# One Step Beyond: Musical Expertise and Word Learning 10.1 A Cortical Framework of Speech and Language Processing 10.1.1 The Faculty of Language and its Neural Substrate

Stefan Elmer, Eva Dittinger, Mireille R Besson

## HAL Id: hal-02437826
## https://amu.hal.science/hal-02437826

# Chapter 10

# One Step Beyond: Musical Expertise and Word Learning

Stefan Elmer, Eva Dittinger,
and Mireille Besson

## 10.1 A Cortical Framework of Speech and Language Processing

### 10.1.1 The Faculty of Language and its Neural Substrate

Speech and language processing constitute a uniquely human faculty that can be distinguished from other forms of communication in the animal kingdom. In fact, even though it has been proposed that different species possess the faculty of language in a broad sense (i.e. mimic, gesture, olfactory cues, etc.), only human language is characterized by a recur-sive structure (Fitch, 2010), the latter referring to the ability to produce an infinite number of phrases from a finite number of entities (i.e. phonemes and words) . Until now, different evolutionary theories have attempted to explain the possible origin of speech and language processing in human beings, ranging from the expansion of brain size relative to body weight, genetics, brain asymmetries, anatomical characteristics of the larynx, and mirror neurons, to cultural and societal aspects (among others) (Fitch, 2010; Fitch & Reby, 2001; Hauser, 2002). From a linguistic perspective, speech and language processing can be sub-divided into different subsystems including phonology, syntax, semantics, morphology, and pragmatics. Phonology refers to knowledge of the sound structure, syntax deals with the rules governing the combination and the order of words in a sentence, semantics ad-dresses the meaning of single words and sentences, morphology is concerned with the structure of words, whereas pragmatics examines language in contexts (e.g. discourse, in-ference, interaction). Furthermore, depending on intonations and stresses, semantics can be influenced.

In the last decades, both lesion studies and neuroimaging techniques have fundamentally contributed to a better understanding of the cortical organization of speech and language

processing. For more than a century, the classical Wernicke-Lichtheim-Geschwind model (Boland, 2014) was considered an accurate representation. This simplistic model was purely based on lesion studies with patients suffering from receptive (i.e. Wernicke's), expressive (i.e. Broca's), or conduction aphasia, and postulated that Broca's area is crucial for language production, whereas Wernicke's area subserves language comprehension functions. In addition, the arcuate fasciculus, a major fibre bundle connecting posterior superior temporal regions (i.e. Brodmann's area 42, Wernicke's area) with Broca's region (i.e. Brodmann's area 44 and 45, pars opercularis and triangularis), was recognized to mediate information exchange between these areas. Even though this historical neurological model enabled the description of a variety of aphasic symptoms, recognizing the contribution of left-sided perisylvian areas to perception and articulation, it is reductive, oversimplified, and relies on brain lesions in-stead of healthy functioning.

Since that time, several branches of research have fundamentally contributed to im-prove and ameliorate the cartography of speech and language processing in the human brain. In analogy to the visual system (Milner & Goodale, 2008; Miskin & Ungerleider, 1982), current models conjointly postulate the involvement of two parallel, bidirectional, and hierarchically-organized processing streams stretching from the auditory-related cortex toward the temporal pole (ventral stream) and the frontal lobe (dorsal stream) (Rauschecker & Scott, 2009), and meshing at two points of convergence, namely in the pos-terior supratemporal plane and in the ventral part of the frontal cortex (see Figure 10.1). However, these models diverge somewhat in the description of the processes supported by the two processing streams, as well as in their linguistic and neurological conceptualiza-tion. In turn, we will summarize some of the most popular frameworks of cortical speech and language processing proposed by Hickok and Poeppel (2007), Bornkessel-Schlesewski and Schlesewski (2013), Friederici (2009, 2011, 2012), and Hagoort (2014). It is important to
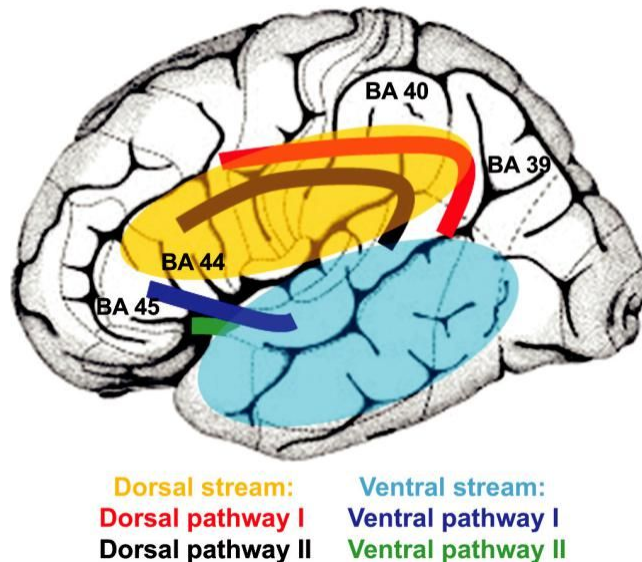


**Figure 10.1** Ventral and dorsal streams of language processing.

**Table 10.1 Summary of current models of cortical speech and language processing proposed by Hickok and Poeppel (2007), Bornkessel-Schlesewski and Schlesewsky (2013), Friederici (2009, 2011, 2012), and Hagoort (2014)**

| Authors | Type of model | Short description |
| --- | --- | --- |
| Hickok and Poeppel | Dual-stream model | The ventral stream is responsible for sound-to-meaning mapping, whereas the dorsal one supports sensory-motor mapping mechanisms and articulation. The ventral stream is bilaterally organized, whereas the dorsal one is lateralized to the left hemisphere. |
| Bornkessel-Schlesewsky and Schlesewsky | Dual-stream model | The ventral stream operates in a time-independent manner by activating and unifying conceptual schemata necessary for creating units of increased complexity and enabling semantic integration. Otherwise, the dorsal stream subserves general time-dependent processes and is engaged in segmenting the input into prosodic words, combining these elements into a syntactic structure, as well as in assessing them into actioncentred representations. |
| Friederici | Dual-stream model | This model postulates a dorsal and a ventral processing stream that is compatible with the underlying white matter architecture. Within this framework, the dorsal and ventral streams are responsible for more than one isolated function, and based on the underlying white matter architecture, each of them can be segregated into two subpathways (i.e. pathway I & II). |
| Hagoort | Dynamic and cognitive model | This model acts on the assumption that for central aspects of language processing the neural infrastructure is shared between comprehension and production systems in the form of dynamic networks, and that this neural substrate is not language-specific. Frontal, temporal, and parietal brain regions are differentially recruited based on task-related network characteristics, meaning that the functional role of a specific brain area is influenced by the other regions of the network depending on information type, processing demands, and cognitive control. |

mention that, besides these models (outlined in Table 10.1), there are a multiplicity of other frameworks that are not discussed in this chapter.

## 10.1.2 Current Models of Speech and Language Processing

The dual-stream model proposed by Hickok and Poeppel (2007) relies on evidence from both lesion-and task-related neuroimaging studies, and postulates a ventral stream that

processes speech signals for comprehension and a dorsal one that maps acoustic speech signals to frontal lobe articulatory networks. When speech signals reach the auditory- related cortex, the model already postulates a division of labour between the two hemi-spheres (Giraud et al., 2007; Zatorre & Belin, 2001), and highlights a relative specialization of the left hemisphere for the processing of transient and fast -changing acoustic cues (i.e. segmental, time windows of about 25 ms, 40 Hz). By contrast, the right-sided counter-part is more sensitive to slow acoustic modulations and frequency information (i.e. supra-segmental, time windows of about 250 ms, 4 Hz). Even though this relative processing asymmetry is controversial (Santoro et al., 2014; Overath et al., 2015), it has previously been associated with a differential spacing between microcolumns, myelination (Harasty et al., 2003; Seldon, 1981), as well as with asymmetric spontaneous neural oscillations in the theta (right -sided asymmetry) and gamma (left-sided asymmetry) frequency range (Giraud et al., 2007). These specific oscillations have been proposed to play an important role in 'packing' the multitime speech signal (i.e. phonemes, words, etc.) into units of the appro-priate temporal granularity (Giraud & Poeppel, 2012) necessary for further processing steps along the ventral and dorsal streams.

In the Hickok and Poeppel (2007) model, the speech signal, after acoustic analysis, spreads along a bilaterally distributed ventral route engaged in meaning extraction (i.e. lexical-semantic processing) through a cascade of hierarchical processes. The posterior supratemporal plane and the posterior part of the superior temporal sulcus (STS) act as an interface between spectrotemporal and phonological processes. At the next hierarch-ical level, phonological information is mapped onto lexical representations in memory by recruiting posterior, middle, and inferior portions of the temporal lobe. Finally, informa-tion converges on a combinatorial network situated in the proximity of the temporal pole and supporting lexical -semantic integration, sentence-level processing, as well as syntactic and semantic nesting. In contrast to the bilateral organization of the ventral route (with a slight left-hemispheric bias), the dorsal stream is strongly left-dominant, maps sensory and phonological representations onto articulatory motor representations in the frontal cortex, and constitutes a bridge between the speech perception and production systems. The dorsal pathway originates from the posterior supratemporal plane, runs through a sensory-motor interface at the parieto-temporal junction, and projects to Broca's area as well as to the dorsal part of the premotor cortex. The model proposed by Hickok and Poeppel is ana-tomically well-defined. However, it only focuses on speech and not generally on language processing, and lacks a certain linguistic depth, especially regarding syntactic and lexical- semantic processes.

Recently, Bornkessel-Schlesewsky and Schlesewsky (2013) presented an alternative dual-stream model that attempts to unify neurobiological assumptions and linguistic sentence comprehension. Similar to Hickok and Poeppel, the authors postulate the engagement of bidirectional ventral and dorsal processing streams, however with important computational differences regarding time dependence. The ventral stream projects from auditory core areas along the superior temporal plane toward the anterior temporal cortex and ventral frontal cortex, whereas the dorsal stream runs, via a relay station situated in the inferior parietal lobe, to the inferior frontal cortex. In their framework, the authors abstain from a conceptual dichotomy between comprehension and articulation and instead propose time-dependent processes along the dorsal stream and time-independent ones along the ventral stream. Thereby, it is assumed that the ventral stream operates in a time-independent manner by

activating and unifying conceptual schemata necessary for creating units of increased com-plexity and enabling semantic integration. Furthermore, the model postulates that the ventral stream enables word-level semantic information, as well as phrase-structure comprehension, by the activation and unification of actor-event schemata (e.g. who and what) that are actor-centred (i.e. focus on persons or objects responsible for a certain events) and category neutral (i.e. nouns or verbs). Unification occurs by integrating one schema (e.g. who) into another one (e.g. what). According to the same model, the dorsal stream does not specifically subserve articulation and repetition but, rather, general time-dependent pro-cesses. Therefore, this stream supports both speech production and comprehension and is engaged in segmenting the input into prosodic words, combining these elements into a syn-tactic structure, as well as in assessing them in actioncentred representations (i.e. who is re-sponsible for a certain event).

Finally, it is important to mention that even though the two processing streams converge in the ventral part of the frontal cortex, this brain region is not assumed to support specific linguistic functions but, rather, accommodates action planning and general executive functions such as verbal and non-verbal memory, inhibitory control, switching, and updating. This framework fundamentally contributes to a better understanding of the cortical implementation of linguistic processes (especially phonology, semantics, and syntax). Otherwise, the model is anatomically vaguely defined and does not explicitly address processing asymmetries along the ventral and dorsal streams.

An alternative dual-stream model is the one proposed by Friederici (2009, 2011, 2012). This model is based on sentence processing, is anatomically as well as linguistically well-defined, and reconciles precise cortical cartography with the underlying white matter pathways. Within this framework, the dorsal and ventral streams are responsible for more than one isolated function, and based on the underlying white matter architecture, each of them can be segregated into two subpathways. The dorsal stream constitutes one fibre bundle connecting the superior temporal cortex to the premotor cortex (via the superior longitudinal fasciculus, dorsal pathway I) as well as by a second pathway linking the temporal cortex to pars opercularis (via the arcuate fasciculus, dorsal pathway II). By contrast, the ventral stream relies on the fibre bundle running from the anterior temporal cortex to pars triangularis (via the extreme capsule, ventral pathway I) as well as on the connection between the anterior supratemporal regions and the frontal operculum (via the uncinate fasciculus, ventral pathway II) (see Figure 10.1).

Initial acoustic and phonological analyses involve the primary and secondary audi-tory cortex, from where activity spreads along the ventral stream to anterior and posterior supratemporal regions. In successive processing steps, initial phrase structure building is conjointly analysed by the left anterior temporal cortex and the frontal operculum (i.e. ventral pathway II). Successively, semantic, grammatical, and thematic relations are processed in a parallel manner. Semantic analyses are supported by middle-posterior areas of the superior and middle temporal gyrus, as well as by pars triangularis and orbitalis (i.e. ventral pathway I). Otherwise, syntactically complex sentences are dependent upon pars opercularis and the posterior temporal cortex (i.e. dorsal pathway II). Furthermore, within this framework, the anterior part of the temporal lobe supports both semantic and syn-tactic processing and subserves combinatorial processes. Finally, the connection between the posterior supratemporal plane and the premotor cortex (i.e. dorsal pathway I) promotes auditory-to-motor mapping mechanisms, whereas the fibre bundle bridging the posterior

temporal cortex and pars opercularis (i.e. dorsal pathway I) supports syntactic processes (especially when sentences are complex). Prosodic information is assumed to be predominantly processed in the right hemisphere and integrated with left-hemispheric syntactic in-formation through the posterior part of the corpus callosum (i.e. isthmus).

The model proposed by Hagoort (2014) overlaps with the models described above in that it posits the engagement of temporal, parietal, and frontal brain regions as the constitutional entities of speech and language processing. However, this model acts on the assumption that for central aspects of language processing, the neural infrastructure is shared between com-prehension and production systems in the form of dynamic networks, and that this neural substrate is not language-specific (see also Friederici & Singer, 2015). Within this frame-work, brain regions situated along the ventral and dorsal streams are differentially recruited based on task-related network characteristics, meaning that the functional role of a specific brain area is influenced by the other regions of the network depending on information type (i.e. phonological, syntactic, and semantic), processing demands, and cognitive control mechanisms.

This perspective has been implemented in a 'Memory, Unification, and Control' (MUC) model that postulates that regions situated in distributed networks in the temporal and in-ferior parietal cortex generally subserve mnemonic representations (i.e. phonological word forms, morphological information, and the syntactic templates associated with nouns, verbs, and adjectives). Otherwise, frontal regions (including Broca's area) are crucially in-volved in unification operations by generating larger structures (i.e. phonologic, semantic, and syntactic) from the templates retrieved from memory. Finally, the model also posits that 'memory' and 'unification' mechanisms are hierarchically subordinated to higher execu-tive control mechanisms that are executed by the dorsolateral prefrontal cortex, the anterior cingulate cortex, as well as by the parietal attention system. This model is compatible with

studies on non-human primates militating that diverse perceptual and cognitive functions are based on similar neural mechanisms, leading to suggestions of a general rather than a language-specific intrinsic organization of the human brain (Rauschecker & Scott, 2009; Friederici & Singer, 2015).

## 10.2   Perceptual and Cognitive Demands onSpeech- and Language-Learning Mechanisms

The investigation of speech- and language-learning mechanisms in infants provides an empirically-based framework for better comprehending the perceptual and cognitive processing underlying this uniquely human faculty. Currently, it is known that the auditory cortex starts functioning at about 24 weeks of gestation and that shortly after birth, infants are characterized by some left-hemispheric language specialization (Dehaene-Lambertz & Spelke, 2015; Perani et al., 2011). These functional-anatomical constraints, in association with neural commitment (Kuhl, 2004), facilitate language-learning mechanisms in infants, including the perception and discrimination of vowels, consonants, phonetic contrasts, and stress patterns in words. Nowadays, it is also recognized that a part of the dorsal stream

(i.e. dorsal pathway I) linking the auditory cortex with the premotor cortex is observable shortly after birth, whereas a second pathway connecting the auditory cortex with Broca's area (i.e. dorsal pathway II) matures only later during development (Brauer et al., 2013).

Speech constitutes a concatenated acoustic signal whose parameters (e.g. pitch and envelope) not only vary in time but also depend on the talker (e.g. gender), speech rate, as well as on the context (e.g. loudness of the environment). Consequently, before acquiring single words, infants have to learn to decrypt the 'speech code' by figuring out the composition of the phonetic categories of a specific language. This mechanism, called 'categorical perception', is (at least partially) mediated by the 'magnet effect', a phenomenon where prototyp-ical phonetic representations stored in memory attract surrounding deviant sounds (Kuhl, 2004). A commonplace example of this effect can be observed in English infants who learn that /r/and / l/pertain to different phonetic categories, whereas Japanese children treat these two phonemes as equivalent because they are not lexically contrastive in Japanese.

Aside from categorical perception, a further important phenomenon is speech segmenta-tion (i.e. the ability to extract meaningful sounds from continuous speech). The recognition of words' boundaries is at least partly based on 'statistical learning'—an implicit faculty that enables infants to analyse statistical distributions and relationships between speech sounds (Kuhl, 2004; Saffran, 2003). For example, within the German language it is more probable (i.e. there is a higher transitional probability) that the consonant /r/follows the consonants /t/and /p/ than /z/. Importantly, infants also strongly rely on prosodic cues (e.g. linguistic stress on the first syllable in German) for the segmentation of a continuous speech signal into different subunits and to identify potential word candidates.

Certainly, speech- and language-learning mechanisms are also strongly influenced by social factors that enable infants to be attracted to infant-directed speech (e.g. motherese) by providing enriched referential information through action-based forms (Kuhl, 2007). In this context, there is evidence showing that language-learning mechanisms rely on the functional contribution of phylogenetically older subcortical reward systems (Péron et al., 2016) possibly involved in reinforcing human motivation to learn a language (Ripollés et al., 2014).

One of the most distinctive differences between infants' and adults speech and language acquisition is that, in the latter, several of the processes described above are established and the brain has already committed to the mother tongue. In fact, in adults, the neural cir-cuits underlying speech processing are fully developed and prefrontal brain regions sup-porting higher cognitive functions (i.e. attention, memory, planning, inhibition, etc.) and explicit learning strategies have reached a maturational ceiling (Gogtay et al., 2004). On the other hand, a common experience, such as being exposed to a new language in a foreign country, brings to light several analogies between speech and language acquisition in in-fants and adults. Similar to infants, adults have to learn to recognize the phonetic reper-toire of a foreign language as well as to segregate continuous speech into subunits in order to recognize words' boundaries and to identify single words. Depending on the phonetic overlap between native and foreign language (e.g. Indo-European, Asian, or neo-Latin), the acquisition process can be either facilitated or hindered. A further point is that not only the phonetic properties of a language but also its spectrotemporal attributes, as well as its syn-tactic complexity, have an influence on the learning process. This is, for example, the case for tonal (e.g. Mandarin or Cantonese Chinese) and quantitative languages (e.g. Finnish) where phonemes vary in pitch (i.e. rising or falling), temporal extension of the vowel (i.e. short or

long), or even by a combination of such spectral and temporal attributes that contribute to differences in word meaning (e.g. Thai).

Currently, there is a significant amount of literature addressing differential aspects of speech-and language-learning mechanisms in both adults and children, ranging from the articulation of foreign speech sounds, categorical perception, speech segmentation, and word learning, to the implicit or explicit acquisition of syntactic knowledge. Even though all these studies cannot be discussed in detail here, it is important to emphasize that the neural circuits underlying different aspects of language-learning mechanisms are the same as those described in Section 10.1, 'A Cortical Framework of Speech and Language Processing'. Therefore, we will only provide a few examples of some of these studies.

Recently, López-Barroso and colleagues (2013) measured a sample of adult participants who performed an artificial language-learning task consisting of segmenting and learning single pseudo-words presented in the form of concatenated speech. Results demonstrated that word-learning ability was related to increased functional and structural connectivity between the left auditory cortex and Broca's region (dorsal pathway). In other studies, Golestani and colleagues reported that French participants who more accurately learnt to discriminate (Golestani and Zatorre, 2004) or pronounce (Golestani and Pallier, 2007) non- native phonetic contrasts were characterized by increased brain activity in left perisylvian areas as well as by enhanced grey-matter density in brain regions supporting speech articula-tion, respectively. Finally, previous electrophysiological studies on vocabulary learning were able to demonstrate lexical-semantic facilitation effects (i.e. as reflected by increased N400 amplitudes) after only a few hours of training (McLaughlin et al., 2004; Perfetti et al., 2005).

In the next section, we will introduce professional musicians, as well as children under-going short- or long-term music training, as a vehicle for better understanding the mu-tual interdependence between perception and cognition during different aspects of speech and language learning. Thereby, we will draw a bridge between functional and structural training-related brain changes, perceptual and cognitive benefits, and several aspects of lan-guage learning.

## 10.3   Music to Speech Transfer Effects

Compared to language research, the neuroscience of music is a relatively new field that has also led to fascinating discoveries. This is at least partly based on the fact that, while all normally-developing children end up being language experts, not all human beings are pro-fessional musicians. In fact, much has been learned about the anatomo-functional organiza-tion of the brain and about brain plasticity by studying the musician's brain and the impact of long-term music training on different perceptual and cognitive functions (Elbert et al., 1995; Jäncke, 2012; Münte et al., 2002; Schlaug et al., 1995; Sluming et al., 2002; Schneider et al., 2002).

Here, we focus on transfer effects, defined as the influence of training in one domain on the level of performance in another domain. Specifically, we address transfer effects from music training to several aspects of language processing (in a broad sense that includes speech processing), to cross-modal integration, and to executive functions (see Asaridou & McQueen, 2013, for the influence of linguistic experience on music processing). The results

described in this section were obtained using different methodologies: behavioural measurements, electrophysiological recordings at the level of the brainstem and at the cortical level (event-relatedpotentials—ERPs), and functional magnetic resonance imaging (fMRI). First, we review the growing evidence for transfer effects; then, we consider two main inter-pretations of such transfers; and finally, we address the questions of the influence of music training or of genetic predispositions for music, and whether music and language processing rely on shared or distinct neural substrates.

## 10.3.1   Growing Evidence for Transfer Effects

Results of many experiments have demonstrated the positive influence of music training on speech perception (Besson et al., 2011; Kraus & Chandrasekaran, 2010). For instance, there is clear evidence that music training influences the segmental processing of speech sounds (consonants, vowels, and syllables) (see Appendix 10.1 and Audio 10.1 for examples of consonant-vowel syllables) at multiple levels of the auditory system from the brainstem (Bidelman & Krishnan, 2010; Musacchia et al., 2007; Wong et al., 2007) to cortical regions (Bidelman et al., 2011; Chobert et al., 2014; Bidelman & Alain, 2015; Elmer et al., 2012; Meyer et al., 2012; Ott, 2011). Music training also positively influences pitch processing in tonal lan-guages such as Mandarin Chinese and Thai, in which pitch variations in vowels change the meaning of words (Wong et al., 2007; Bidleman et al., 2011; Alexander et al., 2005; Bidelman et al., 2013; Lee & Hung, 2008). At the suprasegmental level (couple of syllables, words, and sentences), results have shown that musicians are typically more sensitive than non- musicians to linguistic and emotional prosody (i.e. speech melody and rhythm) (Cason & Schön, 2012; Lima & Castro, 2011; Ma & Thompson, 2015; Magne et al., 2006; Marques et al., 2007; Moreno et al., 2009; Schön et al., 2004; Thompson et al., 2004, 2012; and for contrastive results, Trimmer & Cuddy, 2008), as well as to the timbre of human voices (Chartrand & Belin, 2006).

As reviewed in Section 10.2, categorical perception and speech segmentation are the cornerstones of speech perception. In this context, Bidelman and collaborators (2013) dem-onstrated an influence of music training on the categorical perception of speech sounds (/u/to /a/continuum) at the cortical level: the P2 component of the ERPs was sensitive to between-categories' phonetic boundaries defined by psychometric functions. By contrast, this effect was not significant at the brainstem level. Results of subsequent experiments also showed that younger (Bidelman et al., 2014) and older musicians (Bidelman & Alain, 2015) were faster and showed steeper boundaries between phonetic categories in a vowel cat-egorization task than non-musicians. Increased auditory sensitivity may thus be one of the driving forces behind enhanced categorical perception and enhanced speech processing in musicians.

Speech segmentation is also fundamental to speech comprehension. This is clearly ex-emplified when learning a foreign language that is perceived as a continuous stream of non-sense words. François and colleagues (2013) used a longitudinal approach in children, over a period of two school years, during which the children were trained in music or in painting (45 minutes, twice a week in the first year and once a week in the second year). Children first listened to 5 minutes of an artificial, continuous, sung language in which syllables varied in their transitional probability (as previously described) and was higher within three syllabic

items (hence considered as familiar) than between two consecutive items (hence considered as unfamiliar). Children were then asked which of two items was most familiar. At the be-havioural level, implicit recognition of familiar and unfamiliar items steadily increased over the course of the two years of music training—but not of painting training. At the cortical level, and similarly to adults (François & Schön, 2011), only the music-trained children were characterized by a fronto-central negative component that was larger to unfamiliar than to familiar items. Thus, this longitudinal study demonstrated that music training improved speech segmentation.

Certainly, transfer from music to language is by no means limited to low-level speech pro-cessing, such as categorical perception or speech segmentation, but has also been shown to extend to higher-level language processing. For example, based on the idea that both music and language are structured sequences of events that unfold in time, several studies have investigated the influence of music training on syntactic processing in adults (Fitzroy & Sanders, 2012) and in children (Janus et al., 2016; Jentschke & Koelsch, 2009).. Jentschke and Koelsch (2009) compared the ERPs to violations of linguistic and musical syntax in music-ally trained and untrained 10–11-year-old children. The electrophysiological markers of both types of violations were larger in the former group. Not surprisingly, musically trained chil-dren were more sensitive to harmonic structure than children without such training. What was more surprising is that they also showed more comprehensive knowledge of the syn-tactic structure of sentences, possibly through faster implicit syntactic processing and/or a more efficient use of the prosodic and rhythmic cues that constrain syntactic constructions (Roncaglia-Denissen et al., 2013; Schmidt-Kassow & Kotz, 2009).

A hotly debated issue in the literature is whether the influence of music training on dif-ferent abilities is causally linked to music training or rather results from genetic predis-positions for music. Cross-sectional studies comparing (professional) musicians and non-musicians, children or adults, do not allow for this issue to be addressed since, as pointed out by Schellenberg (2004), correlation is not causality. To our knowledge, the only way to test causality in humans is to use a test–training–retest longitudinal approach that compares two groups of non-musicians (children or adults)—one group trained with music and the other group trained with an equally interesting activity, such as painting or cooking. Participants are pseudo-randomly assigned to one of the two groups, thereby en-suring that no between-group differences on the different tests of interest are found before training. If musically trained participants outperformed painting-/cooking-trained parti-cipants in the retest session, this is evidence that the type of training strongly influenced the results. This approach has been successfully used to demonstrate the influence of music training on the perception of pitch variations in sentence context (Moreno et al., 2009) and on the preattentive processing of the temporal aspects of speech (Chobert et al., 2014) and on speech segmentation (Francois et al., 2013), as previously mentioned.

## 10.3.2  Interpretations of Transfer Effects

Having summarized some of the evidence for music to language transfer effects, we now turn to the most important question: *How can we explain music to speech processing transfer effects?* Two main interpretations, that we refer to as the cascade and multidimensional hy-potheses, have been proposed in the literature.

Following the cascade hypothesis, transfer effects arise because speech and music are auditory signals relying on the same acoustic parameters (i.e. duration, frequency, intensity, and timbre). As musicians are highly trained in perceiving the acoustic structure of sounds, sound encoding is facilitated not only in music but also in speech. Thus, enhanced perceptual encoding and categorization of speech sounds in musicians facilitates higher levels of speech processing. Let us take the example of novel word learning, that will be considered in detail later. If a learner is able to differentiate the subtle acoustic features of different phonemes, he/she may form a more precise phonological representation of the new word. Consequently, it will be easier to associate such a phonological representation with the corresponding word meaning than a less distinct one. This bottom-up interpretation may explain why musicians are more sensitive to the spectro-temporal aspects of speech processing at the segmental and suprasegmental levels, as already reviewed, as well as to other aspects of speech perception (e.g. speech in noise perception, speech segmentation, sentence syntactic structure).

The multidimensional hypothesis is based on the fact that music training is multidimensional. Playing a musical instrument involves auditory and visual perception (the notes on the score), visuo-auditory-motor integration (transforming visual notes into sounds through movements), selective and divided attention (focusing attention on one's own instrument and dividing attention between the different instruments of the orchestra), and motor control (adapting posture and fine distal movements). Playing a musical piece also requires memory (most musicians play by heart), executive functions (switching between visual and auditory codes), inhibitory control (withholding a movement to play at the right moment and up-dating information), and emotion (as translated into the interpretation of the musical piece). Since professional musicians are at their advantage in these different functions, they may outperform non-musicians when these functions are necessary for the task at hand.

Evidence for this multidimensional hypothesis is accumulating from several results showing transfer from music training to cross-modal integration (Lee & Noppeney, 2011; Pantev et al., 2009; Chen et al., 2008) and to executive functions. Executive functions are defined as top-down processes that control behaviour, and typically include selective attention, working memory (WM), short-and long-term memory, and cognitive control (in-hibitory control, cognitive flexibility, updating), although this is still a matter of controversy (Diamond, 2013). In this respect, musicians have been shown to be more efficient at audio- motor learning than non-musicians (Barrett et al., 2013; Lahav et al., 2007; Mathias et al., 2015), possibly because they use different integration strategies, with musicians relying more on auditory and non-musicians more on visual information (Paraskevopoulos et al., 2012, 2014, 2015). In addition, there is evidence that adult musicians outperform non-musicians in WM tasks based on musical stimuli (George & Coch, 2011; Schulze & Koelsch, 2012; Schulze, Mueller, et al., 2011; Schulze, Zysset, et al., 2011; Pallesen et al., 2010; Schulze et al., 2012; Williamson et al., 2010), even though the influence of music training on verbal memory is more controversial (Williamson et al., 2010; Brandler & Rammsayer, 2003; Chan et al., 1998). Importantly, WM, short-term memory, and long-term memory are tightly intertwined, and more work is clearly needed to disentangle the different components of executive functions (Franklin et al., 2008; Jakobson et al., 2008). Research in children also demonstrated that music training can have an influence on executive functions (cognitive flexibility, processing speed, inhibitory control, non-verbal intelligence) as well as on short-term and long-term memory (Janus et al., 2016; Moreno et al., 2011; Bergman Nutley et al., 2014; Roden et al., 2012; Zuk et al., 2014).

In summary, results of a number of studies suggest that music training in adults and in children positively influences several aspects of language processing, executive functions, WM, as well as short-term and long-term memory. Importantly, differences between adult musicians or children with music training and controls are generally larger for the most demanding tasks, when most resources are needed to perform the task at hand (Besson et al., 2011; Diamond, 2013). However, in line with a dynamic and interactive view of human cognition, results also showed that these different functions are intrinsically linked. For instance, Diamond (2013) reported that the training of task-switching abilities transferred to verbal and non-verbal WM, inhibition (Stroop interference), and reasoning tasks. Better understanding of these intricate relationships is an exciting aim of future research, keeping in mind that we need to use well-controlled experimental designs, standardized tests when they are available (e.g. forward digit span for short-term memory, backward digit span for WM), and data-analysis methods that allow controlling for the effects of the many different factors that can also influence the results.

### 10.3.3  Transfer Effects and Shared Neural Networks

Finally, one current hotly debated topic is whether the music to language and cognition transfer effects already reviewed are supported by *shared neural networks.* In fact, the ques-tion of whether brain networks involved in music and language processing are similar or different is a long-standing question that recently received new answers.

Let us take the example of syntactic processing and Broca's area. Using fMRI, early re-sults demonstrated that Broca's area, considered as specifically involved in the processing of linguistic syntactic structures for over a century (Berwick et al., 2013; Friederici et al., 2006), was also activated when processing musical structures (Koelsch et al., 2002; Levitin & Menon, 2003; Maess et al., 2001; Tillmann et al., 2003; Vuust et al., 2006). These re-sults provided evidence that processing syntax in music and language relied on shared neural substrates. This conclusion was further supported by recent results of Abrams et al. (2011) and Rogalsky et al. (2011) showing that similar activations of frontal and temporal regions in both hemispheres are activated by temporal violations in linguistic sentences and melodies. However, results based on a more fine-grained approach— multivariate pattern analysis—showed that the two types of stimuli elicited spatially distinct activity. Thus, based on these results, the authors concluded that temporal structure is encoded differently within the two domains and that distinct cortical networks are activated. Interestingly, brain structures in these networks involved the voice-selective areas identified by Belin and collaborators (2000) and the speech-selective component that emerged from the hypothesis-free voxel decomposition method recently used by Norman-Haignere and colleagues (2015).

In summary, through the example of Broca's and temporal areas, the most studied brain structures in the neuroscience of language, we have seen evidence both for shared and for distinct networks involved in processing syntax and temporal structures in music and language. Importantly, evidence is tightly linked to the specific aspects of music and language that are compared and to the methods chosen for analysis. Depending upon the temporo-spatial resolution of the method, upon the characteristics of the stimuli, and upon the task at hand, results may show overlap of brain regions involved in music and language processing
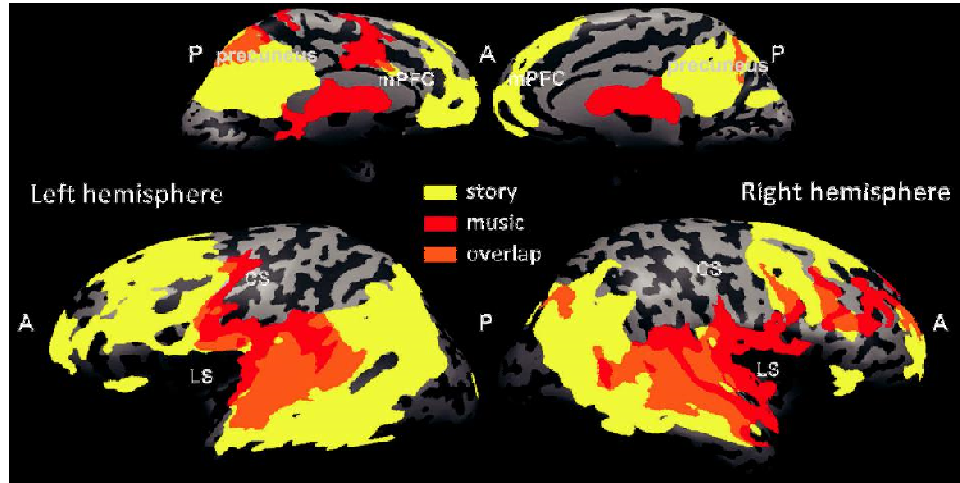
**Figure 10.2** Maps of brain activation whilst attentively listening to an intact story (yellow) or an intact musical excerpt (red) are shown for experienced pianists. Story listening evoked reliable responses in the temporo-parietal junction, angular gyrus, inferior frontal gyrus, lateral and medial prefrontal areas, and orbitofrontal cortex. Reliable responses to music were found in the lateral sulcus, pre-central gyrus, and middle frontal gyrus. Overlapping regions of reliable responses to both stimuli (orange) were evident in early auditory areas along the superior temporal gyrus (STG). mPFC = middle pre-frontal cortex, A = anterior, P = posterior, CS = central sulcus, LS = lateral sulcus.

or distinct local networks involved in specific aspects of language and music processing (see Figure 10.2). There is no doubt that tremendous progress in our understanding of the language–music relationship will be made in the years to come by using finer-grained ana-lyses of the spatio-temporal dynamics of brain networks in well-controlled experiments. Moreover, language and music are complex human functions that are not processed inde-pendently of other cognitive and emotional functions. As a consequence, we consider the cascade and the multidimensional interpretations of transfer effects as strongly complemen-tary. Finally, as we will see in the following section, the ERPs method, that allows us to con-tinuously record on-line changes in brain activity associated to the stimuli and task at hand, has also provided interesting results regarding the spatio-temporal dynamics of music to language transfer effects.

# 10.4 Music Training and Word Learning

In previous parts of this chapter, we considered word learning as an example of a multi-dimensional task relying on both perceptual and cognitive functions. Therefore, in this section, we specifically examine meaning acquisition of novel words, and we illustrate why

word learning constitutes a wonderful opportunity to study the influence of music training on semantic processing, one of the key features of language.

When it comes to learning the meaning of novel words, the learner has to focus attention to the stimuli in order to discriminate spectral and temporal phonetic contrasts, build new phonological representations, and associate these representations with meaning by re-cruiting working-, short-term-, episodic-, and semantic-memory processes. Finally, initial word representations have to be consolidated to build longer-lasting and more robust forms of these representations.

Based on its high temporal resolution, the ERPs method has been frequently used to capture the dynamics of word learning. Thereby, the building up of initial word representations has been shown to be reflected by the rapid emergence of a frontally distributed N400, a negative-going ERP component that develops between 300 and 600 ms after stimulus pres-entation. For instance, McLaughlin and colleagues (2004) were able to show increased N400 amplitudes in native English speakers after 14 hours of training with French words. Perfetti and colleagues (2005) revealed similar results after only 45 minutes of learning the meaning of low-frequency words.

Finally, in the case of learning a novel word's meaning from highly constraining meaningful contexts, Batterink and Neville (2011) showed the integration of such novel meanings into semantic networks after ten repetitions, Mestres-Missé and colleagues (2007) demonstrated the rapid development of an N400 after only three exposures to such words, and Borovsky and collaborators (2010) even after a single exposure to the words. In conclusion, while learning the meaning of novel words may seem to be slow and laborious, initial word representations can be built up within short training sessions and after only a few repetitions depending on the context in which novel words are presented.

In Section 10.3, we provided evidence for transfer from music to a variety of levels of language processing, including the perception of acoustic-phonetic parameters, segmentation, phonology, and syntax. To go one step beyond, we examined whether professional music training also facilitates semantic processing. We tested the hypothesis that professional music training facilitates word learning, designing an ecologically valid series of experi-ments aimed at tracking the electrophysiological dynamics of phonological categorization, semantic acquisition, as well as semantic retrieval (Dittinger et al., 2016) (see Figure 10.3).

Specifically, two groups of adult French speakers, comprising fifteen professional musicians and fifteen nonmusicians,- performed first a phonological categorization task, consisting of identifying nine natural Thai monosyllabic words containing either a simple voicing con-trast, a tonal, a vowel length, or an aspiration contrast. Importantly, two (/ba/and /pa/) out of these nine words were part of the French phonemic repertoire and therefore simple to categorize. By contrast, the other seven words contained contrasts which are linguistically irrelevant for French speakers, but relevant for quantitative or tonal languages (i.e. vowel length, pitch, and aspiration contrasts that are lexically relevant in Thai), resulting in more difficult categorization tasks (Dittinger et al., 2018). Following the categorization task, parti - cipants learned the meaning of these nine words through picture–word associations during a word-learning phase of about 6 minutes. Then, participants were tested for training suc-cess by asking them if a presented picture–word pair matched or mismatched the previously learned association (i.e. matching task). Moreover, to determine whether word learning was restricted to the picture–word pairs learned during the training phase or whether the meaning of the newly learned words was already integrated into semantic networks so that
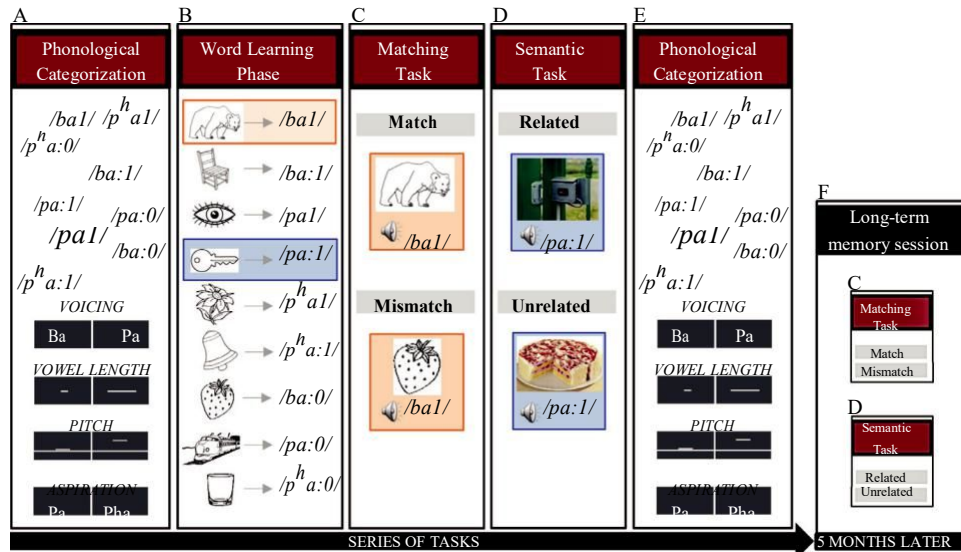
**Figure 10.3** Participants performed a series of tasks. First, in the phonological categorization task (A), nine natural Thai monosyllabic words had to be categorized based on voicing, vowel length, pitch, or aspiration contrasts. Second, in the word-learning phase (B), each word was paired with its respective picture. Third, in the matching task (C), the words were presented with one of the pictures, either matching or mismatching the previously learned associations. Fourth, in the semantic task (D), the words were presented with novel pictures that were either semantically related or unrelated to the novel words. Fifth, participants again completed the four subtasks of the phonological categorization task (E). Finally, participants came back 5 months after the main session to perform again the matching and semantic tasks (F).

priming effects generalized to new pictures, participants performed a semantic task during which novel pictures that had not been seen in the former task were presented in combin-ation with the previously learned words. They were asked to decide whether the picture and the word were semantically related or unrelated. Finally, participants were behaviourally retested 5 months after the main experimental session, to assess first, how long rapidly in-stalled word representations can last, and second, whether professional music training influ-ences long-term memory—two aspects that had not been investigated before.

The originality of this series of experiments is that EEG was simultaneously recorded in all these tasks (except in the long-term-memory session). This allowed us to follow the tem-poral dynamics of word learning from the early stages of word categorization and initial word encoding to subsequent stages of word-meaning retrieval once the novel words had been integrated into pre-existing semantic networks. Thus, we aimed at studying different processes underlying word learning that have previously only been explored in isolation in single experiments, to gain a more complete and integrated view of word learning. Since, as already noted, several studies have evidenced a positive influence of music not only on audi-tory perception, but also on attention, audiovisual integration abilities, as well as memory functions, we expected that professional musicians would be at an advantage to learn these novel words compared to non-musicians.

In line with this hypothesis, results showed that professional musicians learned the meaning of novel words more efficiently than controls, and this result was supported by both behavioural and electrophysiological data. Behaviourally, musicians outperformed non-musicians in word categorization and, as expected, group differences were particularly large for the tonal and aspiration contrasts (Dittinger et al., 2018). While these contrasts are not part of the French phonetic repertoire, the ability to discriminate them is inevitable for the acquisition of several foreign languages. For meaning acquisition, musicians performed similarly to non-musicians in the matching task, but outperformed non-musicians in the semantic task. These results were taken as evidence that musicians had already better integrated the novel words' meanings into semantic networks (i.e. enabling them to generalize the knowledge to novel pictures). Importantly, these behavioural results were supported by group differences in electrophysiological markers. In line with the development of a frontal
N400 during novel-word encoding (McLaughlin et al., 2004; Perfetti et al., 2005; Batterink &Neville, 2011; Mestres-Misse et al., 2007; Borovsky et al., 2010, 2012), all participants showed enhanced N400s over frontal scalp sites after the first half (i.e. only 3 minutes) of the word-learning phase. However, only musicians showed additional N400 increases over left centro-parietal scalp regions after the second half of the word-learning phase, suggesting that musicians were faster in encoding word meaning and integrating novel words into ex-isting semantic networks.

During the test phase (i.e. matching and semantic tasks), musicians were characterized by a typical centro-parietal N400 effect (Kutas & Federmeier, 2011) resulting from larger N400 amplitudes for unexpected (i.e. mismatching or unrelated) than for expected (i.e. matching or related) conditions (see Figure 10.4). By contrast, the N400 was still frontally distributed in non-musicians. In summary, while both groups showed the typical electrophysiological marker of word learning (i.e. the frontal N400) during the learning phase, only musicians showed semantic priming effects during the test phase that were similar to those typically found for known words. In line with this conclusion, a correlation between musical aptitudes and the amplitude of the semantic N400 effect was found for musicians, but not for non- musicians, thereby clearly pointing to a relationship between musicality and word learning. Finally, in the behavioural retest 5 months after the main session, musicians remembered
more words compared to non-musicians, thereby showing evidence for long-lasting word representations and an influence of music training on verbal long-term memory.

In a second step, data from this word-learning experiment was reanalysed by means of functional connectivity (Dittinger et al., 2017). Functional connectivity is defined as the statistical association or dependency among two or more anatomically distinct functional time series (Friston et al., 1996). Functional connectivity is a useful method for studying functional relationships between regions as a function of expertise. In the first part of this chapter, we reviewed some of the current models of speech processing that converge on the
view that two main processing streams, the ventral (i.e. sound-to-meaning mapping) and the dorsal (i.e. sound-to-articulation mapping) pathways, are involved in speech processing (Friederici, 2009). Specifically, the dorsal pathway relies on a fibre tract corresponding to the superior longitudinal fasciculus (SLF), sprawling from inferior parietal and superior- posterior temporal brain regions towards Broca's area and promoting auditory-to-motor mapping mechanisms. Recently, Lopez-Barroso and colleagues (2013) showed that word learning was correlated with the strength of functional and structural connectivity between Broca's and Geschwind's territory (including the angular gyrus (AG) and the supramarginal
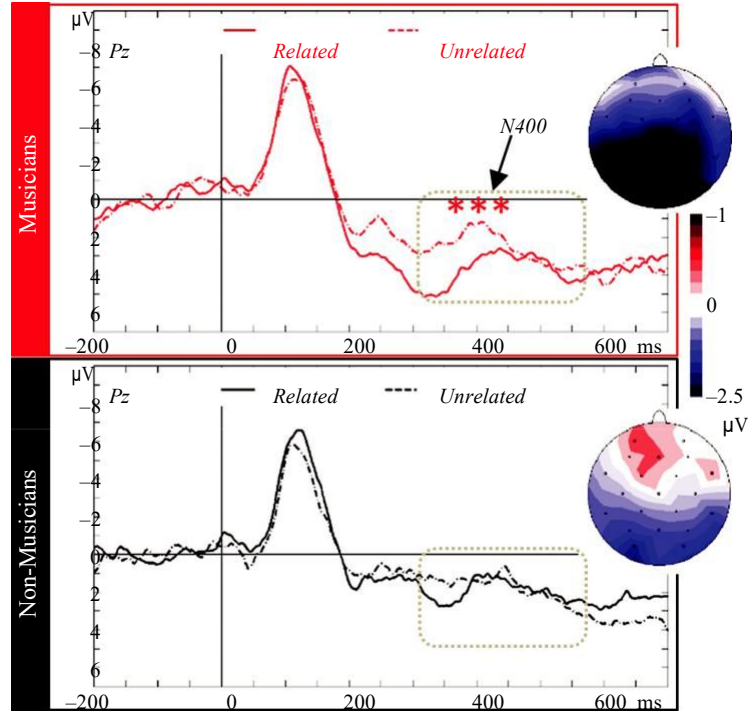
**Figure 10.4** Semantic task. ERPs recorded at parietal sites (Pz) are overlapped for semantically related (solid lines) and unrelated (dotted lines) words, separately for musicians (red) and non-musicians (black). Time in milliseconds is in abscissa, the amplitude of the effects in microvolts is in ordinate, time zero corresponds to word onset, and negativity is plotted upwards. The grey dotted rectangles represent the typical N400 latency window, and the level of significance of the related vs unrelated difference in the two groups is represented by asterisks (with ***p < 0.001). Topographic voltage distribution maps of the unrelated minus related differences in musicians and non-musicians are illustrated for the N400 component and voltage values are scaled from −2.5 to +1.0 μV.

gyrus (SMG)) in the left hemisphere. Moreover, there is evidence that the functional-structural architecture of the left dorsal processing stream is influenced by professional mu-sical training (Halwani et al., 2011; Klein et al., 2016; Oechslin et al., 2009).

Based on these results, we investigated functional connectivity between AG/SMG (region of interest (ROI) 1, Brodmann area (BA) 39/40) and Broca's area (ROI 2, BA 44/45) in the three tasks related to word learning previously described (i.e. word learning phase, matching tasks, and semantic tasks) and compared patterns of connectivity between mu-sicians and the non-musician controls. Specifically, we evaluated non-linear functional connectivity by using lagged coherence, that is, a measure of the variability of time differ-ences between two signals (e.g. coming from ROI 1 and from ROI 2) in a specific frequency band (Lehmann et al., 2006; Thatcher, 2012). We focused on theta (4–7 Hz) oscillations based on previous literature evidencing that theta reflects neuronal communications over long-range circuits and is a reliable frequency band to examine mnemonic processes (Ward, 2003).

As expected, results revealed increased left-hemispheric functional connectivity in mu-sicians compared to controls, but only in the semantic task. In addition, this increased connectivity was correlated with the cumulative number of training years. Results were in-terpreted as showing facilitated feed-forward and feed-backward exchanges between AG/ SMG and Broca's area in musicians, thereby facilitating the rehearsal and learning of novel words in musicians. These results are in line with previous results (López-Barroso et al., 2013; Klein et al., 2016) and with the group differences in ERPs and behaviour described previ-ously. Furthermore, the data indicates a relationship between the musicians' superiority in word learning and the temporal alignment of neural oscillations in the theta frequency band in the left dorsal stream.

Taken together, ERPs and functional connectivity revealed two main findings. First, word learning was reflected in the spatio-temporal dynamics of the N400 component: while ini-tial word learning was reflected by frontally increasing N400s, centro-parietal N400s only developed once the novel words started being integrated into the pool of well-known words. This difference in scalp distribution may reflect different cognitive processes and clearly points to distinct neural generators. For instance, frontal N400s to novel words are compat-ible with results showing that prefrontal and temporal brain regions are associated with the maintenance of novel information in working or short-term memory (Hagoort, 2014) and with the initial building up of episodic memory traces (Rodriguez-Fornells et al., 2009). By contrast, centro-parietally distributed N400s are in line with results showing that semantic representations are possibly stored in the left inferior parietal cortex (Catani & Ffytche, 2005) and/or in the left temporal lobe (Geranmayeh et al., 2015). Furthermore, our experi-mental design allowed us to track the fast spatio-temporal dynamics of word learning that were characterized by a shift in N400 distribution from frontal to parietal networks after only 3–6 minutes of training. These results open new perspectives for further research on brain plasticity and word learning. In addition, we provide the first behavioural evidence for the longevity of these rapidly established word representations, highlighting that even rapid brain plasticity can have long-lasting consequences.

Second, our results revealed that word learning was facilitated by professional music training. How can we account for such a transfer and what could be the implications of these results? As already mentioned in Section 10.3.2, two main interpretations have been pro-posed to explain why musicians learn novel words more efficiently than non-musicians. The first one, in terms of cascading effects, claims that enhanced auditory perception fa-cilitates word learning in musicians. Such an interpretation would be in line with bottom– up accounts of transfer effects. Support for this interpretation was provided by Wong and Perrachione (2007) and by Cooper and Wang (2012) who showed that both tone pitch

identification and musical aptitudes were significantly correlated with word-learning suc-cess in adult English speakers. To directly test for the causality of these effects, Cooper and Wang (2013) trained English non-musicians on the perception of Cantonese tones, and re-sults demonstrated that enhanced perception at the tone level significantly improved word learning. By contrast, the multidimensional interpretation acknowledges potential top– down influences on word learning, as well as interactions between the acoustic properties of sounds, task demands, and expertise of the listener.

While our experimental design does not allow for the disentangling of the cascade and multidimensional accounts, the present results clearly reveal that music training influences the semantic level of language processing, thereby going one step beyond previous studies

on transfer effects. Certainly, further studies are needed to replicate these results, possibly with children and older adults, and to disentangle the respective contribution of perceptive and cognitive functions to word learning, thereby possibly lifting the exciting secret of why musicians seem to be better at learning novel words.

## 10.5 Conclusion

The issue of music to speech transfer effects has generated great interest in the scientific community, as well as in the lay public, probably because music and speech are fascinating domains. The multidimensional aspects of music and speech—how they both rely on perceptual, cognitive, emotional, and motor processes through multiple interactions—are important new avenues for future research. Similarly, much more needs to be done to fully understand how these two abilities are implemented in the brain and whether they rely on shared or distinct neural resources. Based on the current state of knowledge, it is clear that results are tightly linked to the tasks and stimuli that are presented and to the methods that are used for data analysis. Also clear is that music exerts a profound influence on the brain's structural and functional organization, thanks to brain plasticity. It is worth noting in this respect that recent results demonstrated changes in brain electric activity in less than 3 minutes in a novel word-learning task (Dittinger et al., 2017). Taken together, these results open exciting new perspectives for the rehabilitation of patients (children, young adults, and older adults) with various neurological or psychiatric deficits. In this respect, music training may have a strong societal impact.

### Acknowledgements

### References

Abrams, D. A. *et al.* (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cereb. Cortex* **21**, 1507–1518.

Alexander, J. A., Wong, P. C., & Bradlow, A. R. (2005). *Lexical tone perception in musicians and non-musicians*. Paper presented at 9th European Conference on Speech Communication and Technology, Lisbon, Portugal (pp. 397–400).

Asaridou, S. S. & McQueen, J. M. (2013). Speech and music shape the listening brain: evidence for shared domain-general mechanisms. *Front. Psychol.* **4**, article 321.

Barrett, K. C., Ashley, R., Strait, D. L., & Kraus, N. (2013). Art and science: how musical training shapes the brain. *Front. Psychol.* **4**, article 713.

Batterink, L. & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: an event-related potential study. *J. Cogn. Neurosci.* **23**, 3181–3196.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* **403**, 309–312.

Bergman Nutley, S., Darki, F., & Klingberg, T. (2014). Music practice is associated with development of working memory during childhood and adolescence. *Front. Hum. Neurosci.* **7**, 926.

Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends Cogn. Sci.* **17**, 89–98.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *Front. Psychol.* **2**, article 94.

Bidelman, G. M. & Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel per-ception. *J. Neurosci.* **35**, 1240–1249.

Bidelman, G. M. & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res.* **1355**, 112–125.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain Cogn.* **77**, 1–10.

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage* **79**, 201–212.

Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brain-stem and auditory cortex contributes to enhanced categorical speech perception in musi-cians. *Eur. J. Neurosci.* **40**, 2662–2673.

Boland, J. E. (2014). A history of psycholinguistics: the pre-Chomskyan era. *Historiographia Linguistica* 41(1), 168–175.

Bornkessel-Schlesewsky, I. & Schlesewsky, M. (2013). Reconciling time, space and function: a new dorsal-ventral stream model of sentence comprehension. *Brain Lang.* **125**, 60–76.

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Lang. Learn. Dev.* **8**, 278–302.

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: event-related potentials index single-shot contextual word learning. *Cognition* **116**, 289–296.

Brandler, S. & Rammsayer, T. H. (2003). Differences in mental abilities between musicians and non-musicians. *Psychol. Music* **31**, 123–138.

Brauer, J., Anwander, A., Perani, D., & Friederici, A. D. (2013). Dorsal and ventral pathways in language development. *Brain Lang.* **127**, 289–295.

Cason, N. & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia* **50**, 2652–2658.

Catani, M. & Ffytche, D. H. (2005). The rises and falls of disconnection syndromes. *Brain J. Neurol.* **128**, 2224–2239.

Chan, A. S., Ho, Y. C., & Cheung, M. C. (1998). Music training improves verbal memory. *Nature* **396**, 128.

Chartrand, J.-P. & Belin, P. (2006). Superior voice timbre processing in musicians. *Neurosci. Lett.* **405**, 164–167.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J. Cogn. Neurosci.* **20**, 226–239.

Chobert, J., Francois, C., Velay, J.-L., & Besson, M. (2014). Twelve months of active musical training in 8-to 10-year-old children enhances the preattentive processing of syllabic dur-ation and voice onset time. *Cereb. Cortex* **24**, 956–967.

Cooper, A. & Wang, Y. (2013). Effects of tone training on Cantonese tone-word learning. *J. Acoust. Soc. Am.* **134**, EL133–EL139.

Cooper, A. & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *J. Acoust. Soc. Am.* **131**, 4756–4769.

Dehaene-Lambertz, G. & Spelke, E. S. (2015). The infancy of the human brain. *Neuron* **88**, 93–109.

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* **64**, 135–168.

Dittinger, E. *et al.* (2016). Professional music training and novel word learning: from faster semantic encoding to longer-lasting word representations. *J. Cogn. Neurosci.* 28(10), 1584–1602.

Dittinger, E., Valizadeh, S. A., Jäncke, L., Besson, M., & Elmer, S. (2017). Increased functional connectivity in the ventral and dorsal streams during retrieval of novel words in professional musicians. *Hum. Brain Mapp.* **39**, 722–734.

Dittinger, E., D'império, M., & Besson, M. (2018). Enhanced neural and behavioral processing of a non-native phonemic contrast in professional musicians. *Eur. J. Neurosci.* 1–13. doi:10.1111/ ejn.13939

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science* **270**, 305–307.

Elmer, S., Meyer, M., & Jancke, L. (2012). Neurofunctional and behavioral correlates of phon-etic and temporal categorization in musically trained and untrained subjects. *Cereb. Cortex* **22**, 650–658.

Farbood, M. M., Heeger, D. J., Marcus, G., Hasson, U., & Lerner, Y. (2015). The neural pro-cessing of hierarchical structure in music and speech at different timescales. *Front. Neurosci.* **9**, 157.

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press.

Fitch, W. T. & Reby, D. (2001). The descended larynx is not uniquely human. *Proc. R. Soc. B Biol. Sci.* **268**, 1669–1675.

Fitzroy, A. B. & Sanders, L. D. (2012). Musical expertise modulates early processing of syntactic violations in language. *Front. Psychol.* **3**, 603.

Francois, C., Chobert, J., Besson, M., & Schon, D. (2013). Music training for the development of speech segmentation. *Cereb. Cortex* **23**, 2038–2043.

Francois, C. & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cereb. Cortex* **21**, 2357–2365.

Franklin, M. S. *et al.* (2008). The effects of musical training on verbal memory. *Psychol. Music* **36**, 353–365.

Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends Cogn. Sci.* **13**, 175–181.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* **91**, 1357–1392.

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn. Sci.* **16**, 262–268.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain dif-ferentiates human and non-human grammars: functional localization and structural con-nectivity. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2458–2463.

Friederici, A. D. & Singer, W. (2015). Grounding language processing on basic neurophysio-logical principles. *Trends Cogn. Sci.* **19**, 329–338.

Friston, K. J., Frith, C. D., Fletcher, P., Liddle, P. F., & Frackowiak, R. S. (1996). Functional top-ography: multidimensional scaling and functional connectivity in the brain. *Cereb. Cortex* **6**, 156–164.

George, E. M. & Coch, D. (2011). Music training and working memory: an ERP study. *Neuropsychologia* **49**, 1083–1094.

Geranmayeh, F., Leech, R., & Wise, R. J. S. (2015). Semantic retrieval during overt picture description: left anterior temporal or the parietal lobe?*Neuropsychologia* **76**, 125–135.

Giraud, A.-L. & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517.

Giraud, A.-L. *et al.* (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* **56**, 1127–1134.

Gogtay, N. *et al.* (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8174–8179.

Golestani, N. & Pallier, C. (2007). Anatomical correlates of foreign speech sound production. *Cereb. Cortex* **17**, 929–934.

Golestani, N. & Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *NeuroImage* **21**, 494–506.

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Curr. Opin. Neurobiol.* **28**, 136–141.

Halwani, G. F., Loui, P., Rüber, T., & Schlaug, G. (2011). Effects of practice and experience on the arcuate fasciculus: comparing singers, instrumentalists, and non-musicians. *Front. Psychol.* **2**, 156.

Harasty, J., Seldon, H. L., Chan, P., Halliday, G., & Harding, A. (2003). The left human speech- processing cortex is thinner but longer than the right. *Laterality* **8**, 247–260.

Hauser, M. D. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579.

Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402.

Jakobson, L. S., Lewycky, S. T., Kilgour, A. R., & Stoesz, B. M. (2008). Memory for verbal and visual material in highly trained musicians. *Music Percept. Interdiscip. J.* **26**, 41–55.

Jäncke, L. (2012). The relationship between music and language. *Front. Psychol.* **3**, article 123.

Janus, M., Lee, Y., Moreno, S., & Bialystok, E. (2016). Effects of short-term music and second-language training on executive control. *J. Exp. Child Psychol.* **144**, 84–97.

Jentschke, S. & Koelsch, S. (2009). Musical training modulates the development of syntax pro-cessing in children. *NeuroImage* **47**, 735–744.

Klein, C., Liem, F., Hänggi, J., Elmer, S., & Jäncke, L. (2016). The 'silent' imprint of musical training. *Hum. Brain Mapp.* **37**, 536–546.

Koelsch, S. *et al.* (2002). Bach speaks: a cortical 'language-network' serves the processing of music. *NeuroImage* **17**, 956–966.

Kraus, N. & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* **11**, 599–605.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* **5**, 831–843.

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Dev. Sci.* **10**, 110–120.

Kutas, M. & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* **62**, 621–647.

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

Lee, C.-Y. & Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musi-cians and nonmusicians. *J. Acoust. Soc. Am.* **124**, 3235–3248.

Lee, H. & Noppeney, U. (2011). Long -term music training tunes how the brain temporally binds signals from multiple senses. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1441–1450.

Lehmann, D., Faber, P. L., Gianotti, L. R. R., Kochi, K., & Pascual-Marqui, R. D. (2006). Coherence and phase locking in the scalp EEG and between LORETA model sources, and microstates as putative mechanisms of brain temporo-spatial functional organization. *J. Physiol. Paris* **99**, 29–36.

Levitin, D. J. & Menon, V. (2003). Musical structure is processed in 'language' areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage* **20**, 2142–2152.

Lima, C. F. & Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emot. Wash. DC* **11**, 1021–1031.

López-Barroso, D. *et al.* (2013). Word learning is mediated by the left arcuate fasciculus. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13168–13173.

Ma, W. & Thompson, W. F. (2015). Human emotions track changes in the acoustic environ-ment. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14563–14568.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci.* **4**, 540–545.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J. Cogn. Neurosci.* **18**, 199–211.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* **19**, 1453–1463.

Mathias, B., Palmer, C., Perrin, F., & Tillmann, B. (2015). Sensorimotor learning enhances expectations during auditory perception. *Cereb. Cortex* **25**, 2238–2254.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat. Neurosci.* **7**, 703–704.

Mestres-Misse, A., Rodriguez-Fornells, A., & Munte, T. F. (2007). Watching the brain during meaning acquisition. *Cereb. Cortex* **17**, 1858–1866.

Meyer, M., Elmer, S., & Jäncke, L. (2012). Musical expertise induces neuroplasticity of the planum temporale. *Ann. N.Y. Acad. Sci.* **1252**, 116–123.

Milner, A. D. & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia* **46**, 774–785.

Mishkin, M. & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial func-tions of parieto-preoccipital cortex in monkeys. *Behav. Brain Res.* **6**, 57–77.

Moreno, S. *et al.* (2009). Musical training influences linguistic abilities in 8-year-old chil-dren: more evidence for brain plasticity. *Cereb. Cortex* **19**, 712–723.

Moreno, S. *et al.* (2011). Short-term music training enhances verbal intelligence and executive function. *Psychol. Sci.* **22**, 1425–1433.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nat. Rev. Neurosci.* **3**, 473–478.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15894–15898.

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* **88**, 1281–1296.

Oechslin, M. S., Imfeld, A., Loenneker, T., Meyer, M., & Jäncke, L. (2009). The plasticity of the superior longitudinal fasciculus as a function of musical expertise: a diffusion tensor imaging study. *Front. Hum. Neurosci.* **3**, 76.

Ott, C. (2011). Processing of voiced and unvoiced acoustic stimuli in musicians. *Front. Psychol.* **2**, article 195.

Overath, T., McDermott, J. H., Zarate, J. M., & Poeppel, D. (2015). The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nat. Neurosci.* **18**, 903–911.

Pallesen, K. J. *et al.* (2010). Cognitive control in auditory working memory is enhanced in mu-sicians. *PLoS ONE* **5**, e11120.

Pantev, C., Lappe, C., Herholz, S. C., & Trainor, L. (2009). Auditory-somatosensory integration and cortical plasticity in musical training. *Ann. N.Y. Acad. Sci.* **1169**, 143–150.

Paraskevopoulos, E., Kraneburg, A., Herholz, S. C., Bamidis, P. D., & Pantev, C. (2015). Musical expertise is related to altered functional connectivity during audiovisual integration. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12522–12527.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012) . Musical ex-pertise induces audiovisual integration of abstract congruency rules. *J. Neurosci.* **32**, 18196–18203.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2014). Multisensory in-tegration during short-term music reading training enhances both uni-and multisensory cortical processing. *J. Cogn. Neurosci.* **26**, 2224–2238.

Perani, D. *et al.* (2011). Neural language networks at birth. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16056–16061.

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 1281–1292.

Péron, J., Frühholz, S., Ceravolo, L., & Grandjean, D. (2016). Structural and functional con-nectivity of the subthalamic nucleus during vocal emotion decoding. *Soc. Cogn. Affect. Neurosci.* **11**, 349–356.

Rauschecker, J. P. & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12**, 718–724.

Ripollés, P. *et al.* (2014). The role of reward in word learning and its implications for language acquisition. *Curr. Biol. CB* **24**, 2606–2611.

Roden, I., Kreutz, G., & Bongard, S. (2012). Effects of a school-based instrumental music program on verbal and visual memory in primary school children: a longitudinal study. *Front. Psychol.* **3**, article 572.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 3711–3735.

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J. Neurosci.* **31**, 3843–3852.

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech rhythm facili-tates syntactic ambiguity resolution: ERP evidence. *PloS One* **8**, e56000.

Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Curr. Dir. Psychol. Sci.* **12**, 110–114.

Santoro, R. *et al.* (2014). Encoding of natural sounds at multiple spectral and temporal resolu-tions in the human auditory cortex. *PLoS Comput. Biol.* **10**, e1003412.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychol. Sci.* **15**, 511–514.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science* **267**, 699–701.

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

Schneider, P. *et al.* (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* **5**, 688–694.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* **41**, 341–349.

Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Percept. Interdiscip. J.* **29**, 255–267.

Schulze, K. & Koelsch, S. (2012). Working memory for speech and music. *Ann. N.Y. Acad. Sci.* **1252**, 229–236.

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

Seldon, H. L. (1981). Structure of human auditory cortex. I. Cytoarchitectonics and dendritic distributions. *Brain Res.* **229**, 277–294.

Sluming, V. *et al.* (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *NeuroImage* **17**, 1613–1622.

Thatcher, R. W. (2012). Coherence, phase differences, phase shift, and phase lock in EEG/ERP analyses. *Dev. Neuropsychol.* **37**, 476–496.

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19027–19032.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emot. Wash. DC* **4**, 46–64.

Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in mu-sical priming. *Ann. N.Y. Acad. Sci.* **999**, 209–211.

Trimmer, C. G. & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts rec-ognition of emotional speech prosody. *Emot. Wash. DC* **8**, 838–849.

Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., & Østergaard, L. (2006). It don't mean a thing... Keeping the rhythm during polyrhythmic tension activates language areas (BA47). *NeuroImage* **31**, 832–841.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cogn. Sci.* **7**, 553–559.

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short- term memory for verbal and musical sequences: comparing phonological similarity and pitch proximity. *Mem. Cogn.* **38**, 163–175.

Wong, P. C. M. & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Appl. Psycholinguist.* 28(4), 565–585.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* **10**, 420–422.

Zatorre, R. J. & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* **11**, 946–953.

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLOS ONE* **9**, e99868.

**APPENDIX 10.1**

# Auditory Stimuli

The four consonant-vowel (CV) syllables (two natural German consonant-vowel syllables and two reduced-spectrum analogues; see Audio 10.1) were used in three previous publications in order to assess putative advantages of musicians in processing fast-changing phonetic cues. These stimuli consisted of the German CV syllables /ka/(voiceless initial consonant) and /da/ (voiced initial consonant) as well as of its reduced-spectrum analogues. The duration of the syllables was about 350 ms, and the voice-onset time (VOT) of /da/and /ka/was approximately 13 ms and 53 ms, respectively. For the reduced-spectrum analogues, spectral information was removed from the CV syllables by replacing the frequency-specific information in a broad fre-

quency region with band-limited white noise (band 1: 500–1500 Hz, band 2: 2500–3500 Hz). Amplitude and temporal cues were preserved in each spectral band, resulting in double-band- pass filtered noise with temporal CV-amplitude dynamics. A detailed description of the stimuli can be found here:

Elmer, S., Meyer, M., & Jäncke, L. (2012). Neurofunctional and behavioral correlates of phon-etic and temporal categorization in musically trained and untrained subjects. *Cerebral Cortex*, **22**, 650–658.

Elmer, S., Hänggi, J., Meyer, M., & Jäncke, L. (2013). Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex*, **49**, 2812–2821.

Elmer, S., Hänggi, J., & Jäncke, L. (2016). Interhemispheric transcallosal connectivity between the left and right planum temporale predicts musicianship, performance in temporal speech processing, and functional specialization. *Brain Structure and Function*, **221**, 331–344.