicited by a chordal arpeggio (i.e., three consecutive tones) recorded in musician and non-musicians listeners (red and blue, respectively). (B) Expanded time window around the onset response to the chordal third (≈117 ms), the defining note of the arpeggio sequence. Relative to non-musicians, musician responses are both larger and more temporally precise as evident by their shorter duration P-N onset complex (C) and more robust amplitude (D). Musical training thus improves both the precision and magnitude of time-locked neural activity to musical pitch. Error bars = SEM. Data from Bidelman et al. (2011d).

Experience-Dependent Changes in the Psychophysiological Processing of Musical Consonance

At a subcortical level, recent studies have demonstrated more robust and coherent brainstem responses to consonant and dissonant intervals in musically trained listeners relative to their non-musician peers (Lee et al., 2009). Brainstem phase-locking to the temporal periodicity of the stimulus envelope – a prominent correlate of roughness/beating (Terhardt, 1974) – is also stronger and more precise in musically trained listeners (Lee et al., 2009). These results suggest that brainstem auditory processing is shaped experientially so as to refine neural representations of musical pitch in a behaviorally relevant manner (for parallel effects in language, see Bidelman et al., 2011a). They also indicate that subcortical structures provide differential processing of musical pitch above and beyond "innate" representations which might be established in the periphery (Tramo et al., 2001; Bidelman and Heinz, 2011).

Recent work also reveals similar experience-dependent effects at a cortical level.

Consonant chords, for example, elicit differential hemodynamic responses in inferior and middle frontal gyri compared to dissonant chords regardless of an individual's musical experience (Minati et al., 2009). Yet, the hemispheric laterality of this activation differs between groups; while right lateralized for non-musicians, activation is more symmetric in musicians suggesting that musical expertise recruits a more distributed neural network for music processing. Cortical brain potentials corroborate fMRI findings. Studies generally show that consonant and dissonant pitch intervals elicit similar modulations in the early components of the ERPs (P1/N1) for both musicians and non-musicians alike. But, distinct variation in the later waves (N2) are found nearly exclusively in musically trained listeners (Regnault et al., 2001; Itoh et al., 2003, 2010; Schön et al., 2005; Minati et al., 2009). Thus, musicianship might have a differential effect on the time-course of cortical auditory processing; musical training might exert more neuroplastic effects on later, endogenous mechanisms (i.e., N2) than on earlier, exogenous processing (e.g., P1, N1). Indeed, variations in N2 – which covaries with perceived consonance – are exaggerated in musicians (Itoh et al., 2010). These neurophysiological findings are consistent with recent behavioral reports which demonstrate musicians' higher sensitivity and perceptual differentiation of consonant and dissonant pitches (McDermott et al., 2010; Bidelman et al., 2011b,d). Recently, McDermott et al. (2010) have observed a correspondence between a listener's years of musical training and their perceptual sensitivity for harmonicity (but not roughness) of sound. Thus, it is possible that musician's higher behavioral and neurophysiological propensity for musical consonance might result from an experience-dependent refinement in the internalized templates for complex harmonic sounds. Taken together, neuroimaging work indicates that while sensory consonance is coded in both musically trained and untrained listeners, its underlying neural representations can be amplified by musical expertise. In a sense, whatever aspects of musical pitch are governed by innate processing, musical experience can provide an override and exaggerate these brain mechanisms.

Limitations of these reports are worth mentioning. Most studies examining the effects of musical training on auditory abilities have employed cross-sectional and correlational designs. Such work has suggested that the degree of a musicians' auditory perceptual and neurophysiological enhancements is often positively associated with the number of years of his/her musical training and negatively associated with the age at which training initiated (e.g., Bidelman et al., 2013; Zendel and Alain, 2013). These types of correspondences hint that musicians' auditory enhancements might result from neuroplastic effects that are modulated by the amount of musical exposure. It should be noted however, that comparisons between highly proficient musicians and their age-matched non-musician peers offers an imperfect comparison to address questions regarding the role of *experience* on brain and behavioral processing; causality cannot be inferred from these quasi-experimental, cross-sectional designs. To truly gauge the role of musical experience on harmonicity, consonance perception, and brainstem pitch processing, longitudinal experiments with random subject assignment are needed (e.g., Hyde et al., 2009; Moreno et al., 2009). Interestingly, recent training studies with random subject assignment suggests that even short-term auditory training (~1 month) can positively alter brainstem function as indexed via the FFR (Carcagno and Plack, 2011). Presumably, the high intensity and duration of long-term musical training would only act to amplify these plastic effects observed in the short-term supporting

the notion that experience and "nurture" drive the aforementioned plasticity. Future work may also look to developmental studies (e.g., Schellenberg and Trainor, 1996; Trainor et al., 2002) to disentangle the contributions of experiential and innate factors in musical pitch processing.

Is There a Neurobiological Basis for Musical Pitch?

There are notable commonalities (i.e., universals) among many of the music systems of the world including the division of the octave into specific scale steps and the use of a stable reference pitch to establish key structure. In fact, it has been argued that culturally specific music is simply an elaboration of only a few universal traits (Carterette and Kendall, 1999), one of which is the preference for consonance (Fritz et al., 2009). Together, our recent findings from human brainstem recordings (Bidelman and Krishnan, 2009, 2011) and single-unit responses from the AN (Bidelman and Heinz, 2011) imply that the perceptual attributes related to such preferences may be a byproduct of innate sensory-level processing. These results converge with previous behavioral studies with infants which have shown that months into life, newborns prefer listening to consonant rather than dissonant musical sequences (Trainor et al., 2002) and tonal rather than atonal melodies (Trehub et al., 1990). Given that these neurophysiological and behavioral effects are observed in the absence of long-term enculturation, exposure, or music training, it is conceivable that the perception of musical pitch structure develops from domain-general processing governed by the fundamental capabilities of the auditory system (Tramo et al., 2001; McDermott and Hauser, 2005; Zatorre and McGill, 2005; Trehub and Hannon, 2006; Trainor, 2008).

It is interesting to note that musical intervals and chords deemed more pleasant sounding by listeners are also more prevalent in tonal composition (Budge, 1943; Vos and Troost, 1989; Huron, 1991; Eberlein, 1994). A neurobiological predisposition for simpler, consonant intervals/chords – as suggested by our recent studies – may be one reason why such pitch combinations have been favored by composers and listeners for centuries (Burns, 1999). Indeed, the very arrangement of musical notes into a hierarchical structure may be a consequence of the fact that certain pitch combinations strike a deep chord with the architecture of the nervous system.

Conclusion

Brainstem evoked potentials and AN responses reveal robust correlates of musical pitch at subcortical levels of auditory processing. Interestingly, the ordering of musical intervals/chords according to the magnitude of their subcortical representations tightly parallels their hierarchical arrangement as described by Western music practice. Thus, information relevant to musical consonance, dissonance, and scale pitch structure emerge well before cortical and attentional engagement. The close correspondence between subcortical brain representations and behavioral consonance rankings suggests that listeners' judgments of pleasant- or unpleasant-sounding pitch relationships may, at least in part, be rooted in early,

pre-attentive stages of the auditory system. Of the potential correlates of musical consonance described throughout history (e.g., acoustical ratios, cochlear roughness/beating, neural synchronicity), results suggest that the *harmonicity of neural activity* best predicts human judgments. Although enhanced with musical experience, these facets of musical pitch are encoded in non-musicians (and even non-human animals), implying that certain fundamental attributes of music listening exist in the absence of training, long-term enculturation, and memory/cognitive capacity. It is possible that the preponderance of consonant pitch relationships and choice of intervals, chords, and tuning used in modern compositional practice may have matured based on the general processing and constraints of the sensory auditory system.

Abstract

Scales are collections of tones that divide octaves into specific intervals used to create music. Since humans can distinguish about 240 different pitches over an octave in the mid-range of hearing [1], in principle a very large number of tone combinations could have been used for this purpose. Nonetheless, compositions in Western classical, folk and popular music as well as in many other musical traditions are based on a relatively small number of scales that typically comprise only five to seven tones [2]–[6]. Why humans employ only a few of the enormous number of possible tone combinations to create music is not known. Here we show that the component intervals of the most widely used scales throughout history and across cultures are those with the greatest overall spectral similarity to a harmonic series. These findings suggest that humans prefer tone combinations that reflect the spectral characteristics of conspecific vocalizations. The analysis also highlights the spectral similarity among the scales used by different cultures.

Figures

Citation: Gill KZ, Purves D (2009) A Biological Rationale for Musical Scales. PLoS ONE 4(12): e8144. https://doi.org/10.1371/journal.pone.0008144
Editor: Edward Vul, Massachusetts Institute of Technology, United States of America
Received: July 1, 2009; **Accepted:** November 10, 2009; **Published:** December 3, 2009

Funding: NSF (www.nsf.gov) and Duke University (www.duke.edu). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
Competing interests: The authors have declared that no competing interests exist.

Introduction

The most widely employed scales (also called modes) in Western music over the last few centuries have been the major and minor pentatonic and heptatonic (diatonic) scales (Figure 1). The other scales illustrated are commonly found in early liturgical music and, more recently, in folk music, modern jazz and some classical compositions [5], [7]. These same five-note and seven-note collections are also prevalent in traditional Indian, Chinese and Arabic music, although other scales are used as well [2], [3], [8]–[10]. These historical facts present an obvious puzzle: given the enormous number (billions) of possible ways to divide octaves into five to seven tonal intervals, why have only a few scales been so strongly favored?

Download:
- PPT PowerPoint slide
- PNG larger image
- TIFF original image

Figure 1. Pentatonic and heptatonic scales (included tones are indicated by red dots).

The five pentatonic scales are modes of the same set of notes, the only difference being the starting note or tonic. Seven of the nine heptatonic scales shown are also modes that entail the same notes in different arrangements (the exceptions are the harmonic and melodic minor scales). There are three

unique forms of the minor heptatonic scale: the natural, harmonic and melodic (the melodic minor scale shown is designated as ascending since this scale is identical to the natural minor scale when descending). Although the scales shown begin and end on specific notes of the keyboard, each could begin on any note and retain its identity as long as all intervals between notes remained the same. Scale tones are represented on keyboards for didactic purposes only in this and subsequent figures and should not be interpreted as being tuned in equal temperament (see Methods).

https://doi.org/10.1371/journal.pone.0008144.g001

Not surprisingly, a number of investigators have grappled with the general issue of scale structure. One approach has used consonance curves [11] to show that the consonant harmonic scale tones are defined by small integer ratios [12], [13]. This method has not, however, been used to predict any specific scale structures. A different approach to understanding scales has depended on the concept of a generative grammar in linguistics, asking whether musical patterns might define a "musical grammar" [14]. Again, this concept has not been applied to the prediction of preferred scale structures. A third approach has used error minimization algorithms to predict scale structures under the assumption of competing preferences for small integer ratios and equal intervals between successive scale tones [15], [16]. This method can account for the structure of the equal-tempered 12-tone chromatic scale but cannot account for any of the five to seven-tone scales commonly used to make music. Moreover, no basis was provided for the underlying assumptions. Other analyses have predicted scales with as many as 31 intervals, which are rarely used to make music [17], [18]. In short, none of these approaches explains the widespread human preference for a small number of particular scales comprising five to seven tones, or provides a biological rationale for this predilection.

Here we examine the possibility that the thread tying together the scales that have been preferred in music worldwide is their overall similarity to the spectral characteristics of a harmonic series. The comparison of musical intervals to a harmonic series is not new. Helmholtz [19] first proposed that the relative consonance of musical dyads derives from harmonic relationships of the two tones. More recently, Bernstein [20] suggested that scale structure is determined by the appeal of the lower harmonics that occur in naturally generated harmonic series. For example, assuming octave equivalence, the intervals between the tones of the major pentatonic scale are nearly the same as the intervals between the first nine harmonics of a harmonic series. However, a number of flaws were later pointed out in this argument [14]. For one thing, the last note of the major pentatonic scale only roughly approximates the seventh harmonic. Moreover, widely used scales containing a minor second interval are not predicted, as this interval does not occur until the 15^{th} and 16^{th} harmonics of a harmonic series.

The different approach we take here is to quantitatively compare the harmonic

structure that defines each interval in a possible scale to a harmonic series, rather than to consider only the intervals between fundamental frequencies and individual harmonics. Accordingly our analysis does not depend on intervals and scales precisely mimicking a harmonic series, but evaluates degrees of similarity. The average similarity of all intervals in the scale is then used as a measure of the overall similarity of the scale under consideration to a harmonic series. In this way we assess whether the scales with the highest degree of similarity to a harmonic series are in fact the scales commonly used to make music.

Materials and Methods

Measurement of scale similarity to a harmonic series

The degree of similarity between a two-tone combination (a dyad or interval) and a harmonic series was expressed as the percentage of harmonic frequencies that the dyad held in common with a harmonic series defined by the greatest common divisor of the harmonic frequencies in the dyad (Figure 2). Perceptually, the greatest common divisor of the dyad corresponds to its virtual pitch (or missing fundamental) and is used in much the same way as in algorithms that determine virtual pitch [11], [21]. Since the robustness of a virtual pitch depends on how many of the lower harmonics are present in the stimulus [1], [21], this measure of similarity is both physically and perceptually relevant. For example, a dyad whose spectrum comprises 50% of the harmonic frequencies in a harmonic series would evoke a stronger virtual pitch perception than a dyad with only 10% of these frequencies. We refer to this metric as the *percentage similarity* of a dyad. Percentage similarity can be expressed as $((x+y-1)/(x*y))*100$, where x is the numerator of the frequency ratio and y is the denominator of the ratio. For instance, a major third has a frequency ratio of 5:4; since $x=5$ and $y=4$, the percentage similarity is 40%.

Download:

- PPT PowerPoint slide
- PNG larger image
- TIFF original image

Figure 2. The harmonic structure of a tonal dyad (a major third in this example) compared to a harmonic series.

The fundamental frequency of the harmonic series used for comparison with the dyad is given by the greatest common divisor (100 Hz). In this case, the dyad comprises 8 out of the 20 harmonic frequencies in the harmonic series (percentage similarity =40%).

https://doi.org/10.1371/journal.pone.0008144.g002
The overall conformance of a scale to a harmonic series was then determined by calculating the mean percentage similarity of the dyads in the scale in question (Figure 3). Using the mean as an index of similarity between a scale and a harmonic series implies that all possible dyads in the scale are equally relevant. Although in contemporary Western music any two notes in a scale can, in principle, be used together in melody or harmony, in traditional Western voice-leading and in other musical systems (e.g., classical Indian) particular tone combinations are avoided or prohibited [22]–[24], [25], [26]. Nonetheless, there is no universal rule that describes which intervals might be more important in a scale than others; thus we treated all intervals equally.

Download:

- PPT PowerPoint slide
- PNG larger image
- TIFF original image

Figure 3. Determination of the mean percentage similarity of a scale, using the pentatonic minor scale as an example.

A) The 15 possible intervals between the tones of this scale. B) The percentage similarity of each scalar interval compared to a harmonic series (see Figure 2) and the mean percentage similarity of the full scale are indicated. Scale degrees are conventionally indicated as frequency ratios with respect to a fixed tonic.

https://doi.org/10.1371/journal.pone.0008144.g003
Each scale analyzed is bounded by two tonics that are separated by an octave (see Figure 1); thus intervals spanning octaves (e.g., in a natural minor scale, the interval of a major third between the seventh scale degree and the second scale degree in the octave above) are not included in the calculation of the mean percentage similarity. In Western music, intervals spanning octaves are used in melody; however, in particular scales used by other cultures (classical Indian music for example), these intervals are not used [22]–[24]. Given these facts, we do not assume intervals across octaves to be part of any formal scale structure.

Because musical scale tones are not always defined by a single frequency ratio (e.g., the ratios of 7:5 or 10:7 can both represent a tritone), the algorithm we used allowed tones within a specific frequency distance to represent the same scale tone. To our knowledge, there is no psychoacoustical data on the size of the frequency window within which intervals are considered musically equivalent. We thus defined the

window based on musical practice. Twenty-two cents was used because it is the maximum frequency distance between scale tones that are considered musically equivalent in Western music (i.e., the interval between the minor sevenths defined by ratios of 9:5 and 16:9 [7]); it is also the minimum frequency distance between two tones that are considered unique in classical Indian music [3]. Note that 22 cents is significantly larger than the just noticeable frequency difference between tones (around five cents), implying that the size of the window is not based on the resolution of the auditory system. If two or more ratios fell within the 22 cent window, the algorithm defaulted to the ratio yielding the highest percentage similarity from any comparison. For example, if 9:8 or 10:9 represented the second scale degree of a scale being tested (these two intervals are within 22 cents of each other), the algorithm would use 9:8 rather than 10:9 to form an interval with a perfect fifth (3:2) because this choice produces the interval (4:3 versus 27:20) with the higher percentage similarity. Conversely, the algorithm would use 10:9 rather than 9:8 to form an interval with a major sixth (5:3) because this choice produces the interval (3:2 versus 40:27) with the higher percentage similarity.

Numbers of scales evaluated

The number of scales in any given category that we could have analyzed in theory is given by n!/((n-k)!*k!) where k is the number of different tones in the scale and n the number of discriminable tones over an octave in the middle range of human hearing. If we had considered every discriminable interval over an octave as a potential scale tone, the number of possible scales would have been computationally overwhelming. For example, using the value of 240 discriminable tones over an octave given by Zwicker and Fastl [1], the number of possible seven-note combinations is $>10^{11}$. As a compromise between evaluating as many scales as possible while limiting the computational load, we restricted the potential scale tones to 60 tones (i.e., 25% of the number of discriminable tones in an octave; see Table 1). The 60 tones used were those that, as dyadic combinations with a fixed tonic, had the greatest percentage similarity to a harmonic series. The tones in this subset were separated by 20 cents on average, which is much closer than the ~100 cent minimum separation of tones in most scales; even classical Indian microtones (srutis) are never separated by less than 22 cents [3]. This restriction left for analysis 455,126 possible pentatonic scales, 45,057,474 heptatonic scales and 279,871,768,995 dodecatonic (12-note) scales (again for reasons of computational efficiency, we analyzed a random sample of only 10 million possible dodecatonic scales). The numbers of possible scales we analyzed are given by n = 59 and k = 4, 6, and 11; 59 was used rather than 60 because the octave is assumed as a component interval of all scales, and 4, 6 and 11 were used rather than 5, 7, and 12 because we treated the first note as a fixed reference point (i.e. a tonic). Thus the tonic note and the octave above it bounded all the scales analyzed. A MATLAB (The Mathworks Inc., Natick MA) algorithm was written to compute the mean percentage similarity for each potential scale and to rank the scales in descending

order according to their mean percentage similarity.

Download:
- PPT
PowerPoint slide
- PNG
larger image
- TIFF
original image

Table 1. The 60 intervals with the greatest percentage similarity to a harmonic series.

https://doi.org/10.1371/journal.pone.0008144.t001

The 50 pentatonic and heptatonic scales with the highest mean percentage similarity were individually compared to scales from various cultures including Western, Arabic, Indian, and East Asian. Figure 1 shows the common Western scales used for comparison. These same heptatonic and pentatonic scales constitute most of the basic scale structures of Indian and East Asian music, respectively [3], [9], [10]. The ragas of classical Indian music are particular subsets of tones from these seven-tone "parent" scales or *thats*, and the numbers reported in the literature vary from under one hundred to thousands [3], [22]–[24]. Multiple different sources were used to compile a comprehensive list of over 4000 ragas for comparison with the scales shown in Tables 2 and 3 [op cit.]. Arabic music uses some of the same heptatonic scales shown in Figure 1 (e.g., the Ajam scale is equivalent to the major scale) in addition to uniquely Arabic scales [2], [27]. As with ragas, the numbers of Arabic scales reported vary; two sources were used to compile a list of 35 for comparison [op cit.]. The randomly chosen dodecatonic scales were not individually analyzed, as the chromatic scale is the only musical scale in this category.

Download:
- PPT
PowerPoint slide
- PNG
larger image
- TIFF
original image

Table 2. The 50 pentatonic scales whose intervals conform most closely to a harmonic series out of ~4×10^5 possibilities examined.

https://doi.org/10.1371/journal.pone.0008144.t002

Download:

- PPT PowerPoint slide
- PNG larger image
- TIFF original image

Table 3. The 50 heptatonic scales whose intervals conform most closely to a harmonic series out of ~4×10^7 possibilities examined.

https://doi.org/10.1371/journal.pone.0008144.t003

The use of justly tuned intervals

Western music over the last few centuries has been based on equal temperament tuning, which developed as a compromise between the aesthetic value of maintaining justly tuned intervals (i.e., intervals defined by relatively small integer ratios) and the practical need to facilitate musical composition and performance in multiple keys, especially on keyboard instruments [28], [29]. Just intonation is generally considered the most natural tuning system and was the system used before orchestras, composers and instrument makers demanded equal temperament (op cit.). Moreover, just intonation is used in non-Western traditions such as classical Indian music [3], [22]–[24]. The scales analyzed in the present study are therefore justly tuned.

Results

Pentatonic scales

Table 2 lists the 50 five-note scales among the >4×10^5 possibilities evaluated in this category with the highest mean percentage similarity to a harmonic series. The scale topping the list is the minor pentatonic scale, one of the most widely used five-note scales [5]. The second highest ranked is the Ritusen scale, a pentatonic mode used in traditional Chinese and Indian music (see Figure 1; [3], [9], [10], [22]–[24]). The third and fourth ranked pentatonic scales are the ascending forms of two ragas (Candrika todi and Asa-gaudi) used in classical Indian music [3]. Although these two scales are not formally recognized in Western music theory, they can be thought of as the natural minor and major heptatonic scales, respectively, with the second and seventh scale degrees excluded. Thus some Western melodies are likely to use these particular combinations of tones. The fifth ranked pentatonic scale is identical to the Ritusen scale (known as the Durga raga in classical Indian music) except that the fifth scale degree (17:10 in this case) is ~34 cents sharp (i.e., higher in frequency) compared to the 5:3 major sixth in the Ritusen scale. Because a sharp sixth interval is musically acceptable in certain contexts in classical Indian music, this scale may indeed represent the Durga raga (see Discussion). The sixth through eighth ranked five-note

scales are the remaining modes of the major/minor pentatonic scale (see Figure 1), and the ninth ranked scale is the Catam raga [3].

Heptatonic scales

The 50 heptatonic scales with the highest mean percentage similarity among the $>4\times10^7$ possible scales evaluated are shown in Table 3. Three of the seven heptatonic modes (see Figure 1) emerge at the top of this list. The Phrygian mode holds the highest rank followed by the Dorian mode and the Ionian mode (the major scale). The fourth ranked scale is similar to the Phrygian mode but contains a neutral second (12:11) instead of a minor second; this collection is the Husayni scale in Arabic music [27]. The Aeolian mode (the natural minor scale) and Lydian mode are the fifth and sixth ranked scales. The next three scales are similar to the Dorian mode but with slight alterations in one or two scale degrees. The seventh ranked scale may represent the Kafi scale in classical Indian music with an alternative sharp sixth scale degree [22]. The eighth ranked scale is the Kardaniya scale in Arabic music [op cit.]. Although the ninth ranked scale does not represent any well-known musical tone collection, the Mixolydian mode is ranked tenth. The Locrian, which is the least used of the Western modes, is ranked fiftieth. Thus both the five-note and seven-note scales preferred in much music worldwide comprise intervals that conform optimally to a harmonic series.

Dodecatonic scales

A further question is the status of the chromatic scale, which divides octaves into 12 approximately equal intervals (semitones). Both Western and Chinese music theory use the chromatic scale as an organizing principle.

When we compared the chromatic scale to a random sample of 10 million other possible 12-note scales, we found that ~1.5 million had higher mean percentage similarity to a harmonic series, and none of these, to our knowledge, have been used in music. These results are in sharp contrast to the commonly used five- and seven-note scales that rank at or near the top of their respective groupings. This observation suggests that the chromatic scale has no basis in similarity to a harmonic series. This result is consistent with the fact that the full set of 12 tones is not as widely used as the five-and seven-tone subsets shown in Figure 1, and is considered by some to be less accessible to listeners [14], [30]. Nonetheless, modern composers such as Schoenberg, Webern and Berg have used the chromatic scale as a basis for musical compositions.

Discussion

The results we report indicate that musical scale preferences are predicted by the overall similarity of their component intervals to a harmonic series. However, several

caveats and the possible reasons behind this preference deserve mention.

Competing explanations of interval preference

Although the results shown in Tables 2 and 3 suggest that musical intervals that are maximally similar to a harmonic series are favored, a number of other explanations of interval preferences have been proposed over the years. One historically important theory was suggested by Helmholtz [19], who argued that dissonant musical tone combinations are produced by inadequate harmonic overlap. In other words, when the harmonics of two musical tones fall within the minimum frequency distance at which two pure tones can be individually resolved by humans (the critical bandwidth), an unpleasant perception of "beating and roughness" occurs (see also refs. 11, 39–42). Another explanation for interval preferences is based on the relationships among the harmonics produced by the voice or by musical instruments [20], [21]. In this view, the frequency ratios between lower, more powerful harmonics are more readily appreciated, leading to a perceptual preference for dyads whose fundamentals are smaller integer ratios. A third interpretation of scale preferences is based on the elicitation of more harmonious virtual pitches [45]. For example, in addition to the perception of the pitches of the two component tones, a perfect fifth elicits the perception of a virtual pitch an octave below the lower tone. In this theory, such virtual pitches could make an interval more consonant.

Whether any of these theories of dyadic preference could account for scale preferences in music has not been examined. Nonetheless, the rankings of interval preferences predicted by these theories are similar to one another and to the ranking predicted by harmonic series-similarity (see Table 1) [19]–[21], [39]–[42], [45]. This is not surprising since each theory was developed to explain the same generally accepted consonance ranking of dyads. Thus any of these theories could account for scale preferences if the metrics were quantified and used in the algorithm presented here in place of percentage similarity. For example, the scales with the highest mean percentage similarity are likely to be the ones with the highest mean harmonic overlap or lowest mean beating. It is impossible to tease apart the metric or combination of metrics that is responsible for scale preferences using this algorithm alone. It is noteworthy, however, that these theories are all variations of the same general idea, namely a human preference for particular characteristics of harmonic series.

A biological rationale

Why then should this preference exist? Although other explanations cannot be ruled out based on the data we have presented, for the reasons discussed in this section, we favor a biologically based preference for harmonic series as the most plausible explanation for the particular scales used to make music over history and across cultures.

Like any other sensory quality, the human ability to perceive tonal (i.e., periodically repeating) sound stimuli has presumably evolved because of its biological utility. In nature, such sound stimuli typically occur as harmonic series produced by objects that resonate when acted on by a force [19], [31]. Such resonances occur when, for example, wind or water forces air through a blowhole or some other accidental configuration, but are most commonly produced by animal species that have evolved to produce periodic sounds for social communication and ultimately reproductive success (e.g., the sounds of stridulating insects, the vibrations produced by the songbird syrinx, and the vocalizations of many mammals). Although all these harmonic stimuli are present in the human auditory environment, the vocalizations of other humans are presumably the most biologically relevant and frequently experienced.

In humans, vocal stimuli arise in a variety of complex ways, not all of which are harmonic. Harmonic series depend on vocal fold vibrations and are characteristic of the "voiced speech" responsible for vowel sounds and some consonants [1]. Although the relative amplitudes of harmonics are altered by filtering effects of the supralaryngeal vocal tract resonances to produce different vowel phones, the frequencies of harmonics remain unchanged [op cit.]. In consequence, the presence of a harmonic series is a salient feature of human vocalizations and essential to human speech and language. It follows that the similarity of musical intervals to harmonic series provides a plausible biological basis for the worldwide human preference for a relatively small number of musical scales defined by their overall similarity to a harmonic series.

Several lines of evidence accord with this idea. First, humans and other primate species are specifically attracted to conspecific vocalizations, including those with harmonic and even specifically musical characteristics [32]–[38]. Second, the human pinna, ear canal and basilar membrane are all optimized to transmit human vocalizations, suggesting that the human sense of tonality co-evolved to respond to the stimuli generated by the vocal tract [31], [39], [40]. Third, a number of non-musical phenomena in tone perception including perception of the missing fundamental, pitch shift of the residue, spectral dominance and pitch strength can be explained by in terms of the specialization of the human auditory systems for processing vocal sounds [41]–[43]. These observations all support the idea that the musical scales used over human history have resulted from a preference for collections of dyads that most resemble a harmonic series, and therefore human vocalizations.

The biological relevance of other musical features

This interpretation raises the question of whether other features of human vocalizations are, for similar reasons, influential in musical preferences. In addition to harmonicity per se, particular frequency ranges, timbres and prosodic fluctuations make vocalizations specifically human and may be equally or more influential in

musical preferences. In support of this idea, non-human primates have been recently shown to respond affectively to music characterized by frequency ranges and prosody that are similar to their own vocalizations [44]. This evidence accords with the fact that most music, even purely instrumental music, is composed within the human vocal range, and some popular instruments (e.g., the violin) bear a timbral resemblance to the human voice [31]. Moreover, many musical traditions use tones that fall between formal scale tones: in Western music, glissandos involve continuous changes in pitch, blues music depends on "bending" guitar strings to blend the pitches of major and minor thirds, and classical Indian music employs microtonal intervals that fall between the scale tones of ragas [24]. These musical embellishments may reflect the continuous variations in fundamental frequencies that characterize speech prosody. Preferred meters and tempos may also parallel speech and other vocalizations in ways that do not involve tonality at all [6]. Thus while scale preferences seem to be based on the harmonic series that derive from vocal fold vibrations, other aspects of music may be favored because they resemble additional features of the human voice.

The different usage of highly ranked scales

Although many of the highly ranked heptatonic and pentatonic scales in Tables 2 and 3 have been widely used in Western music, some others have not. A possible explanation is that whereas all the modes shown in Figure 1 can be played using the same set of intervals, one or more additional intervals (e.g., a neutral second) would be necessary to play the other highly ranked but little used variations. Since it is relatively easy to play the modes on the same instrument with the same tuning, this property has both practical and theoretical appeal. Nonetheless, some of these other scales are used in non-Western cultures [3], [23], perhaps because their instrumentation favors non-Western tonal relationships. A few highly ranked scales in Tables 2 and 3 may not be used in music simply because they differ so little from the scales that are used. For example, the ninth ranked heptatonic scale is not, to our knowledge, recognized as a scale in its own right. It is, however, nearly the same as the Durga raga with microtonal embellishments, as well as to the Kafi scale and the Dorian mode, both of which have a higher mean percentage similarity.

A related concern is why the ordering of the widely used five-note and seven-note scales from greatest to least mean percentage similarity values in Tables 2 and 3 does not simply follow the order of their popularity, at least in Western music. For example, the major and natural minor heptatonic scales prevalent in Western music today rank below the Phrygian and Dorian modes. One possibility is again instrumentation. For instance, an early explanation for using the Aeolian mode as opposed to another minor mode (e.g., Dorian) was to facilitate performance on particular instruments in certain tunings [46].

A scale that deserves special comment is the Locrian mode, which ranks much lower in Table 3 than the other modes. The Locrian mode is recognized in Western music

theory but rarely used. While reasons cited for the infrequent use of the Locrian mode are its weak tonal center and dissonant tonic chord, it may be less desirable primarily because of the relatively low conformance of its intervals to a harmonic series and thus to the biological signature of voiced speech and other harmonic vocalizations.

Finally, although many of the widely used scales in music worldwide hold high ranks in Tables 2 and 3, scales that are used in the music of a few cultures do not. For example, the sléndro scale used in Javanese gamelan music comprises five approximately equally spaced tones over an octave [47], [48] but is not among the pentatonic scales with the highest mean percentage similarity to a harmonic series. A possible explanation in this case is that the metallophone instruments used by gamelan orchestras (e.g., bells and gongs that are idiosyncratic to a given geographical region) generate non-harmonic frequencies. Thus the present analysis based on harmonic series is not applicable to such instruments or the scales that derive from them. It should also be noted that several Arabic scales examined are not present in Table 3. One possible explanation is that the most commonly used scales are those in Table 3, while the less commonly used scales have lower percentage similarity. However, to our knowledge, there is no consensus about which Arabic scales are most frequently used to make music. Alternatively, harmonic series similarity may not the only factor influencing scale preferences in this culture. By the same token, only a few of the hundreds to thousands of classical Indian ragas are represented among the highly ranked pentatonic and heptatonic scales. However, nearly all the "parent" scales (*thats*) from which all ragas are derived are among the highly ranked heptatonic scales indicated by their Western names in Figure 1 and Table 3.

The relative popularity of five- and seven-tone scales

The fact that most musical scales emphasize five or seven tones raises the question of why such scales are preferred over those with a larger or smaller numbers of tones. As the number of tones in a scale decreases, the similarity of the tone collection to the character of a harmonic series increases (compare the percentage similarity values of the top-ranked pentatonic and heptatonic scales in Tables 2 and 3). Conversely, dividing octaves into a larger number of intervals leads to tonal collections that meet this criterion less well. Thus under the hypothesis that listeners prefer tone collections whose spectra are on average more like a harmonic series, the inclusion of intervals that conform to this criterion relatively poorly would provide an upper bound on the number of preferred scale tones.

Since tone collections with fewer notes have greater similarity to a harmonic series, it is less clear why tone collections smaller than five notes are not preferred. One reason may be that as the number of scale tones that divide an octave decreases, the distance between successive notes necessarily increases. Larger intervals are more difficult to sing [5], presumably because the associated changes in vocal fold tension and vocal tract shape require a greater expenditure of neuromuscular energy and practice to

develop the necessary coordination. Multiple sequential skips (intervals of a third or greater) are discouraged in traditional rules of voice-leading for this reason [25], [26]. Thus beyond a relatively small number of scale tones (e.g., five), a further decrease would increase the difficulty of vocal (or instrumental) performance, outweighing the gain in harmonic series similarity. Moreover, decreasing the number of scale tones decreases the variety of intervals available for musical composition. In short, the number of tones used in popular scales may be a compromise between these competing factors.

A further issue is the place of six-note scales, which seem less frequently used than five- or seven-note scales. In fact, blues scales, which are prevalent in popular music today, are often classified as six-note variants of five- or seven-note scales and considered hexatonic scales by some musicologists [49]–[51]. Six tones are also used in particular Indian ragas [3], [22]–[24]. Melodies using heptatonic scales sometimes use only six out of the seven tones, and melodies using pentatonic scales often use passing tones not included in the scale structure as such [5]. Such compositions could also be interpreted as using six-note scales. Thus there is certainly nothing prohibitive about using a set of six tones to create music; they are simply not recognized as formally as their five- and seven-note counterparts in Western music theory.

The method of analyzing scales

The algorithm we used to analyze scales is unique in that it accounts for every possible interval between scale tones over an octave. Other analyses have focused on intervals between tones and the tonic [15], [16], [20]. Accounting for all possible intervals is essential to our argument and essential to understanding the historical fact that intervals between any two scale tones can be heard as consonant or dissonant and affect the overall appeal of the scale [28], [29]. An algorithm of this sort has the further virtues of being able to incorporate other metrics of interval comparison (see above) and of demonstrating the spectral similarity of the scales commonly used in Western, Indian, Chinese and Arabic music (see Tables 2 and 3).

Conclusions

The analyses we report here show that many of the relatively small number of scales that humans have preferred over history and across cultures comprise intervals that when considered as a set are maximally similar to harmonic series. The basis for these results may be a preference for the biologically significant spectral features that characterize conspecific vocalizations.

Chandrasekaran Abstract | The effects of music training in relation to brain plasticity have caused excitement, evident from the popularity of books on this topic among scientists and the general public. Neuroscience research has shown that music training leads to changes throughout the auditory system that prime musicians for listening challenges beyond music processing. This effect of music training suggests that, akin to physical exercise and its impact on body fitness, music is a resource that tones the brain for auditory fitness. Therefore, the role of music in shaping individual development deserves consideration. Through years of sensory-motor training, often beginning in early childhood, musicians develop an expertise in their instrument or mastery over their voice1 . In the course of training, musicians increasingly learn to attend to the fine-grained acoustics of musical sounds. These include pitch, timing and timbre, the three basic components into which any sound that reaches the human ear — including music or speech — can be broken down2 . Pitch refers to the organization of sound on an ordered scale (low versus high pitch) and is a subjective percept of the frequency of the sound. Timing refers to specific landmarks in the sound (for example, the onset and offset of the sound) and timbre refers to the quality of the sound — a multidimensional attribute that results from the spectral and temporal features in the acoustic signal. Attention to these components is emphasized during music training. For example, a violinist is trained to pay particular attention to pitch cues to effectively tune the violin, an instrumentalist playing in an orchestra has to have a keen sense of timing cues and a conductor needs to rely on timbre cues to differentiate the contribution of various instruments. There is now evidence that music training induces changes in the brain. Indeed, the musician's brain has been used as a model of neuroplasticity3,4. Early studies investigated how music training primes the brain for processing musical sounds and examined the extent to which such plasticity is specific to processing musical sounds1,4,5. These studies revealed that music training induces functional and structural changes in the auditory system6 . For example, compared to non-musicians, pianists show increased neural activity (measured by magnetic source imaging) in the auditory cortex in response to hearing piano notes7 . The strength of neuronal activation to piano notes was found to correlate with the age at which piano training began and with the number of years of music training. This suggests that enhanced functional plasticity reflects experience and is not merely a reflection of innate differences between musicians and non-musicians. Musicians also show structural differences in the brain relative to non-musicians8,9, with larger grey matter volume in areas that are important for playing an instrument. These areas include motor, auditory and visuospatial regions8 . In addition, musical aptitude correlates with the volume of the primary auditory cortex and with neurophysiological responses to sinusoidal tones in this area9 . Moreover, musicians show enhanced electrophysiological responses in the auditory cortex to contour and interval information in melodies10, and in the auditory brainstem11 when listening to musical intervals. Notably, many of these studies used correlational data to infer that functional and structural differences between the brains of musicians and non-musicians are a consequence of years of experience with music. However, causality cannot be derived from correlational analysis — the differences could reflect pre-existing genetic differences between the two groups. To address this issue, longitudinal studies have been conducted in which children were randomly assigned to music training and then periodically assessed over time12,13. Compared with children who were assigned to art training, children who underwent music training showed enhanced brain responses to subtle pitch changes in musical stimuli13. Fifteen months of intense music training has also been shown to induce structural changes in the primary auditory and primary motor areas12. These structural changes were associated with improved auditory and motor skills, respectively. Taken together, these data suggest that music training can cause functional and structural changes in the brain throughout our lifetimes, and that these changes may improve music processing.

Transfer effects The impact of music training on the neural processing of music has now been well documented14. However, are the changes in a musician's brain specific to music processing P erspectives nature reviews | Neuroscience Volume 11 | august 2010 | 599 Nature Reviews | Neuroscience Pattern detection Selective enhancement of sounds Dynamic yet stable representation of sounds Sound to meaning or do they transfer to other domains that involve the processing of pitch, timing and timbre cues? Below, we describe data that support the view that the fine-grained auditory skills of musicians, which are acquired through years of training, percolate to other domains, such as speech, language, emotion and auditory processing6 . Thus, music training improves auditory skills that are not exclusively related to music15–18,22,60,62,64. Music and speech are perceptually distinct but share many commonalities at both an acoustic and cognitive level. At the acoustic level, music and speech use pitch, timing and timbre cues to convey information2 . At a cognitive level, music and speech processing require similar memory and attention skills, as well as an ability to integrate discrete acoustic events into a coherent perceptual stream according to specific syntactic rules19. Musicians show an advantage in processing pitch, timing and timbre of music compared with non-musicians20. Music training also involves a high working-memory load, grooming of selective attention skills and implicit learning of the acoustic and syntactic rules that bind musical sounds together. These cognitive skills are also crucial for speech processing. Thus, years of active engagement with the fine-grained acoustics of music and the concomitant development of 'sound to meaning' connections may result in enhanced processing in the speech and language domains. Indeed, musicians show enhanced evoked potentials in the cortex and brainstem in response to pitch changes during speech processing compared with nonmusicians16,21,22. During speech processing, pitch has extra-linguistic functions (for example, it can help the listener to judge the emotion or intention of a speaker and determine the speaker's identity23) as well as a linguistic function (for example, in tone languages, a change in pitch within a syllable changes the meaning of a word). Musicians are also better able to detect small deviations in pitch contours that can determine whether a speaker is producing a statement or a question (demonstrated behaviourally as well as in terms of event-related potentials recorded over the cortex)16. Furthermore, compared with non-musicians, musicians show a more faithful brainstem representation (measured using the frequency-following response (FFR)) of linguistic pitch contours in an unfamiliar language18,24. These results suggest that longterm training with musical pitch patterns can benefit the processing of pitch patterns of foreign languages21,25. Do such transfer effects occur at automatic, pre-attentive (that is, before conscious perception) levels of auditory processing, that is, in the brainstem? Studies in humans and animals show that brainstem auditory processing (Box1) is shaped by both longterm and short-term experience2,20,22,26,27,30,31. Processing at the level of the brainstem can be non-invasively examined by measuring the onset response and the FFR28,29. Such measurements have shown that the auditory brainstem response to speech reflects the physical properties of sound with such fidelity that when the electrical response recorded from the brainstem is played as a sound file, the response sounds a lot like the stimulus that evoked it28,29. Thus, the onset response and the FFR can be used to Box 1 | Cognitive–sensory interplay in musicians Music training is a demanding task that involves active engagement with musical sounds and the connection of 'sound' to 'meaning', a process that is essential for effective communication through music, language and vocal emotion. Formation of efficient sound-to-meaning relationships involves attending to sensory details that include fine-grained properties of sound (pitch, timing and timbre) as well as cognitive skills that are related to working memory: multi-sensory integration (for example, following and performing a score), stream-segregation (the ability to perceptually group or separate competing sounds), interaction with other musicians and executive function (see the figure, top part). The cognitive–sensory

aspects of music training promote neural plasticity and this improves auditory processing of music as well as of other sounds, such as speech (see the figure, lower part). Sound travels from the cochlea to the auditory cortex (shown by light, ascending arrows) via a series of brainstem nuclei that extract and process sound information. In addition, there are feedback pathways (known as the corticofugal network) that connect the cortex to the brainstem and the cochlea in a top-down manner (shown by dark, descending arrows). In musicians, neuroplastic changes have been observed in the auditory cortex as well as in lower-level sensory regions such as the auditory brainstem. The enhanced subcortical encoding of sounds in the brains of musicians compared to non-musicians is probably a result of the strengthened top-down feedback pathways. Active engagement with music improves the ability to rapidly detect, sequence and encode sound patterns. Improved pattern detection enables the cortex to selectively enhance predictable features of the auditory signal at the level of the auditory brainstem, which imparts an automatic, stable representation of the incoming stimulus. P erspectives 600 | august 2010 | Volume 11 www.nature.com/reviews/neuro Nature Reviews | Neuroscience Amplitude Amplitude Amplitude Amplitude a b Timing Pitch Timbre 360 ms 10 ms 10 ms 10 ms 10 ms Stimulus Response 100 300 500 Time Frequency (Hz) Time Time Quiet Noise 0.15 0.20 0.25 0.30 0.35 Stimulus-to-response (r) Musicians Non-musicians –1 –2 –3 –4 0.1 0.2 0.3 0.4 r = –0.445 p = 0.01 Stimulus-to-response (r) SIN perception (HINT score) understand how the brain represents pitch, timing and timbre (FIG. 1). These responses originate in the brainstem, but they are influenced by cortical structures via corticofugal feedback pathways30. This feedback ensures top-down cortical influences even at the earliest stages of auditory processing20,30,31. To determine whether transfer effects occur at subcortical stages of auditory processing, researchers have measured the brainstem responses as musicians and non-musicians hear speech sounds. These studies have revealed that musicians show brainstem plasticity not only for music stimuli but also for speech stimuli22. Specifically, compared with non-musicians, musicians showed superior representation (greater correspondence between stimulus and neural response) of voice pitch cues — including fundamental frequency as well as harmonic components in speech and timevarying components in speech — at the level of the brainstem17,22,32, and superior encoding of linguistic pitch contours18,33 (FIG. 1). This suggests that music training causes changes in auditory processing in the subcortical sensory circuitry. In all of these studies, the neural encoding of sound was positively correlated with the number of years of music training. This, together with longitudinal data12,13, suggests that experience promotes neuroplasticity. Musicians are also more accurate at judging timbre differences between different instruments, as well as during voice processing34, and auditory brainstem responses from musicians show faster neural responses to the onset of, and to other acoustic landmarks in, the speech sounds that reflect the dynamic transition from a consonant to a vowel17,22. There has been considerable interest and controversy in relation to the effects of musical experience on general cognitive abilities. Although there are indications that music training can enhance cognitive ability35, the extent and specificity — whether the changes are due to music training per se or to the cognitive effort involved in music training — of such improvements are still unclear36–38 and warrant further research. Issues such as these make the use of preattentive neural indices28 (FIG. 1) particularly enticing, as these neural measures do not require active participation or cognitive engagement from participants. Indeed, the auditory brainstem response to sound can be collected even when an individual is sleeping or engaged in another task (for example, watching a subtitled movie). Thus, the auditory brainstem response reflects the current state of the nervous system — the state at that time, formed by an individual's life experience with sound. Through examination of this neural index in musicians (in comparison with a control group of non-musicians), we can examine auditory processing in the absence of attention or working-memory confounds.

Selective enhancement in the brain The effect of music training on brain plasticity is not just a ‘volume-knob effect’ — not every feature of the auditory signal improves to the same extent — but leads to the fine-tuning of auditory signals that are salient (with ‘sound to meaning’ significance) (FIG. 2). Musicians, compared with non-musicians, more effectively represent the most meaningful, information-bearing elements in sounds — for example, the segment of a baby’s cry that signals emotional meaning39, the upper note of a musical chord11,40 or the portion of the Mandarin Chinese pitch contour that corresponds to a note along the diatonic musical scale33. Furthermore, musicians show improvements in auditory verbal memory and auditory attention, but not in visual memory or visual attention41,38. Thus, music training induces an enhancement of the processing of auditory signals, the characteristics of which depend on the nature of the training (for example, conductors show superior peripheral spatial auditory processing relative to pianists42), on the practice strategies (for example, musicians who learn ‘by ear’ show superior auditory encoding of musical Figure 1 | Neural representation of pitch, timing and timbre in the human auditory brainstem. Timing, pitch and timbre are the basic information-bearing elements in music and speech. The auditory brainstem responserepresents a faithful reconstruction of these features and can be recorded in a noninvasive manner in human participants. a | The auditory brainstem response to a speech sound can be studied in the time domain as changes in amplitude across time (top, middle and bottom-left panels) and in the spectral domain as spectral amplitudes across frequency (bottom-right panel). The auditory brainstem response reflects acoustic landmarks in the speech signal with submillisecond precision in timing and phase-locking that corresponds to (and physically resembles) pitch and timbre information in the stimulus. Here, a speech stimulus (/da/) and the brainstem response to this stimulus are shown by black and red traces, respectively. b | A comparison of stimulus-to-response correlations in musicians and nonmusicians. In musicians and non-musicians the brainstem response is positively correlated with the entire speech stimulus. However, when the stimulus is presented in the presence of background noise, musicians represent the sound features more faithfully than non-musicians (top panel). More faithful stimulusto-response correlations in musicians are functionally relevant; individuals who had higher correlations between the stimulus and the brainstem response to the stimulus in the presence of background noise exhibited better speech-in-noise (SIN) perception in standardized tests (for example, the Hearing in Noise Test (HINT)) (bottom panel). Part b, top panel is reproduced, with permission, from REF. 17© (2009) Society for Neuroscience. Part b, bottom panel, data from REF. 17. P erspectives nature reviews | Neuroscience Volume 11 | august 2010 | 601 Nature Reviews | Neuroscience 0 0.0 0.02 0.04 0.06 0.08 100 200 300 400 500 600 700 Amplitude (mV) Frequency (Hz) Musicians Non-musicians a Not a gain effect but selective enhancement 0 50 100 150 200 –1.0 0.0 1.0 Amplitude (mV) Amplitude Time (ms) Salient portion of the cry 250 b sounds relative to those who rely on nonaural strategies43) and on behavioural relevance (for example, the upper note, which often carries the melody in Western music and evokes a stronger neural response in musicians, or the emotion-bearing segment of a baby’s cry) (FIG. 2). A brain wired to regularities An adaptive auditory system is primed to extract sound regularities in a predictive manner44. The ability to extract statistical regularities in soundscapes probably underlies the well-described statistical learning processes that the brain uses to segment linguistic and non-linguistic inputs44,45.56. For example, we are able to track a friend’s voice (a predictable regularity) in a noisy restaurant that has plenty of competing voices. Adaptive sensory processing is especially beneficial in challenging listening conditions, when the incoming auditory information is noisy or unreliable46. The typical auditory system is capable of extracting regularities in the signal implicitly, even without the need for conscious attention44. Subcortical enhancement of stimulus regularities accompanies success with linguistic tasks, such as reading and

hearing speech in noise56. Through training, musicians learn to pick out sound objects from a complex soundscape, and this improves their ability to track regularities in the environment44. Selective enhancement of the sound stimulus in the musician's brain (FIG. 2) may result from a superior ability to encode predictable, relevant events in the incoming sensory stream14,17,44,47. Higher-level cognitive areas assess the relevance and predictability of information-bearing elements in an auditory signal, and these elements are subsequently represented with greater fidelity (greater stimulus-to-response correspondence) in the auditory system via feedback loops46,48 that are provided by corticofugal pathways (BOX 1). In this way, aspects of the signal that are deemed to be important may be enhanced, whereas irrelevant information is suppressed46. Differences between musicians and nonmusicians in the ability to extract relevant information from the incoming signal have been studied using the mismatch negativity (MMN) as an index. An MMN occurs when the brain detects a change (or violation) in a predictable auditory stream (for example, a rarely presented 'oddball' in the context of a frequently occurring and predictable sound event). Detection of a change in pattern requires a strong neural representation of the predictable stimulus. The magnitude of the MMN response has been shown to closely reflect a person's auditory perceptual ability, that is, a larger MMN reflects a greater perceived distance between two sounds44. Musicians show stronger MMN to musical stimuli49, to linguistic pitch contours24 (a transfer effect) and to abstract sound features50 compared with non-musicians. This indicates that music training may promote an efficient top-down feedback system that is continuously (and automatically) engaged to extract and robustly represent regularities in the auditory system. Consistent with this idea, induced oto-acoustic emissions51,52 have revealed evidence that there are stronger efferent (top-down) effects on cochlear biomechanics in musicians than in non-musicians. There is considerable debate regarding the biological utility of music and the part that music has played in human evolution53–55. A recent proposal posits that music has an important role in shaping the brain within an individual person's lifespan54. According to this proposal, engagement with music induces alterations in the brain and thereby provides a direct biological benefit. Consistent with this proposal, we argue that active engagement with music promotes an adaptive auditory system that is crucial for the development of listening skills. An adaptive auditory system that continuously regulates its activity based on contextual demands is crucial for processing information during everyday listening tasks56, 62. Practical implications Does a selective enhancement in auditory processing and an improved ability to extract regularities in sounds place musicians at an advantage during everyday listening conditions? Few studies have examined this question, but their results have important practical implications. Musicians are more successful than nonmusicians in learning to incorporate sound Figure 2 | Transfer effect and selective enhancement in musicians. a | Ensemble neural responses that are recorded from the auditory brainstem show that compared with non-musicians, musicians show enhanced subcortical representations of music (shown by black and red traces, respectively). The figure shows the response in musicians and non-musicians to hearing a chord. The blue circles depict regions in the spectral domain in which musicians show a stronger response than non-musicians. Importantly, spectral enhancement for musicians is only seen for the upper note, which in Western music often carries the melody. No enhancement is observed for lower notes. The green circles depict lower notes to which musicians show no enhanced response. b | Musicians also show an enhanced subcortical representation of non-musical sounds. The response of musicians and nonmusicians to hearing a baby cry is shown by red and black traces, respectively. Crucially, the neural enhancement is highly specific and selective. Musicians represent the most complex, salient portion of a child's cry (shown by the box, right) more strongly than the earlier-occurring portion (shown by the box, left). Thus, music training does not confer an

overall gain effect but rather selective enhancement of key stimulus features that have sound-to-meaning relationships. Part a is reproduced, with permission, from REF. 11© (2009) Society for Neuroscience. Part b is reproduced, with permission, from REF. 32 © (2009) Wiley InterScience. P erspectives 602 | august 2010 | Volume 11 www.nature.com/reviews/ neuro patterns of a new language into words25. This is likely to be a result of functional and structural brain changes in musicians22,57,58. Furthermore, children who are musically trained show stronger neural activation to pitch patterns of their native language21, have a better vocabulary15 and a greater reading ability59,60 compared with children who did not receive music training. This suggests that musicians have an advantage in everyday speech and language tasks. The link between reading ability and auditory skills (for example, in terms of processing time-varying signals, speed of processing and statistical learning) is well-acknowledged60,61. Deficiencies in the neural representation of essential sound elements are associated with poor reading ability56,62,63, whereas this neural representation is enhanced in musicians. Whether music training provides a benefit in listening to speech during challenging listening environments has also been examined. Speech perception in noise is a challenging task for all individuals, particularly for older adults and young children62. For successful perception of speech in noise, individuals must extract relevant signals from other sounds, a task that requires selective attention, sensory representation of sound and various cognitive skills that include auditory stream segregation and voice tagging. Musicians have shown superior performance in each of these skills compared with non-musicians62. For example, musicians show better speech-in-noise perception than non-musicians during experiments in which participants had to repeat sentences word-for-word as background noise parametrically increased until the participant was unable to repeat the sentences successfully17,64. Musicians also show superior workingmemory performance, which positively correlates with performance in the speechin-noise task64. In addition, the neural representation of timing and harmonic features of the speech signal in the presence of background noise is stronger in musicians than non-musicians17 (FIG. 1). Thus, musicians exhibit enhanced cognitive and sensory abilities that give them a distinct advantage for processing speech in challenging listening environments compared with non-musicians. This advantage develops over the lifetime through consistent practice routines64 and is enhanced by music training that starts early in life7,37,59. Future research needs to focus on the time frame of the experience-dependent plasticity. Understanding the temporal trajectory of plastic changes that are induced by music training will allow us to explore the extent and limits of plasticity in the brain. Implications for education Studies that compare musicians and nonmusicians have identified four determinants of music-training related plasticity: age of training onset7 , number of years of continuous training18,22, amount of practice65 and aptitude9 . Plasticity is influenced by the extent to which a person actively engages in music training relatively early in their life66. The importance of the age of onset of music training can be gleaned from a study that controlled for the number of years of music training and practice67. In this study, musicians who began training before the age of 7 showed superior sensory-motor integration (reflected in a motor sequencing task) compared with those who began music training later in life. Neuroplasticity is also determined by the amount of practice, so benefits of music training should occur even in individuals who begin training later in life17,22,32. Aptitude also plays a part, but is not the sole determinant of neuroplasticity. The results of these studies suggest that the benefits of music training may be accessible to everyone and not just to those who show an aptitude towards music. However, in today's society, musicians are often the product of years of private instruction, a luxury that is possible only for a select few. Taking into consideration what we know about the positive effects of music training, it seems imperative that we afford all children an equal opportunity to improve their listening skills through music training. A large-scale effort to

provide music training early in life can only be achieved through the school system. However, there is growing concern in the United States that the quality and extent of music training that is provided at schools is on the decline owing to other curricular demands68. It is possible that this trend may impair academic achievement in the long term. However, instruction in music and the time that is spent participating in music events do not hamper academic achievement69, and we argue that in fact music training may benefit academic achievement by improving learning skills and listening ability, especially in challenging listening environments. Classrooms, for a variety of reasons, are inherently noisy. There is a strong negative relationship between noise levels in classrooms and academic achievement, even after socio-economic factors have been controlled for70. An effective music training program in schools could reduce the negative influences of external noise62 and better prepare a child for everyday listening challenges beyond the challenges that directly relate to music. Children with learning disorders are particularly vulnerable to the deleterious effects of background noise56,62,71–73. Music training seems to strengthen the same neural processes that are often impaired in individuals with developmental dyslexia or who have Glossary Auditory stream segregation The ability to piece together discrete perceptual events into streams. Contour and interval information Aspects of melodic information in music that are related to contour (upward or downward patterns of pitch changes) and interval (pitch distances between successive notes). Frequency-following response A neuronal ensemble response that phase-locks to the incoming stimulus. Fundamental frequency The lowest frequency of a voice, determined by the rate of vibration of the vocal folds. It generally corresponds to the voice's pitch. Harmonic components in speech Aspects of speech that depend on the rate of vibration of the vocal cords. A voice is composed of a fundamental tone and a series of higher frequencies that are called harmonics. Magnetic source imaging The detection of the changing magnetic fields that are associated with brain activity, and their subsequent overlaying onto magnetic resonance images to identify the precise source of the signal. Mismatch negativity A cortical event-related potential, measured using electroencephalography, that is elicited when a sequence of repeated stimuli (standards) is interrupted by an infrequent stimulus that deviates in sensory characteristics, such as intensity, frequency or duration. Onset response A neuronal ensemble response to the onset of sound. Oto-acoustic emissions Sounds that are generated in the inner ear, which can be recorded non-invasively. They serve as acoustic signatures of the cochlear biomechanical activity. Pitch contours Pitch changes that minimally contrast words in a tone language, such as Mandarin Chinese. Time-varying components in speech Dynamically changing acoustic events (for example, formant transitions) that correspond to articulatory changes during speech production. Voice tagging The ability to use voice pitch as a cue to 'tag' a familiar talker amid fluctuating background noise. P erspectives nature reviews | Neuroscience Volume 11 | august 2010 | 603 difficulty hearing speech in noise56,62,63,74. To put this in perspective, music training cannot and should not replace traditional intervention methods for children with learning problems (for example, children with reading difficulties who undergo phonological and/or auditory training). We suggest that, together with traditional remediation approaches, active engagement with music provides a value-added proposition — an enjoyable social experience that improves listening skills. It should be noted that we learn best about things that we care about; therefore, the engagement of the neural circuitry that underlies emotion during music-making is likely to be helpful in this regard75,76. The data discussed in this article suggest that the role of music training in schools should be reassessed. Research into the effect of music training in schools would also benefit our understanding of brain plasticity. Most of the studies that have been carried out so far have examined musicians who have had years of private instruction. This has provided useful insights, but much remains to be learned. What are the effects of the musical education that is

delivered in schools on the nervous system and on learning outcomes? Are musicians predisposed to learning and processing music and other auditory stimuli in a different way to non-musicians? To what extent are the brain changes that are seen in musicians a result of experience-dependent plasticity13? There are some additional issues regarding the effects of music training on the brain that deserve consideration. The considerable diversity in the training and performance profiles of musicians77 yields a relatively heterogeneous population of individuals who are often lumped together as a single group of 'musicians' in neuroscience studies. This makes it hard to examine the effects of specific forms of music training. In addition, there is a potential selection bias in studies that have compared musicians with nonmusicians. It is possible that non-musicians did not continue with music training because they did not experience any training-related benefits, perhaps owing to genetic factors or poor auditory processing abilities. A longitudinal study of children who begin music training as a part of the school curriculum may be an effective way of addressing the issues of innateness and heterogeneity. Further studies should also address the effectiveness of different music training approaches (for example, the Suzuki method, which emphasizes aural learning over sight reading) as well as performance profiles (for example, improvisational versus classical, instrumental versus vocal learning and solo versus group learning) in determining the effects of music training on brain plasticity. In conclusion, music training results in structural and functional biological changes throughout our lifetime. Such neuroplasticity not only benefits music processing but also percolates to other domains, such as speech processing. The musician's brain selectively enhances information-bearing elements of auditory signals — a process that reflects efficient sound-to-meaning relationships — as well as enhancing the extraction of regularities in the signal. Neural changes such as these have practical implications, as they help to prepare people who actively engage with music for the challenges of language learning and everyday listening tasks. The beneficial effects of music training on sensory processing confer advantages beyond music processing itself. This argues for an improvement in the quality and quantity of music training in schools.

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

Implications for Language and Music

Nina Kraus,[a,b] Erika Skoe,[a] Alexandra Parbery-Clark,[a] and Richard Ashley[c]
Author information Copyright and License information Disclaimer

The publisher's final edited version of this article is available at Ann N Y Acad Sci

Go to:

Abstract

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.

Keywords: brain stem, subcortical, musical training, cochlear implant

Go to:

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.[1,2] 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.[3–7] Interestingly, such cortical differences can be seen as early as 1 year after the onset of musical training[8] and extend to speech signals.[9,10] Recently, this line of research has moved to subcortical 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.

Go to:

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

Work from our laboratory[d] 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.[11] 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).[e] 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 property 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 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.[12] 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.[f] By means of commonly employed digital signal processing tools, such as autocorrelation[g] and Fourier analysis,[h] 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.

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 closeup 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.

Subcortical Representation of *Pitch*

Musicians have extensive experience manipulating *pitch* within the context of music. Work by the Kraus Laboratory[9,13,14] shows that lifelong 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.*[14] employed an audiovisual (AV) paradigm to tap into the multisensory nature of music. Given that music performance involves the integration

of auditory, visual, and tactile information, we hypothesized that lifelong 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.[9] 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 affects on subcortical and cortical sensory encoding mechanisms for both musical and speech stimuli and leads to training-induced malleability of sensory processing.

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 ~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.*[9,14]) (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.*[14]) 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.*[13])

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.[15,16] Using a similar paradigm, we explored the idea that musical *pitch* experience can lead to enhanced linguistic *pitch* tracking.[13] 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.[17] 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[18] and enhanced *pitch* elements in musical chords[23] (reviewed below), we concluded that musical training alters subcortical sensory encoding of dynamic *pitch* contours, especially for complex and novel stimuli.

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.*[13])

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.*[19] 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,[20,21] 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 falling *pitch*). Russo *et al*.[22] 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 lifelong 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.[9,18,23] This is evident for a whole host of stimuli including speech and emotionally affective sounds as well as musical sounds. Lee *et al.*[23] 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[i] 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,[24–27] 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 Musacchia *et al.*,[9] 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 representation of *timing*, resulting in earlier (i.e., faster)

and larger onset peaks[14,18] (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.

Go to:

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.[29,30] By showing that the effects of musical experience on the nervous system's response to sound are pervasive and extend beyond music,[9,13,14,18,31] 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.[9,13,14] 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.*[18] (also appearing in this volume[31a]) 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[32,33] 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[31] recorded speech and song samples from musicians and non-musicians. Vocal productions were analyzed using a statistical analysis of frequency ratios.[34] 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 production, indicating a transfer from the auditory to the motor modalities.

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.*[31])

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[13] or certain aspects of the stimuli are enhanced in musicians.[14,18,23] 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.[35–38] 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.[39] 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.*[35]). Importantly, corticofugal modulation of specific auditory information is evident in the earliest stages of auditory processing.[6] 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,[7,40] 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[9,13,14,18,23] (Fig. 3).

Go to:

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

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.

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 impairments showed improved subcortical response *timing* for speech signals presented in background noise.[51] 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.[52]

Go to:

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.[53] 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.[54]–

60 On *timbre* tasks, CI wearers often have a difficult time telling two instruments apart.54–56,58,60 However, despite this well-documented performance, Koelsch and colleagues61 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.54 However, even if implantation occurs later in life, recent work by Guiraud and colleagues,62 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.[54,63]

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.[64,65] 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.[66,67]

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.*,[68] Cooper *et al.*,[69] and Spitzer *et al.*[70]). 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.[71,72] 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,[73] 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.[13,35,39,74] 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.

Go to:

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.[13–16,18,19,23] This top-down influence is likely mediated by the extensive corticofugal circuitry of descending efferent fibers that course from the cortex to the cochlea.[75] 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.