# Toward a neural basis of music perception – a review and updated model

## Stefan Koelsch*

*Cluster of Excellence "Languages of Emotion," Freie Universität Berlin, Berlin, Germany*

Music perception involves acoustic analysis, auditory memory, auditory scene analysis, processing of interval relations, of musical syntax and semantics, and activation of (pre)motor representations of actions. Moreover, music perception potentially elicits emotions, thus giving rise to the modulation of emotional effector systems such as the subjective feeling system, the autonomic nervous system, the hormonal, and the immune system. Building on a previous article (Koelsch and Siebel, 2005), this review presents an updated model of music perception and its neural correlates. The article describes processes involved in music perception, and reports EEG and fMRI studies that inform about the time course of these processes, as well as about where in the brain these processes might be located.

**Keywords: music, brain, fMRI, EEG, ERAN, semantics**

## 1 INTRODUCTION

Music has been proven to be a valuable a tool for the understanding of human cognition, human emotion, and their underlying brain mechanisms. Music is part of the human nature: It appears that throughout human history, in every human culture, people have played and enjoyed music. The oldest musical instruments discovered so far are around 30,000–40,000 years old (flutes made of vulture bones, found in the cave *Hohle Fels* in Geissenklösterle near Ulm in Southern Germany; Conard et al., 2009), but it is likely that already the first individuals belonging to the species *Homo sapiens* made music (about 100,000–200,000 years ago). Only humans learn to play musical instruments, and only humans play instruments cooperatively together in groups. It is assumed by some that human musical abilities played a key phylogenetical role in the evolution of language (e.g., Wallin et al., 2000), and that music-making behavior engaged and promoted evolutionarily important social functions (such as communication, cooperation, and social cohesion; Cross and Morley, 2008; Koelsch et al., 2010, these function as are summarized further below). Ontogenetically, newborns (who do not yet understand the syntax and semantics of words) are able to decode acoustic features of voices and prosodic features of languages (e.g., Moon et al., 1993), and it appears that infants' first steps into language are based in part on prosodic information (e.g., Jusczyk, 1999). Moreover, musical communication in early childhood (such as parental singing) plays a major role in the emotional, presumably also in the cognitive and social development of children (Trehub, 2003). Making music in a group is a tremendously demanding task for the human brain that elicits a large array of cognitive (and affective) processes, including perception, multimodal integration, learning, memory, action, social cognition, syntactic processing, and processing of meaning information. This richness makes music an ideal tool to investigate the workings of the human brain.

This review article presents an update of a previous model of music perception (Koelsch and Siebel, 2005) in which different stages of music perception are assigned to different modules (see

**Figure 1**; for neuroscientific investigations of music production see, e.g., Bangert and Altenmüller, 2003; Katahira et al., 2008; Herrojo-Ruiz et al., 2009, 2010; Maidhof et al., 2009, 2010). Note that these modules are thought of as entities that do *not* exclusively serve the music-perceptual processes described here; on the contrary: They also serve in part the processing of language, and – as will be illustrated in this review – the model presented here overlaps with models for language processing (for a discussion on the term modularity see also Fodor et al., 1991). The following sections will review research findings about the workings of these modules, thus synthesizing current knowledge into a framework for neuroscientific research in the field of music perception.

## 2 AUDITORY FEATURE EXTRACTION

Music perception begins with the decoding of acoustic information. Acoustic information is translated into neural activity in the cochlea, and progressively transformed in the auditory brainstem, as indicated by different neural response properties for the periodicity of sounds, timber (including roughness, or consonance/dissonance), sound intensity, and interaural disparities in the superior olivary complex and the inferior colliculus (Geisler, 1998; Sinex et al., 2003; Langner and Ochse, 2006; Pickles, 2008). It appears, notably, that already the dorsal cochlear nucleus projects into the reticular formation Koch et al. (1992). By virtue of these projections, loud sounds with sudden onsets lead to startle-reactions, and such projections perhaps contribute to our impetus to move to rhythmic music. Moreover, already the inferior colliculi can initiate flight and defensive behavior in response to threatening stimuli [even before the acoustic information reaches the auditory cortex (AC); Cardoso et al., 1994; Lamprea et al., 2002]. From the thalamus (particularly over the medial geniculate body) neural impulses are mainly projected into the AC (but note that the thalamus also projects auditory impulses into the amygdala and the medial orbitofrontal cortex; Kaas et al., 1999; LeDoux, 2000; Öngür and Price,

2000)[1]. Importantly, The auditory pathway does not only consist of bottom-up, but also of top-down projections; nuclei such as the dorsal nucleus of the inferior colliculus presumably receive even more descending than ascending projections from diverse auditory cortical fields (Huffman and Henson, 1990).

During the last years, a number of studies investigated decoding of frequency information in the auditory brainstem using the frequency-following response (FFR; see also contribution by A. Patel, this volume). The FFR can be elicited preattentively, and is thought to originate mainly from the inferior colliculus (but note also that it is likely that the AC is at least partly involved in shaping the FFRs, e.g., by virtue of top-down projections to the inferior colliculus). Using FFRs, Wong et al. (2007) measured brainstem responses to three Mandarin tones that differed only in their (f0) pitch contours. Participants were amateur musicians and non-musicians, and results revealed that musicians had more accurate encoding of the pitch contour of the phonemes (as reflected in the FFRs) than non-musicians. This finding indicates that the auditory brainstem is involved in the encoding of pitch contours of speech information (vowels), and that the correlation between the FFRs and the properties of the acoustic information is modulated by musical training. Similar training effects on FFRs elicited by syllables with a dipping pitch contour have also been observed in native English speakers (non-musicians) after a training period of 14 days (with eight 30-min-sessions; Song et al., 2008). The latter results show the contribution of the brainstem in language learning, and its neural plasticity in adulthood[2].

A study by Strait et al. (2009) also reported musical training effects on the decoding of the acoustic features of an affective vocalization (an infant's unhappy cry), as reflected in auditory brainstem potentials. This suggests (a) that the auditory brainstem is involved in the auditory processing of communicated states of emotion (which substantially contributes to the decoding and understanding of *affective prosody*), and (b) that musical training can lead to a finer tuning of such (subcortical) processing.

## 2.1 ACOUSTICAL EQUIVALENCY OF "TIMBER" AND "PHONEME"

With regard to a comparison between music and speech, it is worth mentioning that, in terms of acoustics, there is *no* difference between a phoneme and the timber of a musical sound (and it is only a matter of convention that some phoneticians rather use terms such as "vowel quality" or "vowel color," instead of "timber")[3].

Both are characterized by the two physical correlates of timber: spectrum envelope (i.e., differences in the relative amplitudes of the individual harmonics) and amplitude envelope (also sometimes called the amplitude contour or energy contour of the sound wave, i.e., the way that the loudness of a sound changes, particularly with regard to the on- and off-set of a sound)[4]. Aperiodic sounds can also differ in spectrum envelope (see, e.g., the difference between /ʃ/ and /s/), and timber differences related to amplitude envelope play a role in speech, e.g., in the shape of the attack for /b/ vs. /w/ and /ʃ/ vs. /tʃ/.

## 2.2 AUDITORY FEATURE EXTRACTION IN THE AUDITORY CORTEX

As mentioned above, auditory information is projected mainly via the subdivisions of the medial geniculate body into the primary auditory cortex [PAC, corresponding to Brodmann's area (BAs) 41] and adjacent secondary auditory fields (corresponding to BAs 42 and 52)[5]. These auditory areas perform that a more fine-grained, and more specific, analysis of acoustic features compared to the auditory brainstem. For example, Tramo et al. (2002) reported that a patient with bilateral lesion of the PAC (a) had normal detection thresholds for sounds (i.e., the patient could say whether there was a tone or not), but (b) had elevated thresholds for determining whether two tones had the same pitch or not (i.e., the patient had difficulties to detect fine-grained frequency differences between two subsequent tones), and (c) had markedly increased thresholds for determining the pitch direction (i.e., the patient had great difficulties in saying whether the second tone was higher or lower in pitch than the first tone, *even though* he could tell that both tones differed[6]. Note that the AC is also involved in a number of other functions, such as auditory sensory memory, extraction of inter-sound relationships, discrimination, and organization of sounds as well as sound patterns, stream segregation, automatic change detection, and multisensory integration (for reviews see Hackett and Kaas, 2004; Winkler, 2007; some of these functions are also mentioned further below).

Moreover, the (primary) AC is involved in the transformation of acoustic features (such as frequency information) into percepts (such as pitch height and pitch chroma)[7]: Lesions of the (right) PAC result in a loss of the ability to perceive residue pitch in both animals (Whitfield, 1980) and humans (Zatorre, 1988), and neurons in the anterolateral region of the PAC show responses to a missing fundamental frequency (Bendor and

---

[1]The lateral nucleus of the amygdala receives impulses from the medial division of the medial geniculate body of the thalamus (as well as from associated regions of the posterior thalamus, and from auditory association cortex), the central nucleus of the amygdala is reciprocally connected with the brainstem. For a review see LeDoux (2000).

[2]For a study using FFRs to investigate effects of musical training on audiovisual integration of music as well as speech see Musacchia et al. (2007); for a study on the development of auditory brainstem responses to speech in 3- to 12-year-old children see Johnson et al. (2008).

[3]When two sounds are heard that match for pitch, loudness, duration, and location, and a difference can still be heard between the two sounds, that difference is called timber (e.g., Moore, 2008). For example: a clarinet, a saxophone, and a piano all play a c' at the same location, with the same loudness and the same duration. Each of these instruments has a unique sound quality. This difference is called timber, tone color, or simply sound quality. There are also many examples of timber differences in speech. For example, two vowels (such as /ɑ/ and /i/) spoken with the same loudness and same pitch differ from one another in timber.

[4]For example, sudden or slow attack or decay, such as in the sounds of plucked vs. bowed stringed instruments.

[5]For a detailed description of primary auditory "core," and secondary auditory "belt" fields, as well as their connectivity, see Kaas and Hackett (2000). With regard to the functional properties of primary and secondary auditory fields, a study by Petkov et al. (2006) showed that, in the macaque monkey, all of the PAC core areas, and most of the surrounding belt areas, show a tonotopic structure, with the tonotopic structure being clearest in the field A1; the tonotopic organization in the (rostral) field R seems weaker than in A1, but stronger than in RT (which is a third field of the PAC, located rostrally of the field R). The majority of belt areas appears to show a tonotopic structure comparable to that of R and RT, and only few belt areas seem to show only weak, or no, tonotopic structure (Petkov et al., 2006).

[6]For similar results obtained from patients with (right) PAC lesions see Johnsrude et al. (2000) and Zatorre (2001).

[7]For example, a sound with the frequencies 200, 300, and 400 Hz is transformed into the pitch percept of 100 Hz.

Wang, 2005). Moreover, magnetoencephalographic data indicate that response properties in the PAC depend on whether or not a missing fundamental of a complex tone is perceived (Patel and Balaban, 2001; data were obtained from humans). Note, however, that combination tones emerge already in the cochlea, and that the periodicity of complex tones is coded in the spike pattern of auditory brainstem neurons; therefore, different mechanisms contribute to the perception of residue pitch on at least three different levels (basilar membrane, brainstem, and AC)[8]. However, the studies by Zatorre (1988) and Whitfield (1980) suggest that, compared to the brainstem or the basilar membrane, the AC plays the a more prominent role for the transformation of acoustic features into auditory percepts (such as the transformation of information about the frequencies of a complex sound, as well as about the periodicity of a sound, into a pitch percept).

Warren et al. (2003) report that changes in pitch chroma involve auditory regions anterior of the PAC (covering parts of the planum polare) more strongly than changes in pitch height. Conversely, changes in pitch height appear to involve auditory regions posterior of the PAC (covering parts of the planum temporale) more strongly than changes in pitch chroma (Warren et al., 2003). Moreover, with regard to functional differences between the left and the right PAC, as well as neighboring auditory association cortex, several studies suggest that the left AC has a higher resolution of temporal

information than the right AC, and that the right AC has a higher spectral resolution than the left AC (Zatorre et al., 2002; Hyde et al., 2008; Perani et al., 2010).
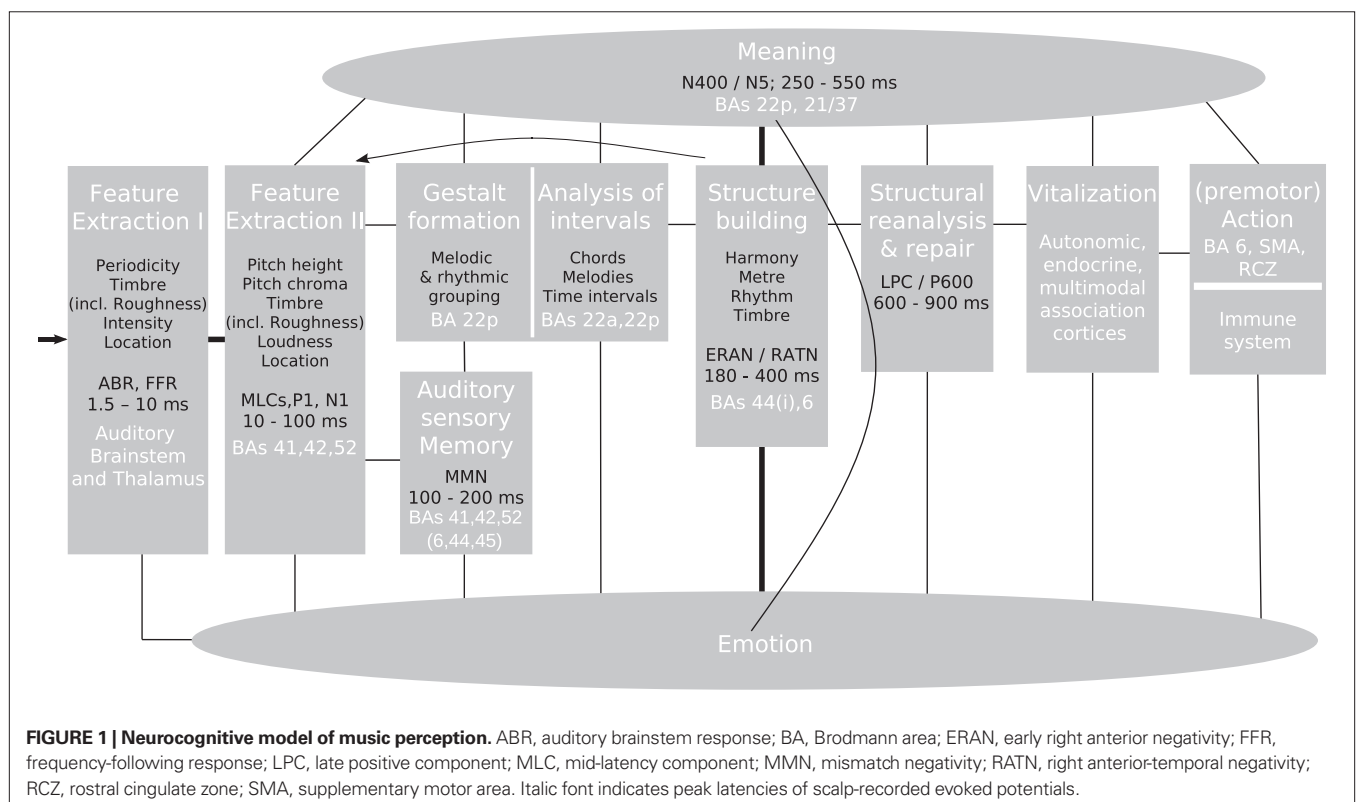
Finally, the AC also prepares acoustic information for further conceptual and conscious processing. For example, with regard to the meaning of sounds, just a short single tone can sound, for example, "bright," "rough," or "dull." That is, single tones are already capable of conveying meaning information (this is indicated by the line connecting the module "Feature Extraction II" and "Meaning" in **Figure 1**; processing of musical meaning will be dealt with further below).

Operations within the (primary and adjacent) AC related to auditory feature analysis are reflected in electrophysiological recordings in ERP components that have latencies of about 10–100 ms, particularly middle-latency responses, including the P1, and the later "exogenous" N1 component (for effects of musical training on feature extraction as reflected in the N1 see, e.g., Pantev et al., 2001).

## 3 ECHOIC MEMORY AND GESTALT FORMATION

While auditory features are extracted, the acoustic information enters the auditory sensory memory (or "echoic memory"), and representations of auditory Gestalten (Griffiths and Warren, 2004; or "auditory objects") are formed (**Figure 1**). Operations of the auditory sensory memory are at least partly reflected electrically in the mismatch negativity (MMN, e.g., Näätänen et al., 2001). The MMN has a peak latency of about 100–200 ms[9], and most presumably

---

[8]Note that responses in the PAC related to the perception of missing fundamental frequencies in the studies by Bendor and Wang (2005) and Patel and Balaban (2001) are presumably in part due to the periodicity information about the missing fundamental frequency coded in the spike pattern of collicular neurons.

[9]For a study differentiating peak latencies of the sensorial and the cognitive part of the MMN see Maess et al. (2007).



**FIGURE 1 | Neurocognitive model of music perception.** ABR, auditory brainstem response; BA, Brodmann area; ERAN, early right anterior negativity; FFR, frequency-following response; LPC, late positive component; MLC, mid-latency component; MMN, mismatch negativity; RATN, right anterior-temporal negativity; RCZ, rostral cingulate zone; SMA, supplementary motor area. Italic font indicates peak latencies of scalp-recorded evoked potentials.

receives its main contributions from neural sources located in the PAC and adjacent auditory (belt) fields, with additional (but smaller) contributions from frontal cortical areas (Giard et al., 1990; Alho et al., 1996; Alain et al., 1998; Opitz et al., 2002; Liebenthal et al., 2003; Molholm et al., 2005; Rinne et al., 2005; Maess et al., 2007; Schonwiesner et al., 2007; for a review see Deouell, 2007). These frontal areas appear to include (ventral) premotor cortex (BA 6), dorsolateral prefrontal cortex near and within the inferior frontal sulcus, and the posterior part of the inferior frontal gyrus (BAs 45 and 44). The frontal areas are possibly involved due to their role in attentional processes, sequencing, and working memory (WM) processes (see also Schonwiesner et al., 2007)[10, 11].

Auditory sensory memory operations are indispensable for music perception; therefore, practically all MMN studies are inherently related to, and relevant for, the understanding of the neural correlates of music processing. As will be outlined below, numerous MMN studies have contributed to this issue (a) by investigating different response properties of the auditory sensory memory to musical and speech stimuli, (b) by using melodic and rhythmic patterns to investigate auditory Gestalt formation, and/or (c) by studying effects of long- and short-term musical training on processes underlying auditory sensory memory operations. Especially the latter studies have contributed substantially to our understanding of neuroplasticity[12], and thus to our understanding of the neural basis of learning. A detailed review of these studies goes beyond the scope of this article (for reviews see Tervaniemi and Huotilainen, 2003; Tervaniemi, 2009). Here, suffice it to say that MMN studies showed effects of long-term musical training on the pitch discrimination of chords (Koelsch et al., 1999)[13], on temporal acuity (Rammsayer and Altenmüller, 2006), on the temporal window of integration (Rüsseler et al., 2001), on sound localization changes (Tervaniemi et al., 2006a), and on the detection of spatially peripheral sounds (Rüsseler et al., 2001). Moreover, using MEG, an MMN study from Menning et al. (2000) showed effects of 3-weeks auditory musical training on the pitch discrimination of tones.

Auditory oddball paradigms were also used to investigate processes of melodic and rhythmic grouping of tones occurring in tone patterns (such grouping is essential for auditory Gestalt formation, see also Sussman, 2007), as well as effects of musical long-term training on these processes. These studies showed effects of musical training (a) on the processing of melodic patterns (Tervaniemi et al., 1997, 2001; Fujioka et al., 2004; Zuijen et al., 2004, in these studies, patterns consisted of four or five tones), (b) on the encoding of the number of elements in a tone pattern (Zuijen et al., 2005), and (c) on the processing of patterns consisting of two voices (Fujioka et al., 2005).

Finally, several MMN studies investigated differences between the processing of musical and speech information. These studies report larger right-hemispheric responses to chord deviants than to phoneme deviants (Tervaniemi et al., 1999, 2000), and different neural generators of the MMN elicited by chords compared to the MMN generators of phonemes (Tervaniemi et al. 1999); similar results have also been shown for the processing of complex tones compared to phonemes (Tervaniemi et al., 2006b, 2009, the latter study also reported effects of musical expertise on chord and phoneme processing). These lateralization effects are presumably due to the different requirements of spectral and temporal processing posed by the chords and the phonemes used as stimuli in that study.

The formation of auditory Gestalten entails processes of perceptual separation, as well as processes of melodic, rhythmic, timbral, and spatial grouping. Such processes have been summarized under the concepts of auditory scene analysis and auditory stream segregation (Bregman, 1994). Grouping of acoustic events follows Gestalt principles such as similarity, proximity, and continuity (for acoustic cues used for perceptual separation and auditory grouping see Darwin, 1997, 2008). In everyday life, such operations are not only important for music processing, but also, for instance, for separating a speaker's voice during a conversation from other sound sources in the environment. That is, these operations are important because their function is to recognize and to follow acoustic objects, and to establish a cognitive representation of the acoustic environment. Knowledge about neural mechanisms of auditory stream segregation, auditory scene analysis, and auditory grouping is still relatively sparse (for reviews see Griffiths and Warren, 2002, 2004; Carlyon, 2004; Nelken, 2004; Scott, 2005; Shinn-Cunningham, 2008; Winkler et al., 2009). However, it appears that the planum temporale (which is part of the auditory association cortex) is a crucial structure for auditory scene analysis and stream segregation (Griffiths and Warren, 2002), particularly due to its role for the processing of pitch intervals and sound sequences (Zatorre et al., 1994; Patterson et al., 2002).

## 4 ANALYSIS OF INTERVALS

Presumably closely linked to the stage of auditory Gestalt formation is a stage of a more fine-grained analysis of intervals, which includes (i) a more detailed processing of the pitch relations between the tones of a chord (required to determine whether a chord with a specific root is a major or minor chord, played in root position, inversion, etc., see also below), or between the tones of a melody; and possibly (ii) a more detailed processing of temporal intervals. With regard to chords, such an analysis is required to specify the "chord form," that is, whether a chord is a major or a minor chord, and whether a chord is played in root position, or in an inversion; note that chords with the same root can occur in several versions (e.g., all of the following chords have the root c: c–e–g, c–e-flat–g, e–g–c, g–c–e, c–e–g–b-flat, e–g–b-flat–c, etc.)[14]. The neural correlates of such processes are not known, but it is likely that both temporal and (inferior) prefrontal regions contribute to such processing; with regard to the processing of melodies, lesion data suggest that the analysis of the contour of a melody (which is part of the auditory Gestalt formation) particularly relies on the posterior part of the right superior temporal gyrus (STG), whereas

---

[10]Moreover, activation of the posterior part of the inferior frontal gyurs (including BA 44) in auditory oddball paradigms (as, e.g., reported in studies by Opitz et al., 2002; Doeller et al., 2003) is perhaps elicited when an auditory stimulus also triggers hierarchical processing of pitch relations.

[11]For a study localizing the generators of the abstract-feature MMN see Korzyukov et al. (2003).

[12]That is, to changes in neuronal structure and function due to experience.

[13]But note that superior attentive pitch discrimination accuracy is not always reflected in the MMN (Tervaniemi et al., 2005).

[14]Listening to a Bach chorale, for example, and specifying on-line for each chord whether it is a major or a minor chord, and specifying the inversion in which it is played, illustrates that this is not a trivial perceptual process.

the processing of more detailed interval information appears to involve both posterior and anterior areas of the supratemporal cortex bilaterally (Liegeois-Chauvel et al., 1998; Patterson et al., 2002; Peretz and Zatorre, 2005)[15].

Melodic and temporal intervals appear to be processed independently, as suggested by the observation that brain damage can interfere with the discrimination of pitch relations but spare the accurate interpretation of time relations, and vice versa (Di Pietro et al., 2004; Peretz and Zatorre, 2005). However, these different types of interval processing have so far not been dissociated in functional neuroimaging studies.

# 5 WHEN INTELLIGENCE COMES INTO PLAY: PROCESSING MUSICAL SYNTAX

The following section deals with the processing of major–minor tonal syntax, particularly with regard to chord functions (i.e., with regard to harmony; for explanation of chord functions see **Figure 2A**). Other aspects of tonal syntax are rhythm, meter, melody (including voice leading), and possibly timbre (these aspects will not be dealt with here due to lack of neurophysiological data).

---

[15]Patients with frontal lesions were not tested in that study. For the processing of chords, more prominent frontal involvement might come into play in the region of (superior) BA 44 (perhaps area 44d according to Amunts et al., 2010) and BA 45, analogous to the activation of these areas during the processing of word form (Longoni et al., 2005).

One possible theoretical description of the syntax of major–minor tonal music is the classical theory of harmony as formulated, e.g., by Rameau (1722), Piston (1948/1987), Schönberg (1969), and Riemann (1971). These descriptions primarily deal with the derivation of chord functions, less so with how chord functions are chained into longer sequences.

## 5.1 FORMAL DESCRIPTIONS OF MUSICAL SYNTAX

Before describing neural correlates of music-syntactic processing, a theoretical basis of musical syntax will be outlined briefly. Heinrich Schenker was the first theorist to deal systematically with the principles underlying (large-scale) structures; this included his thoughts on the *Ursatz*, which implicitly assume (a) a hierarchical structure (such as the large-scale tonic-dominant-tonic structure of the sonata form), (b) that (only) certain chord functions can be omitted, and (c) recursion (e.g., a sequence in one key which is embedded in another sequence in a different key, which is embedded in yet another sequence with yet another different key, etc.). Schenker's thoughts were formalized by the approaches of the "Generative Theory of Tonal Music" (GTTM) by Lerdahl and Jackendoff (1999), which, however, does not provide generative rules (in contrast to what its name says). Nevertheless, the advance of the GTTM lies in the description of Schenker's *Urlinie* as tree-structure (in this regard, GTTM uses terms such as *time-span reduction* and *prolongation*). These tree-structures also parallel the
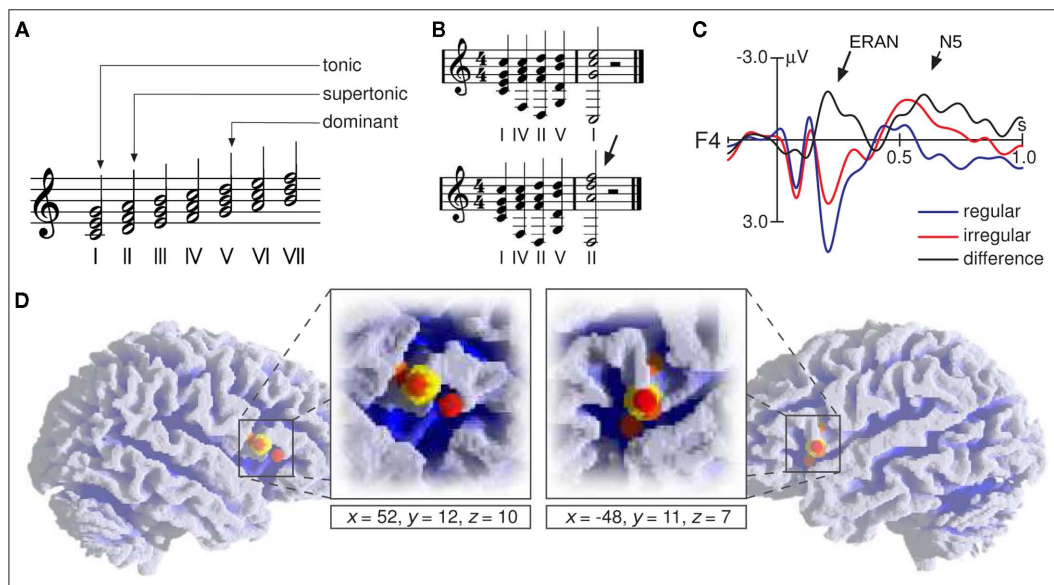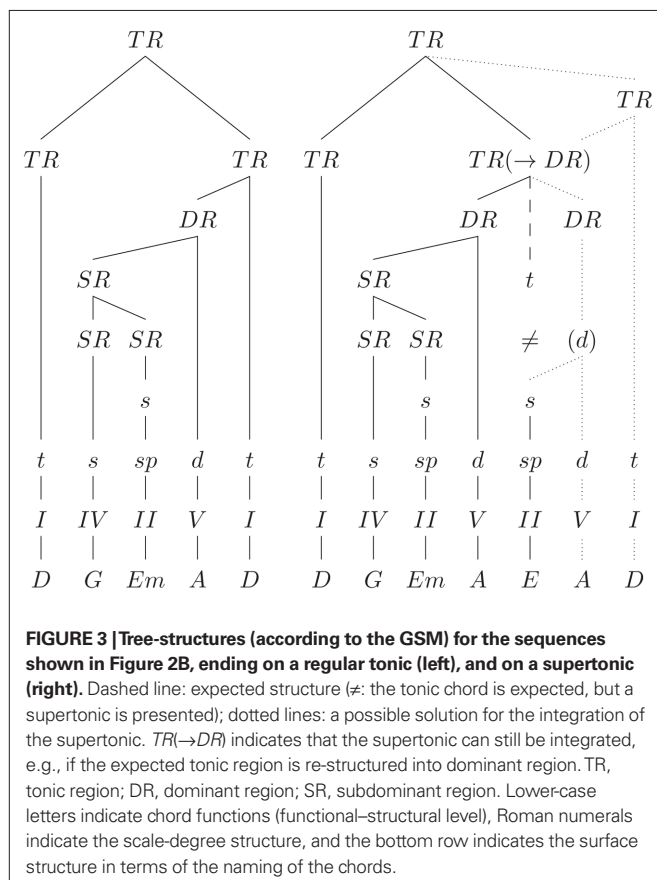
**FIGURE 2 | (A)** Examples of chord functions: The chord built on the first scale tone is denoted as the tonic, the chord on the second tone as the supertonic, and the chord on the fifth tone as the dominant. **(B)** The dominant–tonic progression represents a regular ending of a harmonic sequence (top), the dominant–supertonic progression is less regular and unacceptable as a marker of the end of a harmonic progression (bottom sequence, the arrow indicates the less regular chord). **(C)** ERPs elicited in a passive listening condition by the final chords of the two sequence types shown in **(B)**. Both sequence types were presented in pseudorandom order equiprobably in all 12 major keys. Brain responses to irregular chords clearly differ from those to regular chords (best to be seen in the black difference wave, regular subtracted from irregular chords). The first difference between the two waveforms is maximal around 200 ms after the onset of the fifth chord (ERAN, indicated by the long arrow) and taken to reflect processes of music-syntactic analysis. The ERAN is followed by an N5 taken to reflect processes of harmonic integration (short arrow). **(D)** Activation foci (small spheres) reported by functional imaging studies on music-syntactic processing using chord sequence paradigms (Koelsch et al., 2002, 2005a; Maess et al., 2001; Tillmann et al., 2003) and melodies (Janata et al., 2002a). Large yellow spheres show the mean coordinates of foci (averaged for each hemisphere across studies, coordinates refer to standard stereotaxic space). Reprinted from Koelsch and Siebel (2005).

description of linguistic syntax using tree-structures (see also, e.g., Patel, 2008), although Lerdahl and Jackendoff's GTTM at least implicitly implies that musical and linguistic syntax have only little to do with each other (except perhaps with regard to certain prosodic aspects).

Up to now, no neurophysiological investigation has tested whether individuals perceive music cognitively according to tree-structures; similarly, behavioral studies on this topic are extremely scarce (but see Cook, 1987; Bigand et al., 1996; Lerdahl and Krumhansl, 2007). Note that GTTM is an analytical (not a generative) model; Lerdahl's tonal pitch space (TPS) theory (Lerdahl, 2001) attempts to provide algorithms for this analytical approach, but the algorithms are often not sharp and precise (and need subjective "corrections"). However, TPS provides the very interesting approach to model tension–resolution patterns from tree-structures (e.g., Lerdahl and Krumhansl, 2007).

A more recent approach to model tonal harmony with explicit generative rules according to tree-structures is the *Generative Syntax Model* (GSM) from Rohrmeier (2007) (see also Rohrmeier, 2011). Rohrmeier's GSM distinguishes between four structural levels with increasing abstraction: a level with the surface structure (in terms of the naming of the chords, e.g., C, G, C), a scale-degree structure level (e.g., I, V, I), a functional–structural level (e.g., t, d, t), and a phrase-structure level specifying tonic regions, dominant regions, and subdominant regions (see also **Figure 3**). Each of these levels is described with generative rules of a phrase-structure grammar,



**FIGURE 3 | Tree-structures (according to the GSM) for the sequences shown in Figure 2B, ending on a regular tonic (left), and on a supertonic (right).** Dashed line: expected structure (≠: the tonic chord is expected, but a supertonic is presented); dotted lines: a possible solution for the integration of the supertonic. *TR*(→*DR*) indicates that the supertonic can still be integrated, e.g., if the expected tonic region is re-structured into dominant region. TR, tonic region; DR, dominant region; SR, subdominant region. Lower-case letters indicate chord functions (functional–structural level), Roman numerals indicate the scale-degree structure, and the bottom row indicates the surface structure in terms of the naming of the chords.

the generation beginning on the highest (phrase-structure) level, and propagating information (in the process of derivation) through the sub-ordinate levels. This opens the possibility for the recursive derivation of complex sequences on both a functional and a scale-degree structural level (e.g., "dominant of the dominant of the dominant"). The actual tonality of a tree is propagated from the head ("knot") into the sub-ordinate branches. The tree-structure reflects (a) the formal arrangement of the piece, (b) the phrase structure, (c) the functional aspects of partial phrases and chords, (d) the relation of key-regions, and (e) the degree of relative stability (which is determined by the sub-ordination rules between the branches).

The GSM combines three principles:

(1) Given (a) that musical elements (e.g., a tone or a chord) always occur in relation to other musical elements (and not alone), and (b) that each element (or each group of elements) has a functional relation either to a preceding or to a subsequent element, the simplest way of representing these relations is a tree-structure. This implies that relations of elements that are functionally related, but not adjacent, can be represented by a tree-structure. For example, in a sequence which consists of several chords and which begins and ends on a tonic chord, the first and the last tonic are not adjacent (because there are other chord functions between them), but the last tonic picks up (and "prolongates") the first tonic; this can be represented in a tree-structure in a way that the first and the last tonic chords build beginning- and end-points of the tree, and that the relations of the other chord functions can be represented by dendritic ramification ("prolongating" chords can extend the tonal region, and "progressing" chords determine the progression of tonal functions). This also implies that adjacent elements do not necessarily have a direct structural relation (for example, an initial tonic chord can be followed by a secondary dominant to the next chord: the secondary dominant then has a direct relation to the next chord, but not to the preceding tonic chord). It is not possible that a chord does not have a functional relation to a preceding, nor to a subsequent chord: If in our example the secondary dominant would not be followed by an (implicit) tonic, the sequence would be ungrammatical (with regard to major–minor tonal regularities). Compared to the classical theory of harmony, this GSM approach has the advantage that chord functions are not determined by their tonality (or the derivation from other chords), but by their functional position within the sequence.

(2) The tree-structure indicates which elements can successively be omitted in a way that the sequence still sounds correct. For example, secondary dominants can be omitted, or all chords between first and last tonic chord can be omitted, and a sequence still sounds correct. In a less trivial case, there is a dominant in a local phrase between two tonic chords; the tree-structure then indicates that on the next higher level all chords except the two tonic and the dominant chords can be omitted. This principle is a consequence of the "headedness," in which each knot in the syntax tree dominates sub-ordinate branches. That is, the tree-structure provides a weighting of

the elements, leading to an objectively derivable deep structure (another advantage compared to the classical theory of harmony).

(3) Chord functions are a result of their position within the branches of the tree. For example, a pre-dominant is the subordinate branch preceding a dominant, or there are stable and instable dominants which can be differentiated based on how deeply they are located in the syntax tree. Similarly, a half-cadence is a case in which a dominant is reached as stable local endpoint of a phrase, which must be followed by a phrase that ends on a tonic, thus resolving the open dominant (with regard to the deep structure).

A detailed description of the GSM goes beyond the scope of this article (for details see Rohrmeier, 2011). As mentioned before, the GSM is relatively new, and due to the lack of usable tree-models, previous studies investigating neurophysiological correlates of music-syntactic processing have so far simply utilized the classical theory of harmony as a syntactic rule system: According to the classical theory of harmony, chord functions are arranged within harmonic sequences according to certain regularities (Riemann, 1971, was the first to refer to such regularity-based arrangement as musical syntax). As described above, this regularity-based arrangement implies long-distance dependencies involving hierarchical organization (also referred to as phrase-structure grammar). However, it cannot be excluded that "irregular" chord functions such as the final supertonic in **Figure 2B** are detected as irregular based on a finite state grammar (according to which the tonic is the most regular chord after a I–IV–II–V progression); as will be illustrated below, processing of such chord functions presumably involves processing of both finite state and phrase-structure grammar.

## 5.2 NEURAL CORRELATES OF MUSIC-SYNTACTIC PROCESSING
Neurophysiological studies using EEG and MEG showed that music-syntactically irregular chord functions (such as the final supertonic shown in **Figure 2B**) elicit brain potentials with negative polarity that are maximal at around 150– 350 ms after the onset of an irregular chord, and have a frontal/fronto-temporal scalp distribution, often with right-hemispheric weighting (see also **Figure 2C**). In experiments with isochronous, repetitive stimulation, this effect is maximal at around 150 – 200 ms over right anterior electrodes, and denoted as early right anterior negativity, or ERAN (for a review see Koelsch, 2009b). In experiments in which the position of irregular chords within a sequence is not known (and thus unpredictable), the negativity often has a longer latency, and a more anterior-temporal distribution (also referred to as right anterior-temporal negativity, or RATN; Patel et al., 1998; Koelsch and Mulder, 2002). The ERAN elicited by irregular tones of melodies has a shorter peak latency than the ERAN elicited by irregular chord functions (Koelsch and Jentschke, 2010).

Functional neuroimaging studies using chord sequence paradigms (Maess et al., 2001; Koelsch et al., 2002, 2005a; Tillmann et al., 2003) and melodies (Janata et al., 2002a) suggest that music-syntactic processing involves the pars opercularis of the inferior frontal gyrus (corresponding to BA 44) bilaterally, but with right-hemispheric weighting (see yellow spheres in **Figure 2D**). It seems likely that the involvement of (inferior) BA 44 (perhaps area 44v

according to Amunts et al., 2010) in music-syntactic processing is due to the hierarchical processing of (syntactic) information: This part of Broca's area is involved in the hierarchical processing of syntax in language (e.g., Friederici et al., 2006; Makuuchi et al., 2009), the hierarchical processing of action sequences (e.g., Koechlin and Jubault, 2006; Fazio et al., 2009), and possibly also in the processing of hierarchically organized mathematical formulas and termini (Friedrich and Friederici, 2009, although activation in the latter study cannot clearly be assigned to BA 44 or BA 45). These findings suggest that at least some cognitive operations of music-syntactic and language-syntactic processing (and neural populations mediating such operations) overlap, and are shared with the syntactic processing of actions, mathematical formulas, and other structures based on long-distance dependencies involving hierarchical organization (phrase-structure grammar).

However, it appears that inferior BA 44 is not the only structure involved in music-syntactic processing: additional structures include the superior part of the pars opercularis (Koelsch et al., 2002), the anterior portion of the STG (Koelsch et al., 2002, 2005a), and ventral premotor cortex (PMCv; Parsons, 2001; Janata et al., 2002b; Koelsch et al., 2002, 2005a). The PMCv probably contributes to the processing of music-syntactic information based on finite state grammar: activations of PMCv have been reported in a variety of functional imaging studies on auditory processing using musical stimuli, linguistic stimuli, auditory oddball paradigms, pitch discrimination tasks, and serial prediction tasks, underlining the importance of these structures for the sequencing of structural information, the recognition of structure, and the prediction of sequential information (Janata and Grafton, 2003). With regard to language, Friederici (2004) reported that activation foci of functional neuroimaging studies on the processing of long-distance hierarchies and transformations are located in the posterior IFG (with the mean of the coordinates reported in that article being located in the inferior pars opercularis), whereas activation foci of functional neuroimaging studies on the processing of local structural violations are located in the PMCv (see also Friederici et al., 2006; Makuuchi et al., 2009; Opitz and Kotz, 2011). Moreover, patients with lesion in the PMCv show disruption of the processing of finite state, but not phrase-structure grammar (Opitz and Kotz, 2011).

That is, in the above-mentioned experiments that used chord sequence paradigms to investigate the processing of harmonic structure, the music-syntactic processing of the chord functions probably involved processing of both finite state and phrase-structure grammar. The music-syntactic analysis involved a computation of the harmonic relation between a chord function and the context of preceding chord functions (phrase-structure grammar). Such a computation is more difficult (and less common) for irregular than for regular chord functions (as illustrated by the branch with the dashed line in the right panel of **Figure 3**), and this increased difficulty is presumably reflected in a stronger activation of (inferior) BA 44 in response to irregular chords. In addition, the local transition probability from the penultimate to the final chord is lower for the dominant – supertonic progression than for the dominant – tonic progression (finite state grammar), and the computation of the (less predicted) lower-probability progression is presumably reflected in a stronger activation of PMCv in response to irregular

chords. The stronger activation of both BA 44 and PMCv appears to correlate with the perception of a music-syntactically irregular chord as "unexpected."
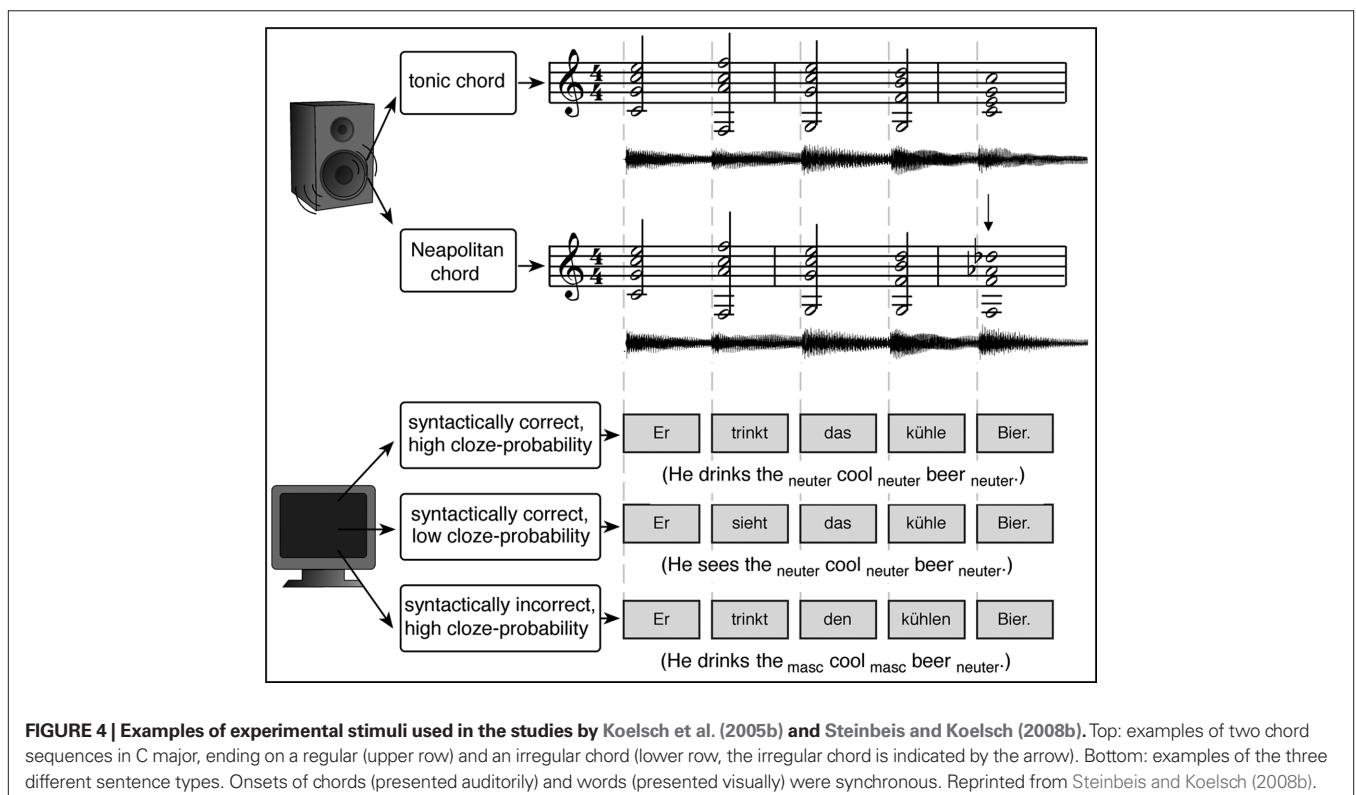
Note that the ability to process phrase-structure grammar is available to humans, whereas non-human primates are apparently not able to master such grammars (Fitch and Hauser, 2004). Thus, it is highly likely that only humans can adequately process music-syntactic information at the phrase-structure level. It is also worth noting that numerous studies showed that even "non-musicians" (i.e., individuals who have not received formal musical training) have a highly sophisticated (implicit) knowledge about musical syntax (e.g., Tillmann et al., 2000). Such knowledge is presumably acquired during listening experiences in everyday life.

## 5.3 INTERACTIONS BETWEEN LANGUAGE- AND MUSIC-SYNTACTIC PROCESSING

As mentioned above, hierarchical processing of syntactic information from different domains (such as music and language) requires contributions from neural populations located in BA 44. However, it is still possible that, although such neural populations are located in the same brain area, entirely different (non-overlapping) neural populations serve the syntactic processing of music and language *within the same area*. That is, perhaps the neural populations mediating language-syntactic processing in BA 44 are different from neural populations mediating music-syntactic processing in the same area. Therefore, the strongest evidence for shared neural resources for the syntactic processing of music and language stems from experiments that revealed interactions between music-syntactic and language-syntactic processing (Koelsch et al., 2005b; Steinbeis

and Koelsch, 2008b; Fedorenko et al., 2009; Slevc et al., 2009). In these studies, chord sequences were presented simultaneously with visually presented sentences (for an example see **Figure 4**). Two of these studies (Koelsch et al., 2005b; Steinbeis and Koelsch, 2008b) used EEG, and three different sentence types: The first type was a syntactically correct sentence in which the occurrence of the final noun was semantically highly probable. The other two sentence types were modified versions of the first sentence type: Firstly, a sentence with a gender disagreement between the last word (noun) on the one hand, and the prenominal adjective as well as the definite article that preceded the adjective on the other; such gender disagreements elicit a left anterior negativity (LAN; Gunter et al., 2000). Secondly, a sentence in which the final noun was semantically less probable; such "low-cloze probability" words elicit an N400 reflecting the semantic processing of words (see also section on musical semantics below). Each chord sequence consisted of five chords (and each chord was presented together with a word, see **Figure 4**), one chord sequence type ending on a music-syntactically regular, and the other type ending on an incorrect chord function.

Both studies (Koelsch et al., 2005b; Steinbeis and Koelsch, 2008b) showed that the ERAN elicited by irregular chords interacted with the LAN elicited by linguistic (morpho-syntactic) violations (Koelsch et al., 2005b; Steinbeis and Koelsch, 2008b): The LAN elicited by words was reduced when the syntactically irregular word was presented simultaneously with a music-syntactically irregular chord (compared to when the irregular word was presented with a regular chord). In the study from Koelsch et al. (2005b) a control experiment was conducted in which the same sentences were presented simultaneously with sequences of single tones. The



**FIGURE 4 | Examples of experimental stimuli used in the studies by** Koelsch et al. (2005b) **and** Steinbeis and Koelsch (2008b). Top: examples of two chord sequences in C major, ending on a regular (upper row) and an irregular chord (lower row, the irregular chord is indicated by the arrow). Bottom: examples of the three different sentence types. Onsets of chords (presented auditorily) and words (presented visually) were synchronous. Reprinted from Steinbeis and Koelsch (2008b).

tone sequences ended either on a standard tone or on a frequency deviant. The physical mismatch negativity (phMMN) elicited by the frequency deviants did not interact with the LAN (in contrast to the ERAN), indicating that the processing of auditory oddballs (as reflected in the phMMN) does not consume resources related to syntactic processing[16]. In addition, the study by Steinbeis and Koelsch (2008b) also showed that the amplitude of the ERAN was reduced when participants processed a syntactically wrong word (**Figure 5A**)[17]. Notably, the ERAN was not affected when words were semantically incongruous (**Figure 5B**), showing that the interaction between ERAN and LAN is specific for syntactic processing (the double interaction between ERAN and syntax, as well as between N5 and semantics, is shown in **Figure 5C**).

The findings of these EEG studies were corroborated by two behavioral studies (Fedorenko et al., 2009; Slevc et al., 2009), as well as by a patient study by Patel et al. (2008). The latter study showed that individuals with Broca's aphasia also show impaired music-syntactic processing in response to out-of-key

---

[16]Whether the abstract-feature MMN consumes such resources remains to be investigated.

[17]The fact that the ERAN was not reduced by the syntactically incorrect words in the study by Koelsch et al. (2005b) is presumably due to the task: Participants focused more attention on the musical stimulus in the study by Steinbeis and Koelsch (2008b), leading to a larger ERAN amplitude (and thus to a better signal-to-noise ratio).
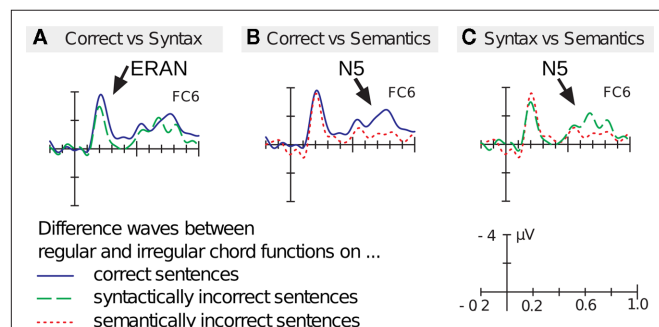


**FIGURE 5 | Grand-average ERPs elicited by the stimuli shown in Figure 4.** Participants monitored whether the sentences were (syntactically and semantically) correct or incorrect; in addition, they had to attend to the timber of the chord sequences and to detect infrequently occurring timber deviants. ERPs were recorded on the final chords/words and are shown for the different word conditions (note that only difference waves are shown). **(A)** The solid (blue) difference wave shows ERAN (indicated by the arrow) and N5 elicited on syntactically and semantically correct words. The dashed (green) difference wave shows ERAN and N5, elicited when chords are presented on morpho-syntactically incorrect (but semantically correct) words. Under the latter condition, the ERAN (but not the N5) is reduced. **(B)** The solid (blue) difference wave is identical to the solid difference wave of **(A)**, showing the ERAN and the N5 (indicated by the arrow) elicited on syntactically and semantically correct words. The dotted (red) difference wave shows ERAN and N5, elicited when chords are presented on semantically incorrect (but morpho-syntactically correct words). Under the latter condition, the N5 (but not the ERAN) is reduced. **(C)** shows the direct comparison of the difference waves in which words were syntactically incorrect (dashed, green line) or semantically incorrect (dotted, red line). These ERPs show that the ERAN is influenced by the morpho-syntactic processing of words, but not by the semantic processing of words. By contrast, the N5 is influenced by the semantic processing of words, but not by the morpho-syntactic processing of words. Data from Steinbeis and Koelsch (2008b).

chords occurring in harmonic progressions (note that all patients had Broca's aphasia, but only some of them had a lesion that included Broca's area).

It is also interesting to note that there are hints for an interaction between the ERAN and the *early left anterior negativity* (ELAN, an ERP component taken to reflect the processing of phrase structure violations during language perception; Friederici, 2002): In a study by Maidhof and Koelsch (2011), participants were presented simultaneously with spoken sentences and chord sequences. Thus, the experiment was similar to the studies by Koelsch et al. (2005b) and Steinbeis and Koelsch (2008b), except that (a) sentences were presented auditorily, (b) that syntactic violations were phrase-structure violations (not morpho-syntactic violations), and (c) that the attention of listeners was directed to either the speech or the music. In that study, the amplitude of the ERAN was slightly smaller when elicited during the presentation of phrase-structure violations occurring in sentences (compared to when elicited during correct sentences); this effect was, however, only marginally significant. Thus, it appears that music-syntactic processing of an irregular chord function (as reflected in the ERAN) also interacts with the processing of the phrase-structure of a sentence, although this interaction seems less obvious than the interaction between ERAN and LAN (i.e., between music-syntactic processing of chord functions and the processing of morpho-syntactic information in language).

In summary, neurophysiological studies show that music- and language-syntactic processes engage overlapping resources (presumably located in the inferior fronto-lateral cortex), and evidence showing that these resources underlie music- and language-syntactic processing is provided by experiments showing interactions between ERP components reflecting music- and language-syntactic processing (in particular LAN and ERAN). Importantly, such interactions are observed in the absence of interactions between LAN and MMN, i.e., in the absence between language-syntactic and acoustic deviance processing (reflected in the MMN), and in the absence of interactions between the ERAN and the N400 (i.e., in the absence of music-syntactic and language–semantic processing). Therefore, the reported interactions between LAN and ERAN are syntax-specific, and cannot be observed in response to any kind of irregularity.

## 5.4 PROCESSING OF PHRASE BOUNDARIES

During auditory music or language perception, the recognition of phrase boundaries helps to decode the syntactic (phrase) structure of a musical phrase, as well as of a sentence (the melodic and rhythmic features that mark a phrase boundary in a spoken sentence are part of the *speech prosody*). The processing of phrase boundaries is reflected in a *closure positive shift* (CPS), during both the processing of musical phrase boundaries (Knösche et al., 2005), and the processing of intonational phrase boundaries during speech perception (Steinhauer et al., 1999; although perhaps with a slightly different latency). The first study on the perception of musical phrase boundaries by Knösche et al. (2005) investigated musicians only, and a subsequent study also reported a CPS in non-musicians (although the CPS had a considerably smaller amplitude than the CPS observed in musicians; Neuhaus et al., 2006). The latter study also reported larger CPS amplitudes for longer pauses as well as longer boundary tones (which make the phrase boundary more

salient). Both studies (Knösche et al., 2005; Neuhaus et al., 2006) used EEG as well as MEG, showing that the CPS can be observed with both methods. In a third study, Chinese and German musicians performed a categorization task with Chinese and Western music (with both groups being familiar with Western music, but only the Chinese group being familiar with Chinese music; Nan et al., 2006). Both groups showed CPS responses to both types of music (with no significant difference between groups). The exact contributions of the processes reflected in the CPS to syntactic processing remain to be specified.

Using fMRI, a study by Meyer et al. (2004) suggests that the processing of the prosodic aspects of the sentences used in the study by Steinhauer et al. (1999) involves premotor cortex of the (right) Rolandic operculum, (right) AC located in the planum temporale, as well as the anterior insula (or perhaps deep frontal operculum) and the striatum bilaterally[18].

A very similar activation pattern was also observed for the processing of the melodic contour of spoken main clauses (Meyer et al., 2002). That is, similar to the right-hemispheric weighting of activations observed in functional neuroimaging studies on music perception, a right-hemispheric weighting of (neo-cortical) activations is also observed for the processing of speech melody.

## 6 STRUCTURAL REANALYSIS AND REPAIR

Following a syntactic anomaly (or an unexpected syntactic structure), processes of structural reanalysis and repair may be engaged. It appears that these processes are reflected in the ERP as positive potentials that are maximal around 600–900 ms (often referred to as P600, or *late positive component*, LPC; Besson and Schön, 2001).

The P600/LPC can be elicited by irregular melody-tones (e.g., Besson and Macar, 1986; Verleger, 1990; Paller et al., 1992; Besson and Faita, 1995; Besson et al., 1998; Miranda and Ullman, 2007; Peretz et al., 2009), as well as by irregular chords (e.g., Patel et al., 1998). It seems that the P600/LPC can only be elicited when individuals attend to the musical stimulus, and that the LPC is partly connected to processes of the conscious detection of music-structural incongruities. In an experiment by Besson and Faita (1995), in which incongruous melody-endings had to be detected, the LPCs had a greater amplitude, and a shorter latency, in musicians compared to "non-musicians," presumably because musicians were more familiar with the melodies than non-musicians (thus, musicians could detect the irregular events more easily). Moreover, the LPC had a larger amplitude for familiar melodies than for novel melodies (presumably because incongruous endings of familiar phrases were easier to detect), and for non-diatonic than for diatonic endings. Diatonic incongruities terminating unfamiliar melodies did not elicit an LPC (presumably because they are hardly to detect), whereas non-diatonic incongruities did (they were detectable for participants by the application of tonal rules).

It is remarkable that in *all* of the mentioned studies on melodies (Besson and Macar, 1986; Verleger, 1990; Paller et al., 1992; Besson and Faita, 1995; Besson et al., 1998; Brattico et al., 2006; Miranda and Ullman, 2007; Peretz et al., 2009), the unexpected tones also elicited an early frontal negative ERP (emerging around the N1

---

[18]Sentences were stripped of phonological information by filtering, and sounded like hummed sentences.

peak, i.e., around 100 ms after stimulus onset). This ERP effect resembles the ERAN (or, more specifically, an early component of the ERAN; for details see Koelsch and Jentschke, 2010), although it presumably overlapped in part with a subsequent N2b due to the detection of irregular or unexpected tones. That is, earlier processes underlying the detection of structural irregularities (as partly reflected in the ERAN) are followed by later processes of structural reanalysis (as reflected in the P600/LPC) when individuals attend to the musical stimulus and detect structural incongruities. It remains to be specified whether the P600/LPC is a late P3, and how the processes of structural reanalysis and repair are possibly related to context-updating (see also Donchin and Coles, 1998; Polich, 2007).

Using sentences and chord sequences, Patel et al. (1998) compared ERPs elicited by "syntactic incongruities" in language and music within subjects (harmonic incongruities were taken as grammatical incongruity in music). Task-relevant (target) chords within homophonic musical phrases were manipulated, so that the targets were either within the key of a phrase, or out-of-key (from a "nearby," or a "distant" key). Both musical and linguistic structural incongruities elicited P600/LPC potentials which were maximal at posterior sites and statistically indistinguishable. Moreover, the degree of a structural anomaly (moderate or high) was reflected in the amplitude of the elicited positivities. Therefore, the results indicated that the P600 reflects more general knowledge-based structural (re-)integration (and/or reanalysis) during the perception of rule-governed sequences. The intersection between the processes of music-syntactic and language-syntactic (re-)integration and repair is referred to as the *shared syntactic integration resource hypothesis* (SSIRH), which states that music and language rely on shared, limited processing resources that activate separable syntactic representations (Patel, 2003). The SSIRH also states that, whereas the linguistic and musical knowledge systems may be independent, the system used for online structural integration may be shared between language and music (see also Fedorenko et al., 2009). This system might be involved in the integration of incoming elements (words in language, and tones/chords in music) into evolving structures (sentences in language, harmonic sequences in music). Beyond the SSIRH, however, the previous section on music-syntactic processing also showed early interactions (between the ERAN and LAN), indicating that processing of music- and language-syntactic information intersects also at levels of morpho-syntactic processing, phrase-structure processing, and possibly word-category information.

## 7 PROCESSING MEANING IN MUSIC

Music is a means of communication, and during music listening, meaning emerges through the interpretation of (musical) information. Previous accounts on musical meaning can be summarized with regard to three fundamentally different classes of meaning emerging from musical information: extra-musical meaning, intra-musical meaning, and musicogenic meaning (for an extensive review see Koelsch, 2011). Extra-musical meaning refers to meaning emerging from reference to the extra-musical world (Meyer, 1956, referred to this class of musical meanings as "designative meaning"). This class of meaning comprises three dimensions: musical meaning due to iconic, indexical, and symbolic sign qualities (Karbusicky, 1986).

(a) Iconic musical meaning emerges from common patterns or forms, such as musical sound patterns that resemble sounds of objects, or qualities of objects. For example, acoustic events may sound "warm," "round," "sharp," or "colorful," and a musical passage may sound, e.g., "like a bird," or "like a thunderstorm." In linguistics, this sign quality is also referred to as *onomatopoetic*.

(b) Indexical musical meaning emerges from action-related patterns (such as movements and prosody) that index the presence of a psychological state, for example an emotion, or an intention. Juslin and Laukka (2003) compared in a meta analysis the acoustical signs of emotional expression in music and speech, finding that the acoustic properties that code emotional expression in speech are highly similar to those coding these expressions in music. With regard to intentions, an fMRI study by Steinbeis and Koelsch (2008c) showed that listeners automatically engage social cognition during listening to music, in an attempt to decode the intentions of the composer or performer (as indicated by activations of the cortical *theory-of-mind* network). That study also reported activations of posterior temporal regions implicated in semantic processing (Lau et al., 2008), presumably because the decoding of intentions has meaning quality.

Cross (2008) refers to this dimension of musical meaning as "motivational–structural" due to the relationship between affective–motivational states of individuals on the one side, and the structural–acoustical characteristics of (species-specific) vocalizations on the other.

(c) Symbolic musical meaning emerges from explicit (or conventional) extra-musical associations (e.g., any national anthem). Note that the meaning of the majority of words is due to symbolic meaning. Cross and Morley (2008) refers to this dimension of musical meaning as "culturally enactive," emphasizing that symbolic qualities of musical practice are shaped by (and shape) culture.

Musical semantics, notably, extends beyond the relations between concepts and the (extra-musical) world in that musical meaning also emerges from the reference of one musical element to another musical element, that is, from intra-musical combinations of formal structures. The term "intra-musical" was also used by Budd (1996), other theorists have used the terms "formal meaning" (Alperson, 1994), "formal significance" (Davies, 1994), or "embodied meaning" (Meyer, 1956). Similar distinctions as the one drawn here between extra- and intra-musical meaning have been made by several theorists (for a review see Koopman and Davies, 2001).

Processing of musical meaning information is reflected in (at least) two negative ERP components: The N400 and the N5. The N400 reflects processing of meaning in both language and music, the N5 has so far only been observed for the processing of musical information. In the following I will present evidence showing that, with regard to the processing of musical meaning, the N400 reflects processing of extra-musical meaning, and the N5 processing of intra-musical meaning.

The N400 has been used to investigate processing of musical meaning (or "musical semantics") in semantic priming paradigms (Koelsch et al., 2004; Steinbeis and Koelsch, 2008a, 2011; Daltrozzo
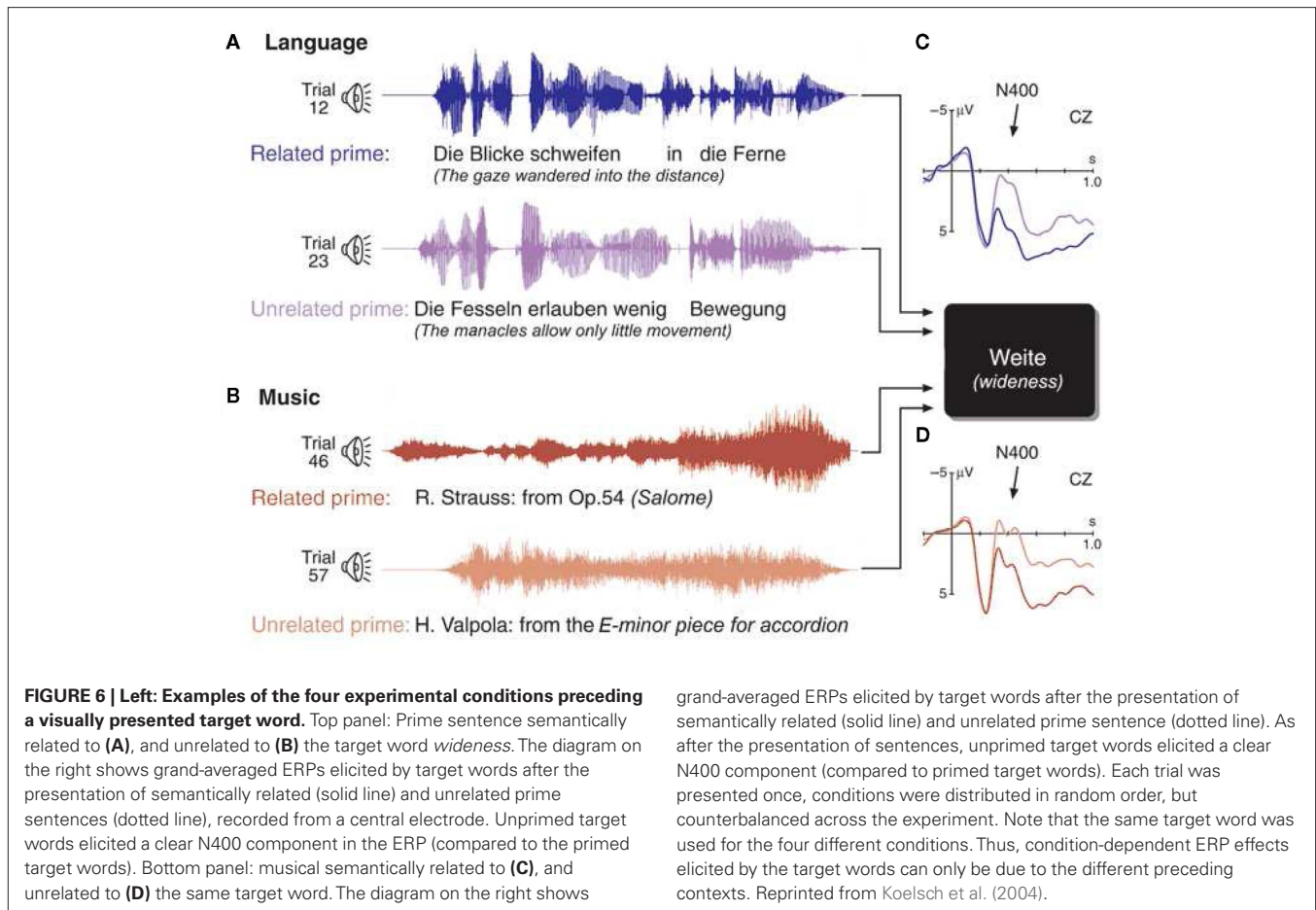
and Schön, 2009a; Grieser-Painter and Koelsch, 2011). In an initial study (Koelsch et al., 2004), sentences and musical excerpts were presented as prime stimuli. The prime stimuli were semantically either related or unrelated to a target word that followed the prime stimulus (see top of **Figure 6**). For example, the sentence "The gaze wandered into the distance" primes the word "wideness" (semantically related), rather than the word "narrowness" (semantically unrelated). Analogously, certain musical passages, for example from a symphony by Mozart, prime the word "angel," rather than the word "scallywag."

In the language condition (i.e., when target words followed the presentation of sentences), unrelated words elicited a clear N400 effect (this is a classical semantic priming effect). This semantic priming effect was also observed when target words followed musical excerpts. That is, target words that were semantically unrelated to a preceding musical excerpt also elicited a clear N400. The N400 effects did not differ between the language condition (in which the target words followed sentences) and the music condition (in which the target words followed musical excerpts), neither with respect to amplitude nor with respect to latency or scalp distribution. A source analysis localized the main sources of the N400 effects, in both conditions, in the posterior part of the medial temporal gyrus bilaterally (BA 21/37), in proximity to the superior temporal sulcus. These regions have been implicated in the processing of semantic information during language processing (Lau et al., 2008).
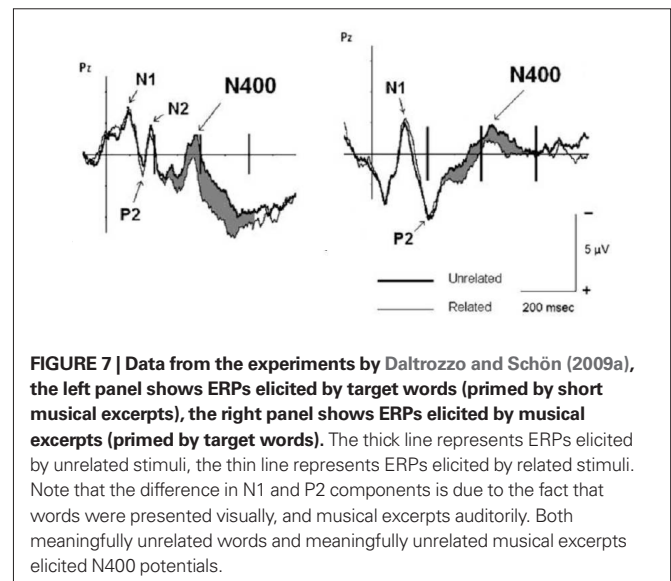
The N400 effect in the music condition demonstrates that musical information can have a systematic influence on the semantic processing of words. The N400 effects did not differ between the music and the language condition, indicating that musical and linguistic priming can have the same effects on the semantic processing of words. That is, the data demonstrated that music can activate representations of meaningful concepts (and that, thus, music is capable of transferring considerably more meaningful information than previously believed), and that the cognitive operations that decode meaningful information while listening to music can be identical to those that serve semantic processing during language perception. The N400 effect was observed for both abstract and concrete words, showing that music can convey both abstract and concrete semantic information. Moreover, effects were also observed when emotional relationships between prime and target words were balanced, indicating that music does not only transfer emotional information.

The priming of meaning by the musical information was due to (a) iconic sign qualities, i.e., common patterns or forms (such as succeeding interval steps priming the word staircase), (b) indexical sign quality, such as the suggestion of a particular emotion due to the resemblance to movements and prosody typical for that emotion (e.g., saxophone tones sounding like derisive laughter), or (c) symbolic sign quality due to meaning inferred by explicit extra-musical associations (e.g., a church anthem priming the word devotion). Unfortunately, is was not investigated in that study whether N400 responses differed between these dimensions of musical meaning (and, so far, no subsequent study has investigated this). However, the results still allow to conclude that processing of extra-musical meaning is associated with N400 effects.

Due to the length of musical excerpts (~10 s), musical information could not be used as target stimulus, thus only words were used as target stimuli (and the N400 was only elicited on words, not on musical information). Hence, a question emerging from that study

**FIGURE 6 | Left: Examples of the four experimental conditions preceding a visually presented target word.** Top panel: Prime sentence semantically related to **(A)**, and unrelated to **(B)** the target word *wideness*. The diagram on the right shows grand-averaged ERPs elicited by target words after the presentation of semantically related (solid line) and unrelated prime sentences (dotted line), recorded from a central electrode. Unprimed target words elicited a clear N400 component in the ERP (compared to the primed target words). Bottom panel: musical semantically related to **(C)**, and unrelated to **(D)** the same target word. The diagram on the right shows

grand-averaged ERPs elicited by target words after the presentation of semantically related (solid line) and unrelated prime sentence (dotted line). As after the presentation of sentences, unprimed target words elicited a clear N400 component (compared to primed target words). Each trial was presented once, conditions were distributed in random order, but counterbalanced across the experiment. Note that the same target word was used for the four different conditions. Thus, condition-dependent ERP effects elicited by the target words can only be due to the different preceding contexts. Reprinted from Koelsch et al. (2004).

was whether musical information can also elicit N400 responses. One study addressing this issue (Daltrozzo and Schön, 2009a) used short musical excerpts (duration was ~1 s) that could be used either as primes (in combination with word targets), or as targets (in combination with word primes). **Figure 7** shows that when the musical excerpts were used as primes, meaningfully unrelated target words elicited an N400 (compared to related target words, as in the study from Koelsch et al., 2004). Importantly, when musical excerpts were used as target stimuli (and words as primes), an N400 was observed in response to excerpts that the participants rated as meaningfully unrelated to the preceding target word (compared to excerpts that were rated as related to the preceding word). This was the first evidence that musical information can also elicit N400 responses. Note that the musical excerpts were composed for the experiment, and thus not known by the participants. Therefore, musical meaning was not due to symbolic meaning, but due to indexical (e.g., "happy") and iconic (e.g., "light") meaning. In the data analysis used in that study (Daltrozzo and Schön, 2009a), the relatedness of prime–target pairs was based on the relatedness judgments of participants. In another article (Daltrozzo and Schön, 2009b) the author showed that even if the data are analyzed based on the un/relatedness of prime–target pairs as pre-defined by the experimenters, a significant N400 was elicited (in the latter study a lexical decision task was used, whereas in the former study participants were asked to judge the conceptual relatedness of prime and targets).



**FIGURE 7 | Data from the experiments by Daltrozzo and Schön (2009a),** the left panel shows ERPs elicited by target words (primed by short musical excerpts), the right panel shows ERPs elicited by musical excerpts (primed by target words). The thick line represents ERPs elicited by unrelated stimuli, the thin line represents ERPs elicited by related stimuli. Note that the difference in N1 and P2 components is due to the fact that words were presented visually, and musical excerpts auditorily. Both meaningfully unrelated words and meaningfully unrelated musical excerpts elicited N400 potentials.

Further studies investigated the processing of musical meaning using only single chords (Steinbeis and Koelsch, 2008a, 2011) or single tones (Grieser-Painter and Koelsch, 2011). These studies showed that even single chords and tones can prime the meaning of words

(in both musicians and non-musicians), and that even chords (in musicians) and tones (in non-musicians and therefore presumably also in musicians) can elicit N400 effects.

## INTRA-MUSICAL MEANING AND THE N5

The previous section dealt with the N400 and extra-musical meaning. However, musical meaning can also emerge from intra-musical references, that is, from the reference of one musical element to at least one other musical element (for example, a *G* major chord is usually perceived as the tonic in *G* major, as the dominant in *C* major, and – in its first inversion – possibly as a Neapolitan sixth chord in *F#* minor). This section will review empirical evidence for this hypothesis, and describe an electrophysiological correlate of the processing of intra-musical meaning (the so-called N5, or N500). The N5 was described first in reports of experiments using chord sequence paradigms with music-syntactically regular and irregular chord functions (Koelsch, 2000), in which the ERAN was usually followed by a late negativity, the N5 (see also **Figure 2C**). Initially (Koelsch, 2000) the N5 was proposed to reflect processes of harmonic integration, reminiscent of the N400 reflecting semantic integration of words, and therefore proposed to be related to the processing of musical meaning, or musical semantics, although the type of musical meaning had remained unclear.

One reason for this proposition was a remarkable resemblance between N5 and N400: *Firstly*, similar to the N400 elicited by open-class words, which declines toward the end of sentences (Van Petten and Kutas, 1990), the amplitude of the N5 (elicited by regular chords) declined toward the end of a chord sequence (Koelsch, 2000). That is, during sentence processing, a semantically correct final open-class word usually elicits a rather small N400, whereas the open-class words preceding this word elicit larger N400 potentials. This is due to the semantic expectedness of words, which is rather unspecific at the beginning of a sentence, and which becomes more and more specific toward the end of the sentence (where people already have a hunch of what the last word will actually be). Thus, a smaller amount of semantic integration is required at the end of a sentence, reflected in a smaller N400. If the last word is semantically unexpected, then a large amount of semantics is required, which is reflected in a larger amplitude of the N400. Similarly, the N5 is not only elicited by irregular chords, but also by regular chords, and the amplitude of the N5 decreases with each progressing chord of the chord sequence. A small N5 elicited by the (expected) final chord of a chord sequence presumably reflects that only a small amount of harmonic integration is required at this position of a chord sequence.

*Secondly*, at the same position within a chord sequence (or a sentence) the N5 (or the N400) is modulated by the degree of fit with regard to the previous harmonic (or semantic) context. A wealth of studies has shown that irregular chords (for reviews see Koelsch, 2004, 2009b) and irregular tones of melodies (e.g., Miranda and Ullman, 2007; Koelsch and Jentschke, 2010) evoke larger N5 potentials than regular ones (see also **Figure 2C**). As mentioned above, the ERAN is related to syntactic processing. In addition, the experiment from Steinbeis and Koelsch (2008b) also showed that the N5 reflects processing of meaning information: In that study, the N5 interacted with the N400 elicited by words with low semantic cloze probability. The N5 was smaller when elicited on words that were semantically less probable ("He sees the cold beer") compared to when elicited on words that were semantically highly probable ("He drinks the cold

beer"; see **Figure 5B**). Importantly, the N5 did not interact with the LAN (i.e., the N5 did not interact with the syntactic processing of words), indicating that the N5 is not simply modulated by any type of deviance, or incongruency, but that the N5 is specifically modulated by neural mechanisms underlying semantic information processing. That is, the N5 potential can be modulated by semantic processes, namely by the activation of lexical representations of words with different semantic fit to a previous context. This modulation indicates that the N5 is related to the processing of meaning. Note that the harmonic relation between the chord functions of a harmonic sequence is an intra-musical reference (i.e., a reference of one musical element to another musical element, and not a reference to anything belonging to the extra-musical world). Therefore, we have reason to believe that the N5 reflects the processing of intra-musical meaning.

The neural generators of the N5 have remained elusive. This is in part due to the difficulty that in most experiments the N5 follows the ERAN (but see also, Poulin-Charronnat et al., 2006), making it difficult to differentiate the neural correlates of N5 and ERAN in experiments using functional magnetic resonance imaging. The N5 usually has a clear frontal scalp distribution, thus the scalp distribution of the N5 is more anterior than that of the N400, suggesting at least partly different neural correlates. Perhaps the N5 originates from combined sources in the temporal lobe (possibly overlapping with those of the N400 in BAs 21/37) and the frontal lobe (possibly in the posterior part of the inferior frontal gyrus). This needs to be specified, for example using EEG source localization in a study that compares an auditory N4 with an auditory N5 within subjects.

The third class of musical meaning (musicogenic meaning) emerges from individual responses to musical information, particularly movement, emotional responses, and self-related memory associations. This dimension of musical meaning, as well as details about the extra- and intra-musical dimensions of musical meaning, are discussed in more detail elsewhere (Koelsch, 2011).

## 8 HOW THE BODY REACTS TO MUSIC

The present model of music perception also takes the potential "vitalization" of an individual into account: vitalization entails activity of the autonomic nervous system (i.e., regulation of sympathetic and parasympathetic activity) along with the cognitive integration of musical and non-musical information. Non-musical information comprises associations evoked by the music, as well as emotional (e.g., happy), and bodily reactions (e.g., tensioned or relaxed). The *subjective feeling* (e.g., Scherer, 2005) requires conscious awareness, and therefore presumably involves multimodal association cortices such as parietal association cortices in the region of BA 7 (here, the musical percept might also become conscious; Block, 2005). Effects of music perception on activity of the autonomic nervous system have mainly been investigated by measuring electrodermal activity and heart rate, as well as the number and intensity of reported "shivers" and "chills" (Sloboda, 1991; Blood and Zatorre, 2001; Khalfa et al., 2002; Panksepp and Bernatzky, 2002; Grewe et al., 2007a,b; Lundqvist et al., 2009; Orini et al., 2010).

Vitalizing processes can, in turn, have an influence on processes within the immune system. Effects of music processing on the immune system have been assessed by measuring variations of (salivary) immunoglobulin A concentrations (e.g., McCraty et al., 1996; Hucklebridge et al., 2000; Kreutz et al., 2004). Interestingly, effects

on the immune system have been suggested to be tied to motor activity such as singing (Kreutz et al., 2004) or dancing (Quiroga Murcia et al., 2009, see also **Figure 1**, rightmost box). With regard to music perception, it is important to note that there might be overlap between neural activities of the late stages of perception and those related to the early stages of action (such as premotor functions related to action planning; Janata et al., 2002b; Rizzolatti and Craighero, 2004).

Music perception can interfere with action planning in musicians (Drost et al., 2005a,b), and premotor activity can be observed during the perception of music (a) in pianists listening to piano pieces (Haueisen and Knösche, 2001), (b) in non-musicians listening to song (Callan et al., 2006), (c) in non-musicians who received 1 week of piano training and listened to the trained piano melody (Lahav et al., 2007; for a detailed review see Koelsch, 2009a). For neuroscience studies related to music production see, e.g., Bangert and Altenmüller (2003), Katahira et al. (2008), Maidhof et al. (2009), Kamiyama et al. (2010), Maidhof et al. (2010).

Movement induction by music perception in the way of dancing, singing, tapping, hopping, swaying, head-nodding, etc., along with music is a very common experience (Panksepp and Bernatzky, 2002); such movements also serve social functions, because synchronized movements of different individuals represent coordinated social activity. Notably, humans have a need to engage in social activities; emotional effects of such engagement include fun, joy, and happiness, whereas exclusion from this engagement represents an emotional stressor, and has deleterious effects on health. Humans making music is an activity involving several social functions, which were recently summarized as the "Seven Cs" (Koelsch, 2010): (1) When we make music, we make *contact* with other individuals (preventing from social isolation). (2) Music automatically engages *social cognition* (Steinbeis and Koelsch, 2008c). (3) Music engages *co-pathy* in the sense that inter-individual emotional states become more homogenous (e.g., reducing anger in one individual, and depression or anxiety in another), thus promoting inter-individual understanding and decrease of conflicts. (4) Music involves *communication* (notably, for infants and young children, musical communication during parent–child singing of lullabies and play-songs is important for social and emotional regulation, as well as for social, emotional, and cognitive development; Trehub, 2003; Fitch, 2006). (5) Music making also involves coordination of movements (requiring the capability to synchronize movements to an external beat; see also Kirschner and Tomasello, 2009; Overy and Molnar-Szakacs, 2009; Patel et al., 2009). The coordination of movements in a group of individuals appears to be associated with pleasure (for example, when dancing together), even in the absence of an explicit shared goal (apart from deriving pleasure from concerted movements). (6) Performing music also requires *cooperation* (involving a shared goal, and increasing inter-individual trust); notably, engaging in cooperative behavior is an important potential source of pleasure Rilling et al. (2002). (7) As an effect, music leads to increased *social cohesion* of a group (Cross and Morley, 2008), fulfilling the "need to belong" (Baumeister and Leary, 1995), and the motivation to form and maintain interpersonal attachments. Social cohesion also strengthens the confidence in reciprocal care (see also the caregiver hypothesis; Trehub, 2003), and the confidence that opportunities to engage with others in the mentioned social functions will also emerge in the future. Music seems to be capable of engaging all of the "Seven Cs" at the same time, which is presumably part of the emotional power of music. These evolutionarily advantageous social aspects of music-making behavior represent one origin for the evolution of music-making behavior in humans.

Action induction by music perception is accompanied by neural impulses in the reticular formation (in the brainstem; for example, for the release of energy to move during joyful excitement). It is highly likely that connections also exist between the reticular formation and structures of the auditory brainstem (as well as between reticular formation and the AC; Levitt and Moore, 1979), and that the neural activity of the reticular formation therefore also influences the processing of (new) incoming acoustic information.

## 9 MUSIC PERCEPTION AND MEMORY

The modules presented in **Figure 1** are associated with a variety of memory functions (for a review see Jäncke, 2008). For example, the auditory sensory memory (along with Gestalt formation) is connected with both WM (Berti and Schröger, 2003) and long-term memory (Näätänen et al., 2001, see above for information about brain structures implicated in auditory sensory memory). Structure building requires WM as well as a long-term store for syntactic regularities, and processing of meaning information is presumably tied to a mental "lexicon" (containing conceptual–semantic knowledge), as well as to a musical "lexicon" containing knowledge about timbers, melodic contours, phrases, and musical pieces (Peretz and Coltheart, 2003). However, the details about interconnections between the different modules and different memory functions remain to be specified.

Neuroimaging studies suggest that the phonological loop of verbal WM and the "tonal loop" of WM for pitch strongly overlap in non-musicians (Hickok et al., 2003; Koelsch et al., 2009; Schulze et al., 2011a), involving PMCv (encroaching Broca's area), dorsal premotor cortex, the planum temporale, inferior parietal lobe, the anterior insula, subcortical structures (basal ganglia and thalamus), as well as the cerebellum. A study by Schulze et al. (2011b) showed that, in contrast to non-musicians, musicians use specific neural subcomponents of WM only during verbal (right insular cortex) or only during tonal WM (right globus pallidus, right caudate nucleus, and left cerebellum). These results revealed the existence of two WM systems in musicians: A phonological loop supporting rehearsal of phonological information, and a tonal loop supporting rehearsal of tonal information. Note that differences between non-musicians and musicians for tonal WM (and between verbal and tonal WM within musicians), were mainly related to structures involved in controlling, programming, and planning of actions, thus presumably reflecting differences in action-related sensorimotor coding of verbal and tonal information. That is, verbal and tonal WM rely strongly on action-related coding (and in this regard sensorimotor coding forms one basis for cognition; for a study on strategic WM processes in musicians see Schulze et al., 2011a).

The knowledge about musical long-term memory is still rather limited. Functional neuroimaging data suggest that access to musical semantic memory involves the (left) middle temporal gyrus, and that musical semantic representations (that is, parts of a musical lexicon) are stored in (left) anterior-temporal areas (Groussard et al., 2009, 2010). In addition, Watanabe et al. (2008) reported that

retrieval of musical information involves the hippocampal formation and the inferior frontal gyrus. Further systematic research in these areas is needed to differentiate memory operations from operations of the modules described in **Figure 1**; it is important to bear in mind that, especially in functional imaging experiments, both types of operations usually co-occur.

## 10 MUSIC AND LANGUAGE

The previous sections provided a number of examples for cognitive processes (and their neural correlates) underlying the processing of music as well as of (spoken) language. Decoding of both music and speech information requires a fine-grained analysis of the spectral and temporal features of acoustic information (boxes "Feature Extraction" in **Figure 1**). The section on auditory feature extraction mentioned the acoustic equivalence of timbre and phoneme, and the *identification* of "phonemes" in language is thus presumably paralleled by the identification of "timbres" in music. However, the segmentation of phonemic information during language perception usually requires a higher temporal resolution compared to the music (because timbral information in music usually does not change as rapidly as phonemic information in language). This probably leads to the left-hemispheric weighting for the segmentation of phonemes during language perception, whereas segmentation of spectral information such as melodic information of speech prosody or musical melodies engages the right AC more strongly than the left AC.

Adequate processing of both music and speech also requires auditory sensory memory, and auditory scene analysis/auditory stream segregation (see boxes "Auditory sensory memory" and "Gestalt formation" in **Figure 1**), particularly in noisy environments (which are more common than the quiet auditory environments typical for experiments on language or music perception)[19]. The processes mediating the identification of chords (e.g., whether a chord is a major or a minor chord, whether a chord is presented in root position or in inversion, etc.; see box "Analysis of intervals" in **Figure 1**) perhaps parallel those underlying the identification of word form (due to both words and chords having a stem, or "root," from which different versions emerge); however, this remains to be specified.

The section on music-syntactic processing (see also box "Syntactic structure building" in **Figure 1**) presented evidence for interactions between the processing of music- and language-syntactic information at levels of morpho-syntactic processing, phrase-structure processing, and possibly word-category information (see the evidence for the interaction between the ERAN and LAN, and the provisional evidence for an interaction between ERAN and ELAN). Beyond these early interactions between music- and language-syntactic processes, it appears that cognitive and neural resources are also shared during later stages of syntactic integration (see box "Structural reanalysis and repair" in **Figure 1**). The notion of shared resources during these later stages is also referred to as SSIRH (Patel, 2003).

Both speech and music perception involve premotor coding [e.g., Liberman and Mattingly, 1985; Koelsch, 2009a, see also box "(premotor) Action" in **Figure 1**], and both music and language give rise to affective processes: With regard to language, the perception

of affective prosody (e.g., Ethofer et al., 2009; Wittfoth et al., 2010) as well as the affective contents of words (e.g., Herbert et al., 2009; Võ et al., 2009) can elicit emotional responses; the neural correlates of music-evoked emotions have been reviewed elsewhere (e.g., Koelsch, 2010).

Finally, this review illustrated that communication of meaning is not exclusively a linguistic domain, but that music can also convey meaningful information: The interpretation of extra-musical sign qualities of musical information can prime representations of meaningful concepts, and such priming can emerge from different cognitive levels (e.g., from auditory feature extraction, or from a more fine-grained analysis of intervals, see also **Figure 1**). Moreover, structural relations give rise to intra-musical meaning (suggested to be reflected in the N500).

Corresponding to these shared processes, there is considerable resemblance, overlap, and interaction between ERP components (and their neural generators) elicited during the perception of music or language: Processing of musical as well as of speech information evokes (a) FFRs originating from the auditory brainstem, (b) P1, N1, and P2 potentials originating from the AC, (c) MMN potentials originating from temporal and frontal cortical areas, (d) potentials of early syntactic processing that interact with each other (ERAN/LAN), and receive main contributions from the (inferior) pars opercularis (BA 44i), (e) potentials of syntactic (re)integration/reanalysis and repair (P600), and (f) N400 effects reflecting semantic processing (probably originating from posterior temporal and inferior frontal cortex).

These overlaps and shared cognitive/neural processes call for more integrative models that synthesize the cognitive processes underlying music and language perception into a common theoretical framework. For example, a comparison of the present model (**Figure 1**) with the language perception model by Friederici (2002) shows several analogies, such as "Feature Extraction"/"Primary acoustic analysis," "Analysis of intervals (chords, melodies, time intervals)"/"Identification of word form," "Syntactic structure building" (both models), "Reanalysis & repair" (both models), or "Meaning"/"Semantic relations." Additional overlaps include WM for verbal and tonal information (as described in the previous section), and processing of rhythm in speech and music (which is not included in this review, but see, e.g., Grahn and Brett, 2007; Schmidt-Kassow and Kotz, 2009).

The illustrated overlaps between music and language also show that "language" and "music" are different aspects of the same domain, or two poles of a rather continuous dimension, rather than being two strictly separate domains (I refer to this continuum as the "music-language-continuum). Spoken language has rhythm, melody, and timbre, and the illusory transformation from speech to song described by Diana Deutsch ("sometimes they behave so strangely"; Deutsch et al., 2011) shows that humans can perceive spoken language also as song (although individuals often do not realize this in everyday life). Once an individual puts emphasis in his/her utterances, the speech becomes more song-like[20], and many art-forms, such as Rap-music or Recitatives, are both song and speech. In addition, as described in this review, music is often structured

---

[19]For a study showing topological shifts in the activation patterns of language perception due to speech degradation (noise-band vocoding) see Obleser et al. (2011).

[20]Martin Luther King's speeches are a nice example of how it is often difficult to say whether someone is singing or speaking.

according to a syntactic system, and during music listening, meaning emerges from the interpretation of musical information. Therefore, any clear-cut distinction between music and language (and thus also any pair of separate definitions for language and music) is likely to be inadequate, or incomplete, and a rather artificial construct[21].

As mentioned already above, even individuals without formal musical training show sophisticated abilities with regard to the decoding of musical information, the acquisition of knowledge about musical syntax, the processing of musical information according to that knowledge, and the understanding of music. This finding supports the notion that musicality is a natural ability

_____

[21]Similar to the distinction between perception and action, or motor and sensory processes, which do not occur in isolation in a living individual.

of the human brain. Such musical abilities are important for the acquisition and the processing of language: infants acquire information about word and phrase boundaries (possibly even about word meaning) in part through different types of prosodic cues. Moreover, tonal languages rely on a meticulous decoding of pitch information, and both tonal and non-tonal languages require an accurate analysis of speech prosody to decode structure and meaning of speech. The assumption of an intimate connection between music and speech is corroborated by the reviewed findings of overlapping and shared neural resources for music and language processing in both adults and children. These findings suggest that the human brain, particularly at an early age, does not treat language and music as strictly separate domains, but rather treats language as a special case of music.

## REFERENCES

Alain, C., Woods, D. L., and Knight, R. T. (1998). A distributed cortical network for auditory sensory memory in humans. *Brain Res.* 812, 23–37.

Alho, K., Tervaniemi, M., Huotilainen, M., Lavikainen, J., Tiitinen, H., Ilmoniemi, R. J., Knuutila, J., and Näätänen, R. (1996). Processing of complex sounds in the human auditory cortex as revealed by magnetic brain responses. *Psychophysiology* 33, 369–375.

Alperson, P. (1994). *What is Music? An Introduction to the Philosophy of Music.* Pennsylvania: State University Press.

Amunts, K., Lenzen, M., Friederici, A., Schleicher, A., Morosan, P., Palomero-Gallagher, N., and Zilles, K. (2010). Broca's region: novel organizational principles and multiple receptor mapping. *PLoS Biol.* 8, e1000489. doi: 10.1371/journal.pbio.1000489

Bangert, M., and Altenmüller, E. (2003). Mapping perception to action in piano practice: a longitudinal dc-eeg study. *BMC Neurosci.* 4, 26. doi: 10.1186/1471-2202-4-26

Baumeister, R., and Leary, M. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol. Bull.* 117, 497–497.

Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature* 436, 1161–1165.

Berti, S., and Schröger, E. (2003). Working memory controls involuntary attention switching: evidence from an auditory distraction paradigm. *Eur. J. Neurosci.* 17, 1119–1122.

Besson, M., and Faita, F. (1995). An event-related potential (ERP) study of musical expectancy: comparison of musicians with nonmusicians. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 1278–1296.

Besson, M., Faita, F., Peretz, I., Bonnel, A. M., and Requin, J. (1998). Singing in the brain: independence of lyrics and tunes. *Psychol. Sci.* 9, 494–498.

Besson, M., and Macar, F. (1986). Visual and auditory event-related potentials elicited by linguistic and non-linguistic incongruities. *Neurosci. Lett.* 63, 109–114.

Besson, M., and Schön, D. (2001). "Comparison between language and music," in *The Biological Foundations of Music*, Vol. 930, eds R. J. Zatorre and I. Peretz (New York: The New York Academy of Sciences), 232–258.

Bigand, E., Parncutt, R., and Lerdahl, J. (1996). Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Percept. Psychophys.* 58, 125–141.

Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn. Sci. (Regul. Ed.)* 9, 46–52.

Blood, A., and Zatorre, R. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11818.

Brattico, E., Tervaniemi, M., Naatanen, R., and Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Res.* 1117, 162–174.

Bregman, A. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound.* Cambridge, MA: The MIT Press.

Budd, M. (1996). *Values of Art.* London: Penguin Books.

Callan, D., Tsytsarev, V., Hanakawa, T., Callan, A., Katsuhara, M., Fukuyama, H., and Turner, R. (2006). Song and speech: brain regions involved with perception and covert production. *Neuroimage* 31, 1327–1342.

Cardoso, S., Coimbra, N., and Brandão, M. (1994). Defensive reactions evoked by activation of NMDA receptors in distinct sites of the inferior colliculus. *Behav. Brain Res.* 63, 17–24.

Carlyon, R. (2004). How the brain separates sounds. *Trends Cogn. Sci. (Regul. Ed.)* 8, 465–471.

Conard, N., Malina, M., and Münzel, S. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature* 460, 737–740.

Cook, N. (1987). The perception of large-scale tonal closure. *Music Percept.* 197–205.

Cross, I. (2008). The evolutionary nature of musical meaning. *Music. Sci.* 179–200.

Cross, I., and Morley, I. (2008). "The evolution of music: theories, definitions and the nature of the evidence," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 61–82.

Daltrozzo, J., and Schön, D. (2009a). Conceptual processing in music as revealed by N400 effects on words and musical targets. *J. Cogn. Neurosci.* 21, 1882–1892.

Daltrozzo, J., and Schön, D. (2009b). Is conceptual processing in music automatic? An electrophysiological approach. *Brain Res.* 1270, 88–94.

Darwin, C. (1997). Auditory grouping. *Trends Cogn. Sci. (Regul. Ed.)* 1, 327–333.

Darwin, C. (2008). Listening to speech in the presence of other sounds. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1011.

Davies, S. (1994). *Musical Meaning and Expression.* Ithaca, NY: Cornell University Press.

Deouell, L. (2007). The frontal generator of the mismatch negativity revisited. *J. Psychophysiol.* 21, 188.

Deutsch, D., Henthorn, T., and Lapidis, R. (2011). Illusory transformation from speech to song. *J. Acoust. Soc. Am.* 129, 2245–2252.

Di Pietro, M., Laganaro, M., Leemann, B., and Schnider, A. (2004). Receptive amusia: temporal auditory processing deficit in a professional musician following a left temporo-parietal lesion. *Neuropsychologia* 42, 868–877.

Doeller, C., Opitz, B., Mecklinger, A., Krick, C., Reith, W., and Schröger, E. (2003). Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *Neuroimage* 20, 1270–1282.

Donchin, E., and Coles, M. G. H. (1998). Context updating and the p300. *Behav. Brain Sci.* 21, 152.

Drost, U., Rieger, M., Brass, M., Gunter, T., and Prinz, W. (2005a). Action-effect coupling in pianists. *Psychol. Res.* 69, 233–241.

Drost, U., Rieger, M., Brass, M., Gunter, T., and Prinz, W. (2005b). When hearing turns into playing: movement induction by auditory stimuli in pianists. *Q. J. Exp. Psychol. A* 58, 1376–1389.

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., and Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *J. Cogn. Neurosci.* 21, 1255–1268.

Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A., Pozzo, T., Calzolari, F., Granieri, E., and Fadiga, L. (2009). Encoding of human action in Broca's area. *Brain* 132, 1980.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., and Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Mem. Cognit.* 37, 1.

Fitch, W. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215.

Fitch, W., and Hauser, M. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377.

Fodor, J., Mann, V., and Samuel, A. (1991). "Panel discussion: the mod-

ularity of speech and language," in *Modularity and the Motor Theory of Speech Perception: Proceedings of a Conference to Honor Alvin M. Liberman*, eds Ignatius G. Mattingly and Michael Studdert-Kennedy, (Hillsdale, NJ: Lawrence Erlbaum Associates), 359.

Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci. (Regul. Ed.)* 6, 78–84.

Friederici, A. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends Cogn. Sci. (Regul. Ed.)* 8, 245–247.

Friederici, A., Bahlmann, J., Heim, S., Schubotz, R., and Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2458.

Friedrich, R., and Friederici, A. (2009). Mathematical logic in the human brain: syntax. *PLoS ONE* 4, e5599. doi: 10.1371/journal.pone.0005599

Fujioka, T., Trainor, L., Ross, B., Kakigi, R., and Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *J. Cogn. Neurosci.* 16, 1010–1021.

Fujioka, T., Trainor, L., Ross, B., Kakigi, R., and Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J. Cogn. Neurosci.* 17, 1578–1592.

Geisler, C. (1998). *From Sound to Synapse: Physiology of the Mammalian Ear.* New York, NY: Oxford University Press.

Giard, M., Perrin, F., and Pernier, J. (1990). Brain generators implicated in processing of auditory stimulus deviance. a topographic ERP study. *Psychophysiology* 27, 627–640.

Grahn, J., and Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* 19, 893–906.

Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2007a). Emotions over time: synchronicity and development of subjective, physiological, and facial affective reactions of music. *Emotion* 7, 774–788.

Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2007b). Listening to music as a re-creative process: physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Percept.* 24, 297–314.

Grieser-Painter, J., and Koelsch, S. (2011). Can out-of-context musical sounds convey meaning? an ERP study on the processing of meaning in music. *Psychophysiology* 48, 645–655.

Griffiths, T., and Warren, J. (2002). The planum temporale as a computational hub. *Trends Neurosci.* 25, 348–353.

Griffiths, T., and Warren, J. (2004). What is an auditory object? *Nat. Rev. Neurosci.* 5, 887–892.

Groussard, M., Viader, F., Hubert, V., Landeau, B., Abbas, A., Desgranges, B., Eustache, F., and Platel, H. (2010). Musical and verbal semantic memory: two distinct neural networks? *Neuroimage*, 49, 2764–2773.

Groussard, M., Viader, F., Landeau, B., Desgranges, B., Eustache, F., and Platel, H. (2009). Neural correlates underlying musical semantic memory. *Ann. N. Y. Acad. Sci.* 1169, 278–281.

Gunter, T., Friederici, A., and Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J. Cogn. Neurosci.* 12, 556–568.

Hackett, T. A., and Kaas, J. (2004). "Auditory cortex in primates: functional subdivisions and processing streams," in *The Cognitive Neurosciences*, ed. M. S. Gazzaniga (Cambridge: MIT Press), 215–232.

Haueisen, J., and Knösche, T. (2001). Involuntary motor activity in pianists evoked by music perception. *J. Cogn. Neurosci.* 13, 786–792.

Herbert, C., Ethofer, T., Anders, S., Junghofer, M., Wildgruber, D., Grodd, W., and Kissler, J. (2009). Amygdala activation during reading of emotional adjectives – an advantage for pleasant content. *Soc. Cogn. Affect. Neurosci.* 4, 35.

Herrojo-Ruiz, M., Jabusch, H., and Altenmüller, E. (2009). Detecting wrong notes in advance: neuronal correlates of error monitoring in pianists. *Cereb. Cortex* 19, 2625.

Herrojo-Ruiz, M., Strübing, F., Jabusch, H. C., and Altenmüller, E. (2010). EEG oscillatory patterns are associated with error prediction during music performance and are altered in musician's dystonia. *Neuroimage* 55, 1791–1803.

Hickok, G., Buchsbaum, B., Humphries, C., and Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J. Cogn. Neurosci.* 15, 673–682.

Hucklebridge, F., Lambert, S., Clow, A., Warburton, D., Evans, P., and Sherwood, N. (2000). Modulation of secretory immunoglobulin A in saliva; response to manipulation of mood. *Biol. Psychol.* 53, 25–35.

Huffman, R., and Henson, O. (1990). The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. *Brain Res. Rev.* 15, 295–323.

Hyde, K., Peretz, I., and Zatorre, R. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia* 46, 632–639.

Janata, P., Birk, J., Van Horn, J., Leman, M., Tillmann, B., and Bharucha, J. (2002a). The cortical topography of tonal structures underlying Western music. *Science* 298, 2167.

Janata, P., Tillmann, B., and Bharucha, J. (2002b). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cogn. Affect. Behav. Neurosci.* 2, 121.

Janata, P., and Grafton, S. T. (2003). Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nat. Neurosci.* 6, 682–687.

Jäncke, L. (2008). Music, memory and emotion. *J. Biol.* 7, 21.

Johnson, K., Nicol, T., Zecker, S., and Kraus, N. (2008). Developmental plasticity in the human auditory brainstem. *J. Neurosci.* 28, 4000.

Johnsrude, I., Penhune, V., and Zatorre, R. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 123, 155.

Jusczyk, P. (1999). How infants begin to extract words from speech. *Trends Cogn. Sci. (Regul. Ed.)* 3, 323–328.

Juslin, P., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814.

Kaas, J., and Hackett, T. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11793.

Kaas, J., Hackett, T., and Tramo, M. (1999). Auditory processing in primate cerebral cortex. *Curr. Opin. Neurobiol.* 9, 164–170.

Kamiyama, K., Katahira, K., Abla, D., Hori, K., and Okanoya, K. (2010). Music playing and memory trace: evidence from event-related potentials. *Neurosci. Res.* 67, 334–340.

Karbusicky, V. (1986). *Grundriß der musikalischen Semantik.* Darmstadt: Wissenschaftliche Buchgesellschaft.

Katahira, K., Abla, D., Masuda, S., and Okanoya, K. (2008). Feedback-based error monitoring processes during musical performance: an ERP study. *Neurosci. Res.* 61, 120–128.

Khalfa, S., Isabelle, P., Jean-Pierre, B., and Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neurosci. Lett.* 328, 145–149.

Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child. Psychol.* 102, 299–314.

Knösche, T., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O., and Friederici, A. D. (2005). Perception of phrase structure in music. *Hum. Brain Mapp.* 24, 259–273.

Koch, M., Lingenhöhl, K., and Pilz, P. (1992). Loss of the acoustic startle response following neurotoxic lesions of the caudal pontine reticular formation: possible role of giant neurons. *Neuroscience* 49, 617–625.

Koechlin, E., and Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron* 50, 963–974.

Koelsch, S. (2000). *Brain and Music – A Contribution to the Investigation of Central Auditory Processing with a New Electrophysiological Approach.* Leipzig: Risse.

Koelsch, S. (2004). *Spatio-temporal Aspects of Processing Syntax and Semantics in Music.* Habilitation thesis, University of Leipzig.

Koelsch, S. (2009a). Music-syntactic processing and auditory memory: similarities and differences between ERAN and MMN. *Psychophysiology* 46, 179–190.

Koelsch, S. (2009b). A neuroscientific perspective on music therapy. *Ann. N. Y. Acad. Sci.* 1169, 374–384.

Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends Cogn. Sci. (Regul. Ed.)* 14, 131–137.

Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Phys. Life Rev.* (in press).

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., and Schlaug, G. (2005a). Adults and children processing music: an fMRI study. *Neuroimage* 25, 1068–1076.

Koelsch, S., Gunter, T., Wittfoth, M., and Sammler, D. (2005b). Interaction between syntax processing in language and in music: an ERP study. *J. Cogn. Neurosci.* 17, 1565–1577.

Koelsch, S., Gunter, T. C., Cramon, D. Y. von, Zysset, S., Lohmann, G., and Friederici, A. D. (2002). Bach speaks: a cortical "language-network" serves the processing of music. *Neuroimage* 17, 956–966.

Koelsch, S., and Jentschke, S. (2010). Differences in electric brain responses to melodies and chords. *J. Cogn. Neurosci.* 22, 2251–2262.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T. C., and Friederici, A. D. (2004). Music, language, and meaning: brain signatures of semantic processing. *Nat. Neurosci.* 7, 302–307.

Koelsch, S., and Mulder, J. (2002). Electric brain responses to inappropriate harmonies during listening to expressive music. *Clin. Neurophysiol.* 113, 862–869.

Koelsch, S., Offermanns, K., and Franzke, P. (2010). Music in the treatment of affective disorders: an exploratory

investigation of a new method for music-therapeutic research. *Music Percept.* 27, 307–316.

Koelsch, S., Schröger, E., and Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport* 10, 1309.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., and Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an FMRI study. *Hum. Brain Mapp.* 30, 859–873.

Koelsch, S., and Siebel, W. (2005). Towards a neural basis of music perception. *Trends Cogn. Sci. (Regul. Ed.)* 9, 578–584.

Koopman, C., and Davies, S. (2001). Musical meaning in a broader perspective. *J. Aesthet. Art Criticism* 59, 261–273.

Korzyukov, O., Winkler, I., Gumenyuk, V., and Alho, K. (2003). Processing abstract auditory features in the human auditory cortex. *Neuroimage* 20, 2245–2258.

Kreutz, G., Bongard, S., Rohrmann, S., Hodapp, V., and Grebe, D. (2004). Effects of choir singing or listening on secretory immunoglobulin A, cortisol, and emotional state. *J. Behav. Med.* 27, 623–635.

Lahav, A., Saltzman, E., and Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J. Neurosci.*, 27, 308–314.

Lamprea, M., Cardenas, F., Vianna, D., Castilho, V., Cruz-Morales, S., and Brandão, M. (2002). The distribution of fos immunoreactivity in rat brain following freezing and escape responses elicited by electrical stimulation of the inferior colliculus. *Brain Res.* 950, 186–194.

Langner, G., and Ochse, M. (2006). The neural basis of pitch and harmony in the auditory system. *Music. Sci.* 10, 185.

Lau, E., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the n400. *Nat. Rev. Neurosci.* 9, 920–933.

LeDoux, J. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23, 155–184.

Lerdahl, F. (2001). *Tonal Pitch Space.* New York, NY: Oxford University Press.

Lerdahl, F., and Jackendoff, R. (1999). *A Generative Theory of Music.* Cambridge: MIT.

Lerdahl, F., and Krumhansl, C. (2007). Modeling tonal tension. *Music Percept.* 24, 329–366.

Levitt, P., and Moore, R. (1979). Origin and organization of brainstem catecholamine innervation in the rat. *J. Comp. Neurol.* 186, 505–528.

Liberman, A., and Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36.

Liebenthal, E., Ellingson, M., Spanaki, M., Prieto, T., Ropella, K., and Binder, J. (2003). Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *Neuroimage* 19, 1395–1404.

Liegeois-Chauvel, C., Peretz, I., Babaie, M., Laguitton, V., and Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain* 121, 1853–1867.

Longoni, F., Grande, M., Hendrich, V., Kastrau, F., and Huber, W. (2005). An fMRI study on conceptual, grammatical, and morpho-phonological processing. *Brain Cogn.* 57, 131–134.

Lundqvist, L., Carlsson, F., Hilmersson, P., and Juslin, P. (2009). Emotional responses to music: experience, expression, and physiology. *Psychol. Music* 37, 61.

Maess, B., Jacobsen, T., Schröger, E., and Friederici, A. (2007). Localizing pre-attentive auditory memory-based comparison: magnetic mismatch negativity to pitch change. *Neuroimage* 37, 561–571.

Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). Musical syntax is processed in the area of Broca: an MEG-study. *Nat. Neurosci.* 4, 540–545.

Maidhof, C., and Koelsch, S. (2011). Effects of selective attention on syntax processing in music and language. *J. Cogn. Neurosci.* [Epub ahead of print].

Maidhof, C., Rieger, M., Prinz, W., and Koelsch, S. (2009). Nobody is perfect: ERP effects prior to performance errors in musicians indicate fast monitoring processes. *PLoS ONE* 4, e5032. doi: 10.1371/journal.pone.0005032

Maidhof, C., Vavatzanidis, N., Prinz, W., Rieger, M., and Koelsch, S. (2010). Processing expectancy violations during music performance and perception: an ERP study. *J. Cogn. Neurosci.* 22, 2401–2413.

Makuuchi, M., Bahlmann, J., Anwander, A., and Friederici, A. (2009). Segregating the core computational faculty of human language from working memory. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8362.

McCraty, R., Atkinson, M., Rein, G., and Watkins, A. (1996). Music enhances the effect of positive emotional states on salivary IgA. *Stress Med.* 12, 167–175.

Menning, H., Roberts, L., and Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *Neuroreport* 11, 817.

Meyer, L. (1956). *Emotion and Meaning in Music.* Chicago: University of Chicago Press.

Meyer, M., Alter, K., Friederici, A., Lohmann, G., and Cramon, D. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum. Brain Mapp.* 17, 73–88.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A., and Cramon, D. (2004). Brain activity varies with modulation of dynamic pitch variances in sentence melody. *Brain Lang.* 89, 277–289.

Miranda, R., and Ullman, M. (2007). Double dissociation between rules and memory in music: an event-related potential study. *Neuroimage* 38, 331–345.

Molholm, S., Martinez, A., Ritter, W., Javitt, D., and Foxe, J. (2005). The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. *Cereb. Cortex* 15, 545.

Moon, C., Cooper, R., and Fifer, W. (1993). Two-day-olds prefer their native language. *Infant Behav. Dev.* 16, 495–500.

Moore, B. (2008). *An Introduction to the Psychology of Hearing*, 5th Edn. Bingley: Emerald.

Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. (2001). Primitive intelli-gence'in the auditory cortex. *Trends Neurosci.* 24, 283–288.

Nan, Y., Knösche, T., and Friederici, A. (2006). The perception of musical phrase structure: a cross-cultural ERP study. *Brain Res.* 1094, 179–191.

Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Curr. Opin. Neurobiol.* 14, 474–480.

Neuhaus, C., Knösche, T., and Friederici, A. (2006). Effects of musical expertise and boundary markers on phrase perception in music. *J. Cogn. Neurosci.* 18, 472–493.

Obleser, J., Meyer, L., and Friederici, A. (2011). Dynamic assignment of neural resources in auditory comprehension of complex sentences. *Neuroimage* 56, 2310–2320.

Öngür, D., and Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* 10, 206.

Opitz, B., and Kotz, S. (2011). Ventral premotor cortex lesions disrupt learning of sequential grammatical structures. *Cortex.* (in press).

Opitz, B., Rinne, T., Mecklinger, A., Cramon, D., and Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory chenage detection: fMRI and ERP results. *Neuroimage* 15, 167–174.

Orini, M., Bailón, R., Enk, R., Koelsch, S., Mainardi, L., and Laguna, P. (2010). A method for continuously assessing the autonomic response to music-induced emotions through HRV analysis. *Med. Biol. Eng. Comput.* 48, 423–433.

Overy, K., and Molnar-Szakacs, I. (2009). Being together in time: musical experience and the mirror neuron system. *Music Percept.* 26, 489–504.

Paller, K. A., McCarthy, G., and Wood, C. C. (1992). Event-related potentials elicited by deviant endings to melodies. *Psychophysiology* 29, 202–206.

Panksepp, J., and Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behav. Processes* 60, 133–155.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., and Ross, B. (2001). Timbre-specific enhancement of auditory cortical representation in musicians. *Neuroreport* 12, 169–174.

Parsons, L. (2001). Exploring the functional neuroanatomy of music performance, perception, and comprehension. *Ann. N. Y. Acad. Sci.* 930, 211–231.

Patel, A. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681.

Patel, A. (2008). *Music, Language, and the Brain.* New York, NY: Oxford University Press.

Patel, A., and Balaban, E. (2001). Human pitch perception is reflected in the timing of stimulus-related cortical activity. *Nat. Neurosci.* 4, 839–844.

Patel, A., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. (1998). Processing syntactic relations in language and music: an event-related potential study. *J. Cogn. Neurosci.* 10, 717–733.

Patel, A., Iversen, J., Bregman, M., and Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* 19, 827–830.

Patel, A., Iversen, J., Wassenaar, M., and Hagoort, P. (2008). Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology* 22, 776–789.

Patterson, R., Uppenkamp, S., Johnsrude, I., and Griffiths, T. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36, 767–776.

Perani, D., Saccuman, M., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., Baldoli, C., and Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4758.

Peretz, I., Brattico, E., Järvenpää, M., and Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain* 132, 1277.

Peretz, I., and Coltheart, M. (2003). Modularity of music processing. *Nat. Neurosci.* 6, 688–691.

Peretz, I., and Zatorre, R. (2005). Brain organization for music processing. *Annu. Rev. Psychol.* 56, 89–114.

Petkov, C., Kayser, C., Augath, M., and Logothetis, N. (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol.* 4, e215. doi: 10.1371/journal.pbio.0040215

Pickles, J. (2008). *An Introduction to the Physiology of Hearing*, 3rd Edn. Bingley: Emerald.

Piston, W. (1948/1987). *Harmony*. New York: Norton.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148.

Poulin-Charronnat, B., Bigand, E., and Koelsch, S. (2006). Processing of musical syntax tonic versus subdominant: an event-related potential study. *J. Cogn. Neurosci.* 18, 1545–1554.

Quiroga Murcia, C., Bongard, S., and Kreutz, G. (2009). Emotional and neurohumoral responses to dancing tango argentino. *Music Med.* 1, 14.

Rameau, J.-P. (1722). *Traité de l'harmonie réduite á ses principes naturels.* Paris: J. B. C. Ballard.

Rammsayer, T., and Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Percept.* 24, 37–48.

Riemann, H. (1877/1971). *Musikalische syntaxis: Grundriss einer harmonischen satzbildungslehre*. Niederwalluf: Sändig.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405.

Rinne, T., Degerman, A., and Alho, K. (2005). Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: an fMRI study. *Neuroimage* 26, 66–72.

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.

Rohrmeier, M. (2007). "A generative grammar approach to diatonic harmonic structure," in *Proceedings of the 4th Sound and Music Computing Conference*, Lefkada, 97–100.

Rohrmeier, M. (2011). A generative grammar approach to tonal harmony. *J. Math. Music* 5.

Rüsseler, J., Altenmüller, E., Nager, W., Kohlmetz, C., and Münte, T. (2001). Event-related brain potentials to sound omissions differ in musicians and non-musicians. *Neurosci. Lett.* 308, 33–36.

Scherer, K. (2005). What are emotions? And how can they be measured? *Soc. Sci. Inform.* 44, 695.

Schmidt-Kassow, M., and Kotz, S. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *J. Cogn. Neurosci.* 21, 1693–1708.

Schöenberg, A.(1969). *Structural functions of harmony.* New York: Norton.

Schonwiesner, M., Novitski, N., Pakarinen, S., Carlson, S., Tervaniemi, M., and Näätänen, R. (2007). Heschl's gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. *J. Neurophysiol.* 97, 2075.

Schulze, K., Mueller, K., and Koelsch, S. (2011a). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *Eur. J. Neurosci.* 33, 189–196.

Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., and Koelsch, S. (2011b). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum. Brain Mapp.* 32, 771–783.

Scott, S. (2005). Auditory processing– speech, space and auditory objects. *Curr. Opin. Neurobiol.* 15, 197–201.

Shinn-Cunningham, B. (2008). Object-based auditory and visual attention. *Trends Cogn. Sci. (Regul. Ed.)* 12, 182–186.

Sinex, D., Guzik, H., Li, H., and Henderson Sabes, J. (2003). Responses of auditory nerve fibers to harmonic and mistuned complex tones. *Hear. Res.* 182, 130–139.

Slevc, L., Rosenberg, J., and Patel, A. (2009). Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychon. Bull. Rev.* 16, 374.

Sloboda, J. A. (1991). Music structure and emotional response: some empirical findings. *Psychol. Music* 19, 110–120.

Song, J., Skoe, E., Wong, P., and Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J. Cogn. Neurosci.* 20, 1892–1902.

Steinbeis, N., and Koelsch, S. (2008a). Comparing the processing of music and language meaning using EEG and FMRI provides evidence for similar

and distinct neural representations. *PLoS ONE* 3, e2226. doi: 10.1371/journal.pone.0002226

Steinbeis, N., and Koelsch, S. (2008b). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cereb. Cortex* 18, 1169.

Steinbeis, N., and Koelsch, S. (2008c). Understanding the intentions behind man-made products elicits neural activity in areas dedicated to mental state attribution. *Cereb. Cortex* 19, 619–623.

Steinbeis, N., and Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *J. Cogn. Neurosci.* 23, 604–621.

Steinhauer, K., Alter, K., and Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat. Neurosci.* 2, 191–196.

Strait, D., Kraus, N., Skoe, E., and Ashley, R. (2009). Musical experience and neural efficiency – effects of training on subcortical processing of vocal expressions of emotion. *Eur. J. Neurosci.* 29, 661–668.

Sussman, E. (2007). A new view on the MMN and attention debate: the role of context in processing auditory events. *J. Psychophysiol.* 21, 164–175.

Tervaniemi, M. (2009). Musicians – same or different? *Ann. N. Y. Acad. Sci.* 1169, 151–156.

Tervaniemi, M., Castaneda, A., Knoll, M., and Uther, M. (2006a). Sound processing in amateur musicians and nonmusicians: event-related potential and behavioral indices. *Neuroreport* 17, 1225.

Tervaniemi, M., Szameitat, A., Kruck, S., Schroger, E., Alter, K., De Baene, W., and Friederici, A. D. (2006b). From air oscillations to music and speech: functional magnetic resonance imaging evidence for fine-tuned neural networks in audition. *J. Neurosci.* 26, 8647.

Tervaniemi, M., and Huotilainen, M. (2003). The promises of change-related brain potentials in cognitive neuroscience of music. *Ann. N. Y. Acad. Sci.* 999, 29–39.

Tervaniemi, M., Ilvonen, T., Karma, K., Alho, K., and Näätänen, R. (1997). The musical brain: brain waves reveal the neurophysiological basis of musicality in human subjects. *Neurosci. Lett.* 226, 1–4.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., and Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Exp. Brain Res.* 161, 1–10.

Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., and Friederici, A. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *Eur. J. Neurosci.* 30, 1636–1642.

Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R., and Näätänen, R. (1999). Functional specialization of the human auditory cortex in processing phonetic and musical sounds: a magne-toencephalographic (MEG) study. *Neuroimage* 9, 330–336.

Tervaniemi, M., Medvedev, S., Alho, K., Pakhomov, S., Roudas, M., Zuijen, T., and Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Hum. Brain Mapp.* 10, 74–79.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R., and Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learn. Mem.* 8, 295.

Tillmann, B., Bharucha, J., and Bigand, E. (2000). Implicit learning of tonality: a self-organized approach. *Psychol. Rev.* 107, 885–913.

Tillmann, B., Janata, P., and Bharucha, J. (2003). Activation of the inferior frontal cortex in musical priming. *Brain Res. Cogn. Brain Res.* 16, 145–161.

Tramo, M., Shah, G., and Braida, L. (2002). Functional role of auditory cortex in frequency processing and pitch perception. *J. Neurophysiol.* 87, 122.

Trehub, S. (2003). The developmental origins of musicality. *Nat. Neurosci.* 6, 669–673.

Van Petten, C., and Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Mem. Cognit.* 18, 380–393.

Verleger, R. (1990). P3-evoking wrong notes: unexpected, awaited, or arousing? *Int. J. Neurosci.* 55, 171–179.

Võ, M., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M., and Jacobs, A. (2009). The Berlin affective word list reloaded (BAWL-R). *Behav. Res. Methods* 41, 534–538.

Wallin, N. L., Merker, B., and Brown, S. (eds). (2000). *The Origins of Music.* Cambridge, MA: MIT Press.

Warren, J. D., Uppenkamp, S., Patterson, R. D., and Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10038–10042.

Watanabe, T., Yagishita, S., and Kikyo, H. (2008). Memory of music: roles

of right hippocampus and left inferior frontal gyrus. *Neuroimage* 39, 483–491.

Whitfield, I. (1980). Auditory cortex and the pitch of complex tones. *J. Acoust. Soc. Am.* 67, 644.

Winkler, I. (2007). Interpreting the mismatch negativity. *J. Psychophysiol.* 21, 147–163.

Winkler, I., Denham, S., and Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci. (Regul. Ed.)* 13, 532–540.

Wittfoth, M., Schröder, C., Schardt, D., Dengler, R., Heinze, H., and Kotz, S. (2010). On emotional conflict: interference resolution of happy and angry prosody reveals valence-specific effects. *Cereb. Cortex* 20, 383.

Wong, P., Skoe, E., Russo, N., Dees, T., and Kraus, N. (2007). Musical experience shapes human brain-stem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.

Zatorre, R. (1988). Pitch perception of complex tones and human temporal-lobe function. *J. Acoust. Soc. Am.* 84, 566–572.

Zatorre, R. (2001). Neural specializations for tonal processing. *Ann. N. Y. Acad. Sci.* 930, 193–210.

Zatorre, R., Belin, P., and Penhune, V. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci. (Regul. Ed.)* 6, 37–46.

Zatorre, R., Evans, A., and Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* 14, 1908.

Zuijen, T., Sussman, E., Winkler, I., Näätänen, R., and Tervaniemi, M. (2004). Grouping of sequential sounds-an event-related potential study comparing musicians and nonmusicians. *J. Cogn. Neurosci.* 16, 331–338.

Zuijen, T., Sussman, E., Winkler, I., Näätänen, R., and Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity–a mismatch negativity study comparing musicians and non-musicians. *Brain Res. Cogn. Brain Res.* 23, 270–276.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.