

## Oxford Handbooks Online

### Music and Language

Daniele Schön and Benjamin Morillon

The Oxford Handbook of Music and the Brain

*Edited by Michael H. Thaut and Donald A. Hodges*

Subject: Psychology, Cognitive Neuroscience Online Publication Date: Oct 2018

DOI: 10.1093/oxfordhb/9780198804123.013.16

### Abstract and Keywords

This chapter begins with a historical approach that explains why music and language have long been considered as modular and as such independent brain functions. It shows some methodological limitations of the neuropsychological and neuroimaging approach in interpreting differences between music and language neural substrates. The chapter advocates for the use of a truly comparative approach to study differences and commonalities between music and language. It then highlights several common functions and operations necessary in both language and music processing and presents the sharing resource hypothesis. The study of the effects of music training and music stimulation on language processing is a privileged avenue to unravel the relation of common operations and their specific nature at the algorithmic level. The chapter concludes with a focus on the temporal dimension and the dynamic nature of oscillatory activity and their role in temporal prediction in music and language.

Keywords: modularity, shared resources, comparative approach, oscillatory activity, brain dynamics, temporal prediction, rhythm

---

# Introduction

While music and language may differ in terms of their structures and functions, they share the distinctive feature of being dynamically organized in time; the information they carry is intrinsically contained in the temporal dimension. A frequently asked question is whether music and language are processed by similar or different brain regions, neural networks, or cortical oscillatory processes, and to what extent the brain circuitry is specialized compared to other stimuli. In order to tackle these issues, it is worth keeping in mind some principles. Nikolaas Tinbergen and David Marr described different levels of analysis that must, in their valuable opinion, be taken into account if one wants to understand behavior and complex systems (Marr, 1982; Tinbergen, 1963). Marr's three levels of analysis (computational, algorithmic, and implementational) are particularly suited to study brain functions. Because music and language differ in terms of surface acoustic features and convey different purposes, the computations needed to process them differ. On the other hand, at the implementation level, the same organ and a myriad of cells process both music and language. The key program in modern cognitive neurosciences is thus to tackle the algorithmic level (Poeppel, 2012): Are similar or different algorithms involved in the processing of music and language? And what are they? In this chapter, we will begin with a historical perspective, where the human brain is described from a phrenological viewpoint. Then, we will describe the common functions and operations in music and language, the methodological limitations in current approaches, and portray the resource-sharing hypothesis. We will then describe the interdependency between music and language, notably how musical training improves language skills, before trying to bridge music and language in a single context. We will conclude by describing a promising avenue: studies that adopt a dynamical standpoint to understand music and language.

## On Modularity of Music and Language

From a historical perspective, the study of the comparison of music and language brain functions dates back to the early observations of deficits acquired following a brain lesion. Since then, language and musical disorders are described with different terms: aphasia and amusia. This distinction comes along with a deeper distinction between language and musical domains that at the end of the nineteenth century had been the object of structural or historical formalization following very different paths. Language is analyzed as a formal system of different elements, while music is viewed in a historical perspective as an artistic behavior. Language and music are thus viewed as two highly distinct human domains. In this context, the observation of selective impairment of

language or musical abilities fits in very well and also complies with the idea that different functions are implemented in different brain regions.

The birth of cognitive sciences is strongly influenced by this vision of language as a specific and uniquely human function with dedicated neural structures and music as a different human “artistic” function. At the end of the 1950s Noam Chomsky was convinced that the principles underlying language structure are biologically determined: every individual has the same language potential because it is genetically transmitted, independently of socio-cultural differences. This scientific and political view of language development has had a tremendous impact in the field of linguistics, cognitive sciences, and neurosciences. It stands in clear contrast with that of another giant of psychology, B. F. Skinner. Skinner considered the mind as a *tabula rasa* whereon only experience could add knowledge. The two giants faced each other in an intellectual duel. The most famous attack of Chomsky (1959) is the argument on the poverty of the stimulus: the child exposed to a limited amount of linguistic stimuli is able to generalize to new linguistic constructions using the rules acquired on the initial set. According to Chomsky, the trial and error learning mechanisms defended by the behaviorists would not be an appropriate model to language acquisition since language is acquired by listening to correct sentences. This observation, as well as the fact that a confined brain lesion such as in Broca’s area may induce a specific language deficit (agrammatism), led to Chomsky’s suggestion that syntactic knowledge may be partly innate. Curiously, Chomsky did not remark that music acquisition follows very similar principles as language acquisition: early acquisition, generativity, and learning from correct structures.

Chomsky’s work strongly inspired that of Jerry Fodor who in the early 1980s wrote *The Modularity of Mind* (1983). The mind (and the brain) would be organized in independent modules with specific functions. Again Fodor’s view is strongly influenced by and reinforces the results of the neuropsychological literature, digging deeper and deeper into specific deficits following focal brain lesions. The functioning of the brain seems quite simple: every region has a specific functional role and the lesion causes a deficit that may be very specific, for instance affecting independently nouns and verbs processing (Hillis & Caramazza, 1995).

It is within this context that the field of neuropsychology of music develops, beyond previous anecdotal accounts. As every new field, the desire to gain identity and acknowledgment is strong. Music is thus studied as a special human faculty with dedicated brain areas. This vision is also constrained by some intrinsic limitations of the field of neuropsychology of music. First, research on musical skills in brain lesion patients requires a neurologist or neuropsychologist with a musical background. Indeed, while testing language skills may appear a simple task, assessing musical abilities definitely requires special skills, even more so in the era of “pencil and paper.” The second limitation is the Western idea that music is the prerogative of a few people, called musicians, and thus it only make sense to assess musical abilities in experienced musicians such as composers, conductors, or performers with musical education (Basso & Capitani, 1985; Luria, Tsevetkova, & Futer, 1965). Altogether this gives access to a

limited amount of data strongly influenced by the modular approach, with musical functions clearly distinct from other human abilities. This is the vision that is well summarized in the article entitled “Modularity of Music Processing” (Peretz & Coltheart, 2003): several single case studies are used to defend not only the hypothesis of modularity of music and language, but also the modularity of different levels of music processing.

However, focusing on a single function, even more when using a single methodological approach (for instance, brain lesions) will systematically lead toward a modular interpretation of reality. In other words, focusing only on language syntactic processing in Broca’s aphasics will necessarily lead one to conclude that the left inferior frontal operculum is involved (or not) in syntactic processing. This may be in turn interpreted in a modular perspective: syntax is independently and specifically processed in the left frontal operculum. By contrast a comparative approach will give a broader and more complex picture. Patel (2003), considering language and musical syntax, claims that, while these may seem very different, there are several commonalities, such as the need to build an integrated flow of information that takes into account a certain number of rules. Here we can clearly see all the power of the comparative approach that requires us to go beyond a circular definition of cognitive function (e.g., syntax is syntax) in order to compare apparently different function (e.g., syntax and harmony) that can possibly be redefined in terms of a more elementary function with a greater psychobiological validity. In the case of syntax and harmony, finding common substrates requires one to redefine the object of the study (i.e., some elementary operation common to both).

With the advent of the neuroimaging era, while the first two decades have been dominated by a modular approach, the last decade has put the accent on the importance of the network and its connections. Cognitive neurosciences have also gained access to the functioning of non-pathological brains using highly sophisticated experimental designs. This has allowed a breakdown of both language and music processing into more elementary operations. If the search of biomarkers has somewhat consolidated the innatist model, several major criticisms have been developed further. For instance, studies on the zebra finch, a species of bird well known for their ability to learn new songs, showed that it is the learning process that alters the neuronal circuits. The maturation of synaptic inhibition onto premotor neurons is correlated with learning but not age (Vallentin, Kosche, Lipkind, & Long, 2016). This shows that even in a species wherein one could think that the rules governing song acquisition are genetically encoded, the environment plays an important role. Of course putting a zebra finch in a cage with a cat (and assuming that the cat did not eat the bird) would not allow the bird to learn how to meow, that is to say that the genes do play a major role, of course. When considering the case of language and music, two extremely refined forms of communication, while the human species specificity is certainly genetically encoded, this does not imply that whatever allows the development of language is specific to language and not shared with music. In other words, if language and music are specific to humans

in their capacity to convey an extraordinary amount of information, one should not misinterpret this in terms of different evolutionary or developmental trajectories of language and music.

Psychology of music and neurosciences of music are recent fields of research. The major limitation of new disciplines (and of humans) is their strong desire to build their own identity, which often occurs to the detriment of considering neighboring disciplines (and identities). Our field has also yielded to this temptation by building musical cognitive models that have initially ignored other potentially inspiring and similar domains, such as language for instance. We will see now that music shares several cognitive operations with language.

## Common Functions and Operations in Music and Language

Both language and music serve a highly sophisticated communicative function. While we will refrain here from giving a definition of what is language and what is music, it is important to keep in mind that they require a huge amount of different perceptual and cognitive operations.

To perceive both music and language, one of the first operations that need to be implemented is the possibility to discriminate sounds. The two phonemes [d] and [t] are quite similar but need to be distinguished as it is the case for a C and B in music or for the same pitch played by an oboe or a bassoon. Sounds can be characterized in terms of a limited number of spectral features and these features are relevant to both musical and linguistic sounds. The analysis of the acoustic features of sounds takes place in the cochlea and in several subcortical relays up to the primary auditory cortex. There is a suggestion that the auditory cortex may be asymmetric in the use of temporal windows of analysis, with the left auditory cortex preferring short windows of integration and the right auditory cortex preferring longer windows (Giraud et al., 2007; Poeppel, 2003; Zatorre, Belin, & Penhune, 2002). This hypothesis has been used to defend the idea that language, requiring short windows of analysis to discriminate consonants, is preferentially processed in the left hemisphere, while music, requiring longer windows of analysis to discriminate pitch, is preferentially processed in the right hemisphere. While the debate is still open, one should keep in mind that language perception is not just consonant discrimination, but also requires us to take into account other features, such as pitch in tone or stresses, that require longer windows of analysis. On the other side, music is often considered in our Western society and by non-musicians as mostly relying on pitch discrimination. However, any good musician will claim that an extremely important feature of music is the sound quality, which is not stationary as pitch and requires short analysis windows. The scenario is thus more complicated than it is often

depicted and the idea of the cortex performing parallel processing on any acoustic input, yielding to the extraction of complementary piece of information, seems necessary to overcome the simplistic monolithic distinction between language and music.

Generating different patterns of neuronal responses to every sound would yield, in everyday life, to an infinite number of sound representations. This is why sounds are categorized. Two acoustically different tokens of [b] will thus be perceived as a unique [b]. Two different high Es of the violin will be perceived as E, even if one is slightly lower than the other; an A and a C note of a piano will be perceived as “piano” sounds. Categorization is necessary and common to both language and music and it allows us to make sense of the world, by reducing its intrinsic variety to a finite and limited number of categories. Categorical representations of sounds are possibly distributed across neuronal populations within the human auditory cortices, including primary auditory areas (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Rauschecker & Scott, 2009; Rauschecker & Tian, 2000; Staeren, Renvall, De Martino, Goebel, & Formisano, 2009), although motor regions seem also to play a role in representing, for instance, phonemic acoustic features (Cheung, Hamilton, Johnson, & Chang, 2016).

We rarely perceive sounds in isolation, but rather in a complex flow. This requires us to build a structure that evolves in time, taking into account the different phonemes of a sentence or tones of a melody. Building such a structure requires at the very least a working memory capacity that allows manipulating sound representations. Sounds are grouped into larger units and this grouping depends upon our previous experience with these sounds. In other words we take advantage of our previous experience with the world and build multiple statistical distributions of sounds. Different distributions will account for different grouping strategies: for instance, streaming a specific voice or musical instrument in a cocktail party or in a musical ensemble (Elhilali & Shamma, 2008); grouping phonemes together or tones to build words or melodies according to the transitional probabilities of phonemes or tones (Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Schön et al., 2008). These statistical distributions are built on the memory traces of what we have previously perceived and strongly influence our upcoming perception of the world. In fact, following these statistical distributions, several rules may emerge that allow us to simplify even more the complex and continuous auditory flow. Importantly in both language and music, the distributions can also be computed onto symbolic unit.

These distributions or internal models at different feature levels have two major consequences. The first, cited earlier, is that they allow us to generate new sequences having similar statistical properties—in other words, new sentences or melodies complying to the rules of the musical or linguistic system. The second is that they allow us to make accurate predictions on upcoming events. Listening to a person speaking or playing we will be able to anticipate, to a certain degree, what and when is going to be said/played. Considering the very fast and changing nature of the auditory flow, this ability is of utmost importance and it explains why sounds (phonemes or tones) missing

from a speech or musical signal can be restored by the brain and appear to be heard (DeWitt & Samuel, 1990). In this respect music is particularly challenging because it may require us to anticipate simultaneously several distinct streams of features. For instance, when listening to a symphony orchestra or a string quartet, several melodic lines take place at the same time and need to be anticipated in order to perceive a sense of continuity in the music.

Overall, language and music are characterized by a limited set of acoustic features, categorized by the human brain into a limited set of representations, and subjected to similar rules of statistical learning.

## Overlap and Resource Sharing

Since most research in cognitive neuroscience has been guided by the assumption that brain regions are specialized for a given function, studies on music and language have addressed the question of whether music and language share common neural substrates. This has been often referred as the notion of overlap (Patel, 2011). The idea is simple. If one could show that there is a strong overlap for music and language processing, this would go against a modular and domain-specific view. However, there are more problems with this approach than one might imagine. We will review them briefly in the following section together with some neuroimaging findings.

The first problem is of purely methodological order. Indeed, many published works using fMRI, including those comparing music and speech processing, use a subtraction logic. Namely, results are a statistical contrast that only allows us to see which areas show a greater signal compared to another condition. This is referred to the tip of the iceberg problem. Indeed, it may well be that by contrasting a language and a music task one finds a peak in a given region. This is then interpreted as a specific area dedicated to language (or to music, depending upon the direction of the subtraction; see for instance Rogalsky & Hickok, 2011). However, this completely ignores the possibility that there is a large common substrate that is invisible when making the subtraction (e.g., 100 and 101 share 100, but 101–100 only shows 1). This approach is Manichean and suffers from its lack of quantitative descriptions. These studies have, therefore, a methodological bias toward highlighting differences rather than commonalities.

A second series of problems is the experimental designs that have been used. Indeed, only a few studies have directly used the same participants and the same experiment music and language processing. Comparing results across studies will also tend to show differences that may not be due to brain computations but to differences in the populations, acquisition, or analysis pipelines. Even when assessing music and language processing in the same participant, there remains the challenge of comparing comparable conditions. This goes beyond the fact that speech and music stimuli by nature are different acoustically, insofar as if this was the only difference it should affect only the

primary auditory cortex. The real challenge is to define the proper elementary operation and balance the difficulty level of the task across linguistic and musical stimuli. Defining the operation is already quite challenging because it requires a “good” model of what to compare. Of course comparing music and language does not make any sense, because there is no such a thing as a function for music in the brain. Thus, music and language need to be reduced to more elementary functions as described earlier. But even comparing syntactic processing is not trivial. Indeed, one needs to choose which syntactic level to compare in language (syntactic embedding or gender agreement do not imply the same operations) and find the good analogy in music. Then, the researcher is still left with a complicated issue, that of the difficulty level. For instance, in comparing the role of pitch in music and in language prosody, one should ascertain that the difficulty level of the task is comparable across material rather than using a fixed criterion (e.g., detect a 15 percent pitch change) that may be trivial with music but not with speech (Schön, Magne, & Besson, 2004).

Another important issue is raised by Peretz and colleagues:

It is important to keep in mind that neural overlap does not necessarily entail neural sharing. The neural circuits established for musicality may be intermingled or adjacent to those used for a similar function in language and yet be neurally separable. For example, mirror neurons are interspersed among purely motor-related neurons in pre-motor regions of the macaque cortex (Rizzolatti & Craighero, 2004). Similarly, the neurons responsible for the computation of some musical feature may be interspersed among neurons involved in similar aspects in speech.

(Peretz, Vuvan, Lagrois, & Armony, 2015, p. 3)

The problem that is raised here is the scale problem of human anatomy. Historically, there has been a very rough distinction of music and language in terms of hemispheric dominance and this led many people to believe that language is processed by the left hemisphere and music by the right hemisphere. We now clearly know that this is not the case (Lindell, 2006; Vigneau et al., 2011). Then, there have been more specific claims that the left Broca’s area would be language specific, but this has also been falsified, by showing for instance that musical harmony (Koelsch et al., 2002; Maess, Koelsch, Gunter, & Friederici, 2001) and rhythm processing (Herdener et al., 2012) are mediated by the same regions processing language syntax (Friederici & Kotz, 2003). Further work based on multivariate pattern analysis has shown that within overlapping regions, distinct brain patterns of responses can be found to linguistic and musical sounds (Abrams et al., 2010; Fedorenko, McDermott, Norman-Haignere, & Kanwisher, 2012). However, these differences could be accounted for in terms of differences in the stimuli manipulation or in the task. For instance, Abrams et al. (2010) compared scrambled versions of music and speech to normal music and speech and used a fixed scrambling window of 350 ms. As the authors acknowledge, it could be that music and speech have inherently different acoustical regularities and structures, rendering one material more “scrambled” than the



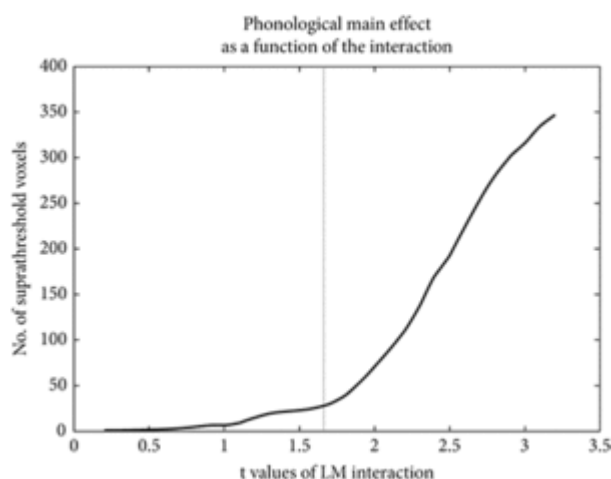
other. Also, different patterns of activation in common brain areas may result from the same neural population reacting differently to music and language (Kunert & Slevc, 2015).

The argument raised by Peretz advocates for the possibility of music dedicated neurons, adjacent to language dedicated neurons. While this is of course a non-falsifiable hypothesis for the moment, one should not think of music or language as a whole, but in terms of precisely defined elementary operations. If these operations are required with both language and music material, then there would be no reason for the brain to produce two extremely intermingled networks computing the same algorithm. On the other side it is clear that the rules determining gender agreement or those affecting tonality modulations are necessarily represented in different neural networks. Thus, claiming that differences may always subsist at a smaller scale is a recursive argument that does not really add much to the debate (besides the fact that at a quantum level, music and language can be described by the same equations). In our view, the major advances will not come from single unit recordings showing specific neurons to the last chord of a precise Haydn piano sonata, but rather from neurocomputational models precisely describing what particular operations are subtended by a given neural network when listening to speech and to music.

A more promising approach seems to us to study whether two different levels of music and language processing interact or not. Indeed, the interaction is a measure of the extent to which two processes influence each other and as such it can be used to infer that one process is not independent of the other. Several studies have tackled this issue by using interference paradigms. For instance Slevc and colleagues (Slevc, Rosenberg, & Patel, 2009) measured the reading time of garden path sentences and found that it was influenced by simultaneous presentation of irrelevant harmonically unexpected chords while it was not affected by timbrally unexpected chords (e.g., a different instrument). These results have been interpreted as evidence for shared music-language resources processing structural (syntactic) relations. The task-irrelevant music being processed automatically, it uses some resources resulting in a suboptimal processing of the language syntactic relations. Other studies have used this approach to show an interaction between melodic and syntactic processing (Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009), harmonic and syntactic processing but not semantic processing (Hoch, Poulin-Charronnat, & Tillmann, 2011), and harmonic processing and word recall (Fiveash & Pammer, 2014). This has also been coupled to electrophysiological measures, confirming that melodic or harmonic unexpected events affect the syntax-related left anterior negativity (Carrus, Pearce, & Bhattacharya, 2013; Koelsch, Gunter, Wittfoth, & Sammler, 2005). Interestingly Sammler et al. (2013) showed a co-localization of early components elicited by musical and linguistic syntactic deviations using intracranial recordings.

Surprisingly, few neuroimaging studies have exploited the possibly most natural setting to compare music and language which is a stimulus that combines both speech and music: song. The use of songs has the clear advantage of solving the problem of using different

stimuli in the language and musical task. Schön et al. (2010) used an interference paradigm based on sung sentences and showed that the processing demands of melodic and lexical/phonological processing interact in a large network including bilateral temporal cortex and left inferior frontal cortex. Importantly, most voxels sensitive to the lexical/phonological manipulation are also sensitive to the interaction between the lexical/phonological and the melodic dimensions. In other words there seem to be very few voxels that are involved in lexical/phonological and are not influenced by melodic structure (see Fig. 1).



[Click to view larger](#)

*Figure 1.* Number of surviving voxels for the main effect of lexical/phonological dimension as a function of the threshold of the interaction between phonological and melodic dimensions. The dotted vertical line indicates the p-value of 0.05 for the mask. The right edge corresponds to a very conservative p-value (adapted from Schön et al., 2010).

Similarly, Sammler et al. (2010) using an adaptation paradigm, showed a strong integration between melodic and phonological levels in song in the dorsal pathway with a degree of integration decaying toward anterior regions of the left STS, possibly resulting from the processing of meaning of words. This integration of melodic and phonological dimension is also in line with the findings that a sung language is more easily learned than a spoken language (Schön et

al., 2008). Kunert and colleagues (Kunert, Willems, Casasanto, Patel, & Hagoort, 2015) showed an effect of musical harmonic deviancy on language syntax processing in the left inferior frontal gyrus. Notably this effect was not present when the deviancy in the musical stimulus was limited to the acoustic level (louder sound). Interestingly the authors also showed, in a behavioral study, an effect of the syntactic structure of sentences on the performance of a musical harmonic judgment task, confirming the idea of shared resources.

One may wonder how to combine these data suggesting shared resources with the “older” data issued from the neuropsychological studies pointing rather to a specificity and independence of several levels of language and music processing. However, very few studies have tried to systematically assess the co-existence of language and musical deficits, even for the most studied language deficit following a lesion in Broca’s area. Ani Patel was the first to investigate brain-damaged individuals and more specifically aphasic individuals with grammatical comprehension problems in language in order to see whether they also have a deficit in processing structural musical relations (Patel, Iversen, Wassenaar, & Hagoort, 2008). Broca’s aphasic patients and controls had to judge whether

a set of sentences contained or not a grammatical or semantic error. A similar task was used with harmonic error introduced into musical chord sequences. In a second experiment participants were tested using an implicit harmonic priming procedure. Both experiments showed that the aphasic patients have an impaired musical syntactic processing. Importantly, this took place in absence of low-level deficits, and with a preserved short-term memory for pitch patterns. This scenario is complicated by the fact that not all agrammatic patients may necessarily show a musical deficit (Slevc, Faroque-Shah, Saxena, & Okada, 2016). On a similar line, Sammler and colleagues (Sammler, Koelsch, & Friederici, 2011) showed a reduction or extinction of the typical electrophysiological marker of musical syntax processing in agrammatic patients with a lesion in the left inferior frontal cortex. These results are consistent with the hypothesis that Broca's area computes a rather domain-general "syntactic" processing but still a huge amount of work remains to be done with brain-lesioned patients.

## Music Training and Language Skills

We have seen that the approach of studying music and language brain correlates is limited by a number of methodological problems that render the interpretation of the results in terms of sharing or not of the resources rather complex. Another way to address the sharing resources hypothesis is to investigate whether music training affects the way the brain processes language, and vice versa. The reasoning is the following. Musical expertise requires an intense training often starting at an early age. As a result of learning, all the operations required by music perception and production will be affected by this training and become more efficient. If some of these operations are also required by language perception and production, then one should be able to observe a more efficient processing whenever the appropriate language processing levels are investigated. By contrast with the approach described above, the validation of this hypothesis does not necessarily require brain imaging data, insofar as behavioral differences can be taken as evidence that resource sharing exists. Psychologists and some neuroscientists often use the term "transfer of learning." This term is, however, rather vague as it seems to point to some sort of magic transfer of learning from one domain to another or from one function to another function without specifying how this transfer would actually take place. However, an alternative explanation is to hypothesize that these so-called transfer effects are simply due to an elementary function that is shared by both music and language processing. According to this view there is no transfer taking place, but only sharing of functions and resources. Importantly, while there is no clear way of showing how transfer could be possibly implemented, shared elementary operations can be defined via careful experimental manipulations.

Considering the early steps of sound analysis helps to clarify this point. The group of Nina Kraus has studied for many years the effect of music training on sound perception in general, including speech. Using EEG and focusing on high frequency (>200 Hz) neural

responses, possibly principally occurring at the subcortical level, this group of researchers has shown that, compared to non-musicians, musicians have a stronger representation of several features of speech sounds, including the fundamental frequency (Wong, Skoe, Russo, Dees, & Kraus, 2007), the harmonics (Kraus & Chandrasekaran, 2010), and rapid transients that may be important in distinguishing consonants (Parbery-Clark, Tierney, Strait, & Kraus, 2012). Overall, the correlation between the neural response and the stimulus is greater in musicians than in non-musicians and this independently of whether the stimulus is a music or a speech sound (Musacchia, Sams, Skoe, & Kraus, 2007). Most importantly this correlation is more resistant to acoustic noise in musicians. In other words, musicians seem to be able to filter out the noise better than non-musicians (Parbery-Clark, Skoe, & Kraus, 2009). Interestingly some of these differences can be observed in adults that had a few years of music training during childhood, thus showing that these changes last in time and do not necessarily require a long-lasting and intense training (Skoe & Kraus, 2012).

Moreover, these differences induced by music training are not simply due to a better processing of any sound feature. Indeed, results of a recent experiment show that music training can facilitate the selective processing of certain relevant features of speech. In this study, Intartaglia and colleagues (Intartaglia, White-Schwoch, Kraus, & Schön, 2017) compared French and American participants listening to an American phoneme, not existing in French. The comparison of the neural signatures showed that American participants had a more robust representation compared to French participants. The differences concerned the high formant frequencies that are necessary to encode the specific features of consonants and vowels. They then tested French musicians and the differences with the Americans disappeared. In other words, music training seems to allow a better encoding of the relevant features of speech sounds, even when these sounds are not familiar.

When interpreting these overall results one should keep in mind that two possible non-exclusive explanations co-exist. First, the subcortical relays may be more efficient in sound processing due to massive bottom-up processing. In this case one can clearly see that there is no need to advocate for a transfer effect. There is a dedicated auditory subcortical network that processes both musical and linguistic sounds. If this network becomes more efficient via intensive musical training, then speech processing will also benefit from the enhanced efficiency. Second, the cortical regions are known to send efferent signals to the subcortical relays and these modulatory top-down signals may play a role in enhancing the representation of certain features of sounds or in reducing the noise (Strait, Kraus, Parbery-Clark, & Ashley, 2010; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). In this perspective, the changes are possibly due to an enhanced connectivity that allows a finer modulatory activity of cortical over subcortical activity.

Independently of whether these enhanced subcortical representations reflect a bottom-up or a top-down modulation, these results are important in interpreting the differences that may be observed at a more integrated level. Indeed, differences observed at a phonological, syntactic, or prosodic level may result from a cascade effect of early

auditory processing differences. The studies on prosody and phoneme perception in musicians are particularly sensitive to this issue. Indeed, pitch is important in speech at the supra-segmental level, by signaling the emotional content of an utterance (Kotz et al., 2003), the linguistic structure (Steinhauer, Alter, & Friederici, 1999), and certain syntactic features such as to determine whether the utterance is a question or not (Astésano, Besson, & Alter, 2004). Pitch contour also plays a role at the segmental level in tone languages: it plays a linguistically contrastive function.

Musicians are more accurate in detecting subtle pitch variations in both music and speech prosody. These variations in speech prosody are detected earlier by musicians' brains and elicit more distinguishable event-related potentials compared to normal speech (Schön et al., 2004). This has been replicated with 8-year-old musician children (Magne, Schön, & Besson, 2006). Music lessons also seem to promote sensitivity to emotions conveyed by speech prosody. Indeed, musically trained adults perform better than untrained adults in discrimination and identification of emotional prosody (Thompson, Schellenberg, & Husain, 2004). Finally, musicians are more accurate at identifying, reproducing, or discriminating Mandarin tones (Gottfried & Riester, 2000; Gottfried, Staby, & Ziemer, 2004; Marie, Delogu, Lampis, Belardinelli, & Besson, 2011). However, as previously stated, it is difficult to know to what extent these differences are due to cortical or subcortical plasticity. Considering that anatomical differences have been observed at the cortical level in the auditory cortex (Benner et al., 2017; Kleber et al., 2016; Schlaug, Jäncke, Huang, Staiger, & Steinmetz, 1995; Shahin, Bosnyak, Trainor, & Roberts, 2003), it seems reasonable to believe that the whole auditory network is modified by music training, thus affecting speech processing at multiple levels.

Interestingly, previous studies provided evidence for a positive relationship between the function or the anatomy of the planum temporale and performance during syllable categorization (Elmer, Hänggi, Meyer, & Jäncke, 2013). Recently, Elmer and colleagues (Elmer, Hänggi, & Jäncke, 2016) provided evidence for a relationship between planum temporale connectivity, musicianship, and phonetic categorization. They found an increased connectivity between the left and right plana temporalia in musicians compared to non-musicians. This increased connectivity positively correlated with the performance in a phonetic categorization task as well as with musical aptitudes. Indeed, music training seems to affect the sensitivity to some acoustic features that are important to categorization of syllables, in particular temporal features such as voice-onset time (Chobert, Marie, François, Schön, & Besson, 2011; Zuk et al., 2013).

Very few studies have examined whether musical expertise influences the processing of the speech temporal structures. While isochrony is absent in speech, several nested temporal hierarchies are present in speech (Cummins & Port, 1998; Ghitza, 2011; Giraud & Poeppel, 2012). Musicians outperform non-musicians when asked to judge the lengthening of a syllable in a sentence (Marie, Magne, & Besson, 2011). Also, independently of whether musicians direct attention to the temporal or semantic content, they are more sensitive to subtle changes in the temporal structure of speech than non-musicians (Magne et al., 2006; Marie, Delogu, et al., 2011). Milovanov et al. (2009)

reported a positive correlation between musical aptitudes and sensitivity to syllable discrimination in children. In artificial language learning, speech segmentation results from the capacity to parse a continuous stream of syllables and to build and maintain probabilistic relationship of the different elements (syllables) that compose words. François and Schön (2011) showed that musicians have improved segmentation skills compared to non-musicians. Indeed, when listening to a new stream of an artificial language, they are faster and more accurate at segmenting the continuous stream. Children, after only one year of music training already show an improvement in speech segmentation (François, Chobert, Besson, & Schön, 2012). This ability, namely discovering word boundaries in the continuous stream of natural speech, is of utmost importance during language learning in the first years of life (Saffran et al., 1996).

The evidence concerning an effect of music training on language semantic and syntactic levels is rather scarce. One study showed that music training seems to influence semantic aspects of language processing (Dittinger et al., 2016). However, in this study, French participants had to learn new words that were in Thai language. Thus, differences may be due to the difficulty of the task at the perceptual level in terms of discriminating Thai tokens that differed in terms of pitch or vowel length. At the neural level, results indicate an increased functional connectivity in the ventral and dorsal streams of the left hemisphere during retrieval of novel words in musicians compared to non-musicians (Dittinger, Valizadeh, Jäncke, Besson, & Elmer, 2018). An effect of musical expertise on syntactic processing was shown by Jentschke and Koelsch (2009) with earlier and larger evoked responses to syntactic errors in children with musical training. However, others described that differences are absent at the behavioral level and that musical expertise does not modulate the amplitude of responses evoked by syntactic violations but only the topographical distribution (Fitzroy & Sanders, 2013). Thus, the evidence that music training affects language semantic and syntactic processing is not yet compelling and further studies are awaited.

Overall, while the theoretical framework of transfer of learning remains uncertain, there is a rather massive amount of data pointing to an improvement induced by music training at different levels of speech and language processing. Patel (2014) has tried to formalize the conditions under which music training may be beneficial to speech processing. In the OPERA hypothesis (Overlap, Precision, Emotion, Repetition, and Attention) he suggests that, in order for music training to enhance speech processing, music and speech need to share sensory or cognitive processing mechanisms and music must place higher demands on these mechanisms compared to speech. These mechanisms are tightly bound to the music emotional rewards system (Salimpoor et al., 2013). The last ingredients of music-induced and speech-related neural plasticity would be the fact that music training requires a repetition of sound patterns and gestures for an enormous amount of time under conditions of highly focused attention.

# Bridging Music and Language

When considering the effects of music training on speech and language abilities, one should keep in mind that most of the studies described here compared adult professional musicians to a group of adult non-musicians. This comparison has two methodological weaknesses. The first concerns the possibility of pre-existing differences, namely musicians already differed from non-musicians before starting to make music. The second is that music training is a complex activity, often involving individual lessons, group activities, theory classes, and so on. This makes it impossible to know what factors in music training had an impact on speech and language abilities.

Both criticisms can be addressed by running longitudinal studies assessing the absence of differences before the beginning of music training (Chobert, François, Velay, & Besson, 2012; François et al., 2012), and comparing the music-training group with a control group involved in an activity with a similar setting (e.g., visual arts, theater). However, this approach is time and cost consuming, insofar as it requires following two groups of children for a long period of time (often one year), testing them at least twice and coordinating the two training programs.

There is an alternative methodological approach that is somewhat in between the interference or interaction approach and the group comparison described earlier. The idea is to test the effect of music stimulation on speech perception. This has proven particularly successful in the temporal domain. Indeed, the structure of speech and music have a similar hierarchical temporal scaffolding (Haegens & Golumbic, 2018; Schön & Tillmann, 2015). A series of studies has shown that priming the temporal structure of speech using a music rhythmic prime can induce a speech processing benefit (Cason, Astésano, & Schön, 2015; Cason & Schön, 2012; Chern, Tillmann, Vaughan, & Gordon, 2018; Przybylski et al., 2013). These studies showed a benefit of rhythmic priming both in phoneme detection and in a grammaticality judgment task. This approach has been particularly efficient with language-impaired population. For instance, passive listening to a rhythmic regular prime improved the performance in a grammaticality judgment task in children with dyslexia or specific language impairment (SLI, Bedoin, Brisseau, Molinier, Roch, & Tillmann, 2016; Przybylski et al., 2013) and patients with a basal ganglia lesion (Kotz, Gunter, & Wonneberger, 2005). While these results support the importance of temporal predictions, the ability to anticipate in time upcoming events, for language processing, it is not clear whether the benefit at the grammatical level is mediated by a selective effect at the syntactic level or by improved speech perception. For instance Cason and colleagues (Cason, Hidalgo, Isoard, Roman, & Schön, 2015) have shown that priming the temporal structure of a sentence with music improved phoneme perception in hearing-impaired children.

Most of these studies have used a passive listening approach. However, an active approach, requiring the intervention of the audio-motor network seems to have a stronger effect than passive listening (Cason, Astésano, & Schön, 2015; Morillon & Baillet, 2017; Morillon, Schroeder, & Wyart, 2014). An interesting avenue for the future is to test the effect of a single session of music training on several levels of speech and language processing. This seems to us a good compromise between all the above-mentioned approaches insofar as it prevents the criticisms of pre-existing differences, and it allows a strict control of the content of the training session without “reducing” music to a passive listening of an isochronous metronome. Recently, Hidalgo and colleagues (Hidalgo, Falk, & Schön, 2017) used this type of approach to investigate temporal adaptation in speech interaction in hearing-impaired children. They showed that a 30 minute session of active rhythmic training facilitated the access to the temporal structure of verbal interactions and improved performance in a simple turn-taking task.

One of the factors prompting research in the domain of music and language is the possibility to use music to remediate language impairment. Thus, the fundamental research supports the therapeutic approach of using music to recover impaired functions by defining what aspects of music training benefit language processing and at which levels of processing. While it is not the aim of this chapter to review this literature (see chapter by McCain et al., this volume), it is important to note that the underlying neuroscientific models supporting the use of music in language rehabilitation have changed. For instance, the development of melodic intonation therapy to recover language function in non-fluent aphasic patients was somewhat driven by the idea that patients can learn a new way to speak through singing by using the right hemisphere (Albert, Sparks, & Helm, 1973; Zumbansen, Peretz, & Hébert, 2014). Forty years later our knowledge of the spatiotemporal dynamics subtending music and language on one side and of the pathophysiology of language disorders on the other side has been refined. Stahl and colleagues (Stahl, Kotz, Henseler, Turner, & Geyer, 2011) have shown, concerning non-fluent aphasia, that rhythmic training may be the most relevant aspect of the musical intervention, rather than the melodic aspect, especially when patients present a basal ganglia lesion, a subcortical structure involved in motor coordination and the processing of temporal information (Kotz & Schwartz, 2010).

Interestingly, several recent works on the use of music for language rehabilitation point to an important role of the rhythmic aspect of music. More precisely, musical training targeted toward improving rhythmic perception and production, resulted in improved phonological and reading skills (Bhide, Power, & Goswami, 2013; Cogo-Moreira, de Avila, Ploubidis, & de Jesus Mari, 2013; Flaunacco et al., 2015; Moore, Branigan, & Overy, 2017; Overy, 2000). These results suggest a shared substrate and point to temporal processing as playing a major role in language processing. This fits with the temporal sampling framework proposed by Goswami (2011) for dyslexia and by extension for SLI. Building on the neural resonance theory positing internal oscillators guiding attention over time (Large & Jones, 1999), Goswami suggests that deficits in syllabic segmentation and other sequential processes may result from impaired rhythmic entrainment leading to difficulties in sampling information over time. Along a similar line, Tierney and Kraus



## Music and Language

---

(2014) proposed the precise auditory timing hypothesis (PATH) that suggests that neural entrainment in auditory and motor cortex, and the interaction between them, underlies many of the behavioral aspects of both language and music processing. We will now describe music and language with a temporal focus.

## A Temporal Focus into Music and Language

As of today, the most promising approach to understand information processing seems to us the adoption of a dynamical emphasis, focusing on the temporal dimension. This perspective can be operationalized in complementary ways, ranging from portraying temporal regularities within sensory inputs to investigating time-resolved neural patterns of activity implicated in sensory processing, both in terms of frequency-resolved neural oscillations and neural networks dynamics. The underlying motivation is to describe information processing at the algorithmic (or representational) level, as first proposed by David Marr (Marr, 1982; Poeppel, 2012)—in other words to understand how the system does what it does, and more precisely what representations it uses, how they emerge, and how they are manipulated. Describing the time constants or the temporal profile of activity of each of these neural algorithms constitutes a preliminary stage toward this ultimate goal. While this approach can be carried out separately for music and language, a direct comparison of the two is also useful to delimitate general processing steps from more specific ones.

In the speech domain, David Poeppel has theorized this approach in the “asymmetric sampling in time” hypothesis (Giraud and Poeppel, 2012; Poeppel, 2003). Basically, speech can be described as a multi-timescale signal, with a hierarchical organization composed of phonemic, syllabic, and prosodic information (among others). At the neural level, both parallel and sequential processing occurs, with gamma ( $\sim 30$  Hz), theta ( $\sim 5$  Hz), and delta ( $\sim 2$  Hz) oscillations being specifically engaged by these multi-timescale, quasi-rhythmic properties of speech, and tracking its dynamics. Giraud and Poeppel argue that such neural oscillations “are foundational in speech and language processing, ‘packaging’ incoming information into units of the appropriate temporal granularity” (Giraud & Poeppel, 2012, p. 511). Interestingly, music is also characterized by a multi-timescale structure, with rhythm and meter hierarchically organized (Vuust & Witek, 2014). However, in an acoustic characterization of the temporal modulations in music and speech, Ding and colleagues (2017) recently highlighted that their temporal modulation rates differ. While the main tempo of music is around 2 Hz (120 bpm), a temporal modulation around 5 Hz primarily characterizes speech, which corresponds to the syllabic rate. At least two complementary avenues can be drawn from this result.

First, the distinction between music and speech modulation properties could be at the origin of some of their computational differences. In a fascinating paradigm, Oded Ghitza showed that intelligibility of time-compressed speech can be greatly enhanced if periods of silence of the appropriate duration are inserted (Ghitza, 2011; Ghitza & Greenberg, 2009). Oscillation-based models of speech perception best explain these data, where optimum intelligibility is achieved when the syllable rhythm is within the range of the theta-frequency brain rhythms ( $\sim 4$ – $10$  Hz), comparable to the rate at which segments and syllables are articulated in conversational speech. Follow-up experiments were performed in the music domain, where participants had to identify the musical key of time-

compressed short melodic sequences (Farbood, Marcus, & Poeppel, 2013; Farbood, Rowland, Marcus, Ghitza, & Poeppel, 2015). This highlighted that insertion of silence gaps was beneficial to performance, in unison with the speech experiments, providing compelling clues into possible oscillatory mechanisms underlying segmentation of auditory information. However, the two experiments in the music domain were not conclusive with regard to the preferred rate of processing, observed at 2–3 Hz or 5–7 Hz, respectively. While the former result would be compatible with the fact that the main tempo of music is around 2 Hz, suggesting that the distinctions between music and speech acoustic modulation properties are a productive attribute of their respective perceptual analysis, the latter would be compatible with the idea that the auditory cortex parses information at the theta rate, and that such sampling operates rather independently of the nature of the acoustic signal (music or speech).

Second, the most shared characteristic between music and language acoustic signals is that both of them have strong temporal constraints (i.e., a salient main modulation rate, at ~2 and 5 Hz, respectively), leading to strong temporal predictions. Temporal predictions are believed to play a fundamental role in the way we sample sensory information, in particular in the auditory domain (Jones, 1976; Nobre & van Ede, 2018; Schroeder & Lakatos, 2009). Behavioral experiments show that anticipating the moment of occurrence of an upcoming event optimizes its processing by improving the quality of sensory information (Jaramillo & Zador, 2011; Morillon, Schroeder, Wyart, & Arnal, 2016; Rohenkohl, Cravo, Wyart, & Nobre, 2012). Current theories and empirical findings suggest that this enhancement is achieved by the entrainment of low-frequency neuronal oscillations, which temporally modulates the excitability of task-relevant neuronal populations (Cravo, Rohenkohl, Wyart, & Nobre, 2013; Large & Jones, 1999; Schroeder & Lakatos, 2009). Such entrainment, principally observed in sensory cortices (Besle et al., 2011; Lakatos et al., 2013), would be possible thanks to the downward propagation of temporal prediction signals, recently shown to originate in the motor system (Morillon & Baillet, 2017). These signals would be responsible for the predictive alignment of the neuronal excitability phase of ongoing oscillations in sensory cortex with upcoming events, possibly through top-down phase-reset (e.g., Park, Ince, Schyns, Thut, & Gross, 2015; Stefanics et al., 2010).

A recent proposition by Arnal and colleagues (Rimmele, Morillon, Poeppel, & Arnal, submitted) is that time estimation relies on the neural recycling of action circuits (Coull, 2011) and is implemented by internal, non-conscious “simulation” of movements in most ecological situations (Arnal, 2012; Arnal & Giraud, 2012; Schubotz, 2007). On this view, temporal predictions correspond to a covert form of active sensing (Morillon, Hackett, Kajikawa, & Schroeder, 2015; Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010). In other words, the efferent motor signals that are generated when synchronizing our actions to predictable events are also generated during the passive perception of such regularities (Arnal, 2012; Patel & Iversen, 2014). When temporal regularities occur in the timescale of natural actions/movements, the motor system is recruited (Chen, Penhune, & Zatorre, 2008; Du & Zatorre, 2017; Grahn & Rowe, 2012; Merchant, Grahn, Trainor, Rohrmeier, & Fitch, 2015; Teki, Grube, Kumar, & Griffiths, 2011; Zatorre, Chen, &

Penhune, 2007). The great richness of the repertoire of motor schemes (gestures) makes it possible to simulate (and predict) the occurrence of sensory events with great accuracy and to treat them with greater precision (Morillon et al., 2016; Schubotz, 2007), offering a flexible tool to precisely predict “when” and select relevant information in time. Given the finesse of our motor expertise and the amazing complexity of our repertoire of actions, this means that we can use internal simulation of action to anticipate temporal trajectories. This conception is compatible with various forms of “motor theories” of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), in which the covert simulation of actions can lead to a given sensory configuration.

The role of temporal predictions, while being a critical role in both music and language, differs in multiple ways. First, music is much more rhythmic than speech, hence predictions are more precise. Second, while temporal predictions have primarily a contextual role in language, helping to optimize the extraction of relevant information, they serve a much more fundamental purpose in music. Indeed, musical rhythm has a remarkable capacity to move our minds and bodies. This is because it is part of the information content in itself, rather than being a contextual cue (as in language). In a compelling review article, Vuust and Witek (2014) hypothesize that music would exploit general principles of brain functioning, notably its structuration as a Bayesian, predictive system, to optimize our pleasure and desire to move. In any case, these distinctions highlight that music stimulates the dorsal auditory stream much more than language, as this pathway is involved in audio-motor transformation (Hickok & Poeppel, 2007) and temporal information processing (Morillon & Baillet, 2017). As a consequence, musical training or musical stimulation strengthen the connectivity between auditory and motor cortices, which has beneficial effects for speech comprehension (Falk, Lanzilotti, & Schön, 2017), especially in noisy conditions (Du & Zatorre, 2017), and phonological and reading skills in children (Flaugnacco et al., 2015), as described earlier. Overall, while music and language have both different structure and function, they share the specificity to be temporal in essence. Adopting a dynamical approach seems thus the most promising avenue to understand how the human brain interacts with this type of multisensory environment.

## References

- Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., & Menon, V. (2010). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex* 21(7), 1507–1518.
- Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Archives of Neurology* 29, 130–131.
- Arnal, L. (2012). Predicting “when” using the motor system’s beta-band oscillations. *Frontiers in Human Neuroscience* 6, 225. Retrieved from <https://doi.org/10.3389/fnhum.2012.00225>

Arnal, L., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences* 16(7), 390–398.

Astésano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research* 18(2), 172–184.

Basso, A., & Capitani, E. (1985). Spared musical abilities in a conductor with global aphasia and ideomotor apraxia. *Journal of Neurology, Neurosurgery & Psychiatry* 48(5), 407–412.

Bedoin, N., Brisseau, L., Molinier, P., Roch, D., & Tillmann, B. (2016). Temporally regular musical primes facilitate subsequent syntax processing in children with specific language impairment. *Frontiers in Neuroscience* 10. Retrieved from <https://doi.org/10.3389/fnins.2016.00245>

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403(6767), 309–312.

Benner, J., Wengenroth, M., Reinhardt, J., Stippich, C., Schneider, P., & Blatow, M. (2017). Prevalence and function of Heschl's gyrus morphotypes in musicians. *Brain Structure and Function* 222(8), 1–17.

Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., ... Schroeder, C. E. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *Journal of Neuroscience* 31(9), 3176–3185.

Bhide, A., Power, A., & Goswami, U. (2013). A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind, Brain, and Education* 7(2), 113–123.

Carrus, E., Pearce, M. T., & Bhattacharya, J. (2013). Melodic pitch expectation interacts with neural responses to syntactic but not semantic violations. *Cortex* 49(8), 2186–2200.

Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: Rhythmic priming and audio-motor training affect speech perception. *Acta Psychologica* 155, 43–50.

Cason, N., Hidalgo, C., Isoard, F., Roman, S., & Schön, D. (2015). Rhythmic priming enhances speech production abilities: Evidence from prelingually deaf children. *Neuropsychology* 29(1), 102.

Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia* 50(11), 2652–2658.

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex* 18(12), 2844–2854.

- Chern, A., Tillmann, B., Vaughan, C., & Gordon, R. L. (2018). New evidence of a rhythmic priming effect that enhances grammaticality judgments in children. *Journal of Experimental Child Psychology* 173, 371–379.
- Cheung, C., Hamilton, L. S., Johnson, K., & Chang, E. F. (2016). The auditory representation of speech sounds in human motor cortex. *eLife* 5, e12577.
- Chobert, J., François, C., Velay, J. L., & Besson, M. (2012). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex* 24(4), 956–967.
- Chobert, J., Marie, C., François, C., Schön, D., & Besson, M. (2011). Enhanced passive and active processing of syllables in musician children. *Journal of Cognitive Neuroscience* 23(12), 3874–3887.
- Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior*. *Language* 35(1), 26–58.
- Cogo-Moreira, H., de Avila, C. R. B., Ploubidis, G. B., & de Jesus Mari, J. (2013). Effectiveness of music education for the improvement of reading skills and academic achievement in young poor readers: A pragmatic cluster-randomized, controlled clinical trial. *PLoS ONE* 8(3), e59984.
- Coull, J. T. (2011). Discrete neuroanatomical substrates for generating and updating temporal expectations. In S. Dehaene & E. Brannon (Eds.), *Space, time and number in the brain: Searching for the foundations of mathematical thought* (pp. 87–101). Amsterdam: Elsevier.
- Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *Journal of Neuroscience* 33(9), 4002–4010.
- Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26(2), 145–171.
- DeWitt, L. A., & Samuel, A. G. (1990). The role of knowledge-based expectations in music perception: Evidence from musical restoration. *Journal of Experimental Psychology: General* 119(2), 123–144.
- Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews* 81(B), 181–187.
- Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional music training and novel word learning: From faster semantic encoding to longer-lasting word representations. *Journal of Cognitive Neuroscience* 28(10), 1584–1602.

Ditinger, E., Valizadeh, S. A., Jäncke, L., Besson, M., & Elmer, S. (2018). Increased functional connectivity in the ventral and dorsal streams during retrieval of novel words in professional musicians. *Human Brain Mapping* 39(2), 722–734.

Du, Y., & Zatorre, R. J. (2017). Musical training sharpens and bonds ears and tongue to hear speech better. *Proceedings of the National Academy of Sciences* 5, 201712223. Retrieved from <https://doi.org/10.1073/pnas.1712223114>

Elhilali, M., & Shamma, S. A. (2008). A cocktail party with a cortical twist: How cortical mechanisms contribute to sound segregation. *Journal of the Acoustical Society of America* 124(6), 3751–3771.

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(1), 331–344.

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(10), 2812–2821.

Falk, S., Lanzilotti, C., & Schön, D. (2017). Tuning neural phase entrainment to speech. *Journal of Cognitive Neuroscience* 29(8), 1378–1389.

Farbood, M. M., Marcus, G., & Poeppel, D. (2013). Temporal dynamics and the identification of musical key. *Journal of Experimental Psychology Human Perception & Performance* 39(4), 911–918.

Farbood, M. M., Rowland, J., Marcus, G., Ghitza, O., & Poeppel, D. (2015). Decoding time for the identification of musical key. *Attention, Perception, & Psychophysics* 77(1), 28–35.

Fedorenko, E., McDermott, J. H., Norman-Haignere, S., & Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *Journal of Neurophysiology* 108(12), 3289–3300.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition* 37(1), 1–9.

Fitzroy, A. B., & Sanders, L. D. (2013). Musical expertise modulates early processing of syntactic violations in language. *Frontiers in Psychology* 3, 603. Retrieved from <https://doi.org/10.3389/fpsyg.2012.00603>

Fiveash, A., & Pammer, K. (2014). Music and language: Do they draw on similar syntactic working memory resources? *Psychology of Music* 42(2), 190–209.

- Flaunghacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS ONE* 10(9), e0138715.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
- François, C., Chobert, J., Besson, M., & Schön, D. (2012). Music training for the development of speech segmentation. *Cerebral Cortex* 23(9), 2038–2043.
- François, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex* 21(10), 2357–2365.
- Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage* 20, S8–S17.
- Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology* 2. Retrieved from <https://doi.org/10.3389/fpsyg.2011.00130>
- Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66, 113–126.
- Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56(6), 1127–1134.
- Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience* 15(4), 511–517.
- Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences* 15(1), 3–10.
- Gottfried, T. L., & Riester, D. (2000). Relation of pitch glide perception and Mandarin tone identification. *Journal of the Acoustical Society of America* 108(5), 2604.
- Gottfried, T. L., Staby, A. M., & Ziemer, C. J. (2004). Musical experience and Mandarin tone discrimination and imitation. *Journal of the Acoustical Society of America* 115(5), 2545.
- Grahn, J. A., & Rowe, J. B. (2012). Finding and feeling the musical beat: Striatal dissociations between detection and prediction of regularity. *Cerebral Cortex* 23(4), 913–921.
- Haegens, S., & Golumbic, E. Z. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews* 86, 150–165.



- Herdener, M., Humbel, T., Esposito, F., Habermeyer, B., Cattapan-Ludewig, K., & Seifritz, E. (2012). Jazz drummers recruit language-specific areas for the processing of rhythmic structure. *Cerebral Cortex* 24(3), 836–843.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience* 8, 393–402.
- Hidalgo, C., Falk, S., & Schön, D. (2017). Speak on time! Effects of a musical rhythmic training on children with hearing loss. *Hearing Research* 351, 11–18.
- Hillis, A. E., & Caramazza, A. (1995). Representation of grammatical categories of words in the brain. *Journal of Cognitive Neuroscience* 7(3), 396–407.
- Hoch, L., Poulin-Charronnat, B., & Tillmann, B. (2011). The influence of task-irrelevant music on language processing: Syntactic and semantic structures. *Frontiers in Psychology* 2. Retrieved from <https://doi.org/10.3389/fpsyg.2011.00112>
- Intartaglia, B., White-Schwoch, T., Kraus, N., & Schön, D. (2017). Music training enhances the automatic neural processing of foreign speech sounds. *Scientific Reports* 7(1), 12631.
- Jaramillo, S., & Zador, A. M. (2011). The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nature Neuroscience* 14, 246–251.
- Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *NeuroImage* 47(2), 735–744.
- Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review* 83(5), 323–355.
- Kleber, B., Veit, R., Moll, C. V., Gaser, C., Birbaumer, N., & Lotze, M. (2016). Voxel-based morphometry in opera singers: Increased gray-matter volume in right somatosensory and auditory cortices. *NeuroImage* 133, 477–483.
- Koelsch, S., Gunter, T. C., Cramon, D. Y. V., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical “language-network” serves the processing of music. *NeuroImage* 17(2), 956–966.
- Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience* 17(10), 1565–1577.
- Kotz, S. A., Gunter, T. C., & Wonneberger, S. (2005). The basal ganglia are receptive to rhythmic compensation during auditory syntactic processing: ERP patient data. *Brain and Language* 95(1), 70–71.

- Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain & Language* 86(3), 366–376.
- Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences* 14(9), 392–399.
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience* 11(8), 599–605.
- Kunert, R., & Slevc, L. R. (2015). A commentary on: “Neural overlap in processing music and speech.” *Frontiers in Human Neuroscience* 9. Retrieved from <https://doi.org/10.3389/fnhum.2015.00330>
- Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca’s area: An fMRI study. *PloS One* 10(11), e0141069.
- Lakatos, P., Musacchia, G., O’Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). the spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77, 750–761.
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review* 106(1), 119–159.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review* 74, 431–461.
- Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex* 15(10), 1621–1631.
- Lindell, A. K. (2006). In your right mind: Right hemisphere contributions to language processing and production. *Neuropsychology Review* 16(3), 131–148.
- Luria, A. R., Tsevetkova, L. S., & Futer, D. S. (1965). Aphasia in a composer. *Journal of the Neurological Sciences* 2(3), 288–292.
- Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca’s area: an MEG study. *Nature neuroscience*, 4(5), 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. *Journal of Cognitive Neuroscience* 18(2), 199–211.
- Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience* 23(10), 2701–2715.

- Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience* 23(2), 294–305.
- Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.
- Merchant, H., Grahm, J., Trainor, L., Rohrmeier, M., & Fitch, W.T. (2015). Finding the beat: A neural perspective across humans and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370(1664), 20140093. doi:10.1098/rstb.2014.0093
- Milovanov, R., Huotilainen, M., Esquef, P. A., Alku, P., Välimäki, V., & Tervaniemi, M. (2009). The role of musical aptitude and language skills in preattentive duration processing in school-aged children. *Neuroscience Letters* 460(2), 161–165.
- Moore, E., Branigan, H. & Overy, K. (2017). Exploring the role of auditory-motor synchronisation in the transfer of music to language skills in dyslexia. Outstanding Poster Award talk at Neurosciences and Music VI conference.
- Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proceedings of the National Academy of Sciences* 114(42), E8913–E8921.
- Morillon, B., Hackett, T. A., Kajikawa, Y., & Schroeder, C. E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. *Current Opinion in Neurobiology* 31, 230–238.
- Morillon, B., Schroeder, C. E., & Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nature Communications* 5, 5255.
- Morillon, B., Schroeder, C. E., Wyart, V., & Arnal, L. H. (2016). Temporal prediction in lieu of periodic stimulation. *Journal of Neuroscience* 36(8), 2342–2347.
- Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences* 104(40), 15894–15898.
- Nobre, A. C., & van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience* 19, 34–38.
- Overy, K. (2000). Dyslexia, temporal processing and music: The potential of music as an early learning aid for dyslexic children. *Psychology of Music* 28(2), 218–229.
- Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience* 29(45), 14100–14107
- Parbery-Clark, A., Tierney, A., Strait, D. L., & Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience* 219, 111–119.

- Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology* 25(12), 1649–1653.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience* 6(7), 674–681.
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology* 2, 142. doi:10.3389/fpsyg.2011.00142
- Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research* 308, 98–108.
- Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in System Neuroscience* 8, 57. Retrieved from <https://doi.org/10.3389/fnsys.2014.00057>
- Patel, A. D., Iversen, J. R., Wassenaar, M., & Hagoort, P. (2008). Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology* 22(7–8), 776–789.
- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience* 6(7), 688–691.
- Peretz, I., Vuvan, D., Lagrois, M. É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370(1664), 20140090.
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as “asymmetric sampling in time.” *Speech Communication* 41(1), 245–255.
- Poeppel, D. (2012). The maps problem and the mapping problem: Two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology* 29(1–2), 34–55.
- Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Léculier, L., ... Tillmann, B. (2013). Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology* 27(1), 121–131.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience* 12, 718–724.
- Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences* 97(22), 11800–11806.

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (submitted). The proactive and flexible sense of timing.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience* 27, 169–192.

Rogalsky, C., & Hickok, G. (2011). The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience* 23(7), 1664–1680.

Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. (2012). Temporal expectation improves the quality of sensory information. *Journal of Neuroscience* 32(24), 8424–8428.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274(5294), 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70(1), 27–52.

Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., & Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* 340(6129), 216–219.

Sammler, D., Baird, A., Valabrègue, R., Clément, S., Dupont, S., Belin, P., & Samson, S. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: A functional magnetic resonance adaptation study. *Journal of Neuroscience* 30(10), 3572–3578.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H. J., ... Friederici, A. D. (2013). Co-localizing linguistic and musical syntax with intracranial EEG. *NeuroImage* 64, 134–146.

Sammler, D., Koelsch, S., & Friederici, A. D. (2011). Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? *Cortex* 47(6), 659–673.

Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia* 33(8), 1047–1055.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition* 106(2), 975–983.

Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J. L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage* 51(1), 450–461.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology* 41(3), 341–349.

- Schön, D., & Tillmann, B. (2015). Short- and long-term rhythmic interventions: Perspectives for language rehabilitation. *Annals of the New York Academy of Sciences* 1337, 32–39.
- Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences* 32(1), 9–18.
- Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology* 20, 172–176.
- Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences* 11(5), 211–218.
- Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience* 23(13), 5545–5552.
- Skoe, E., & Kraus, N. (2012). A little goes a long way: How the adult brain is shaped by musical training in childhood. *Journal of Neuroscience* 32(34), 11507–11510.
- Slevc, L. R., Farooqi-Shah, Y., Saxena, S., & Okada, B. M. (2016). Preserved processing of musical structure in a person with agrammatic aphasia. *Neurocase* 22(6), 505–511.
- Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review* 16(2), 374–381.
- Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology* 19(6), 498–502.
- Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: Why singing may not hold the key to recovery from aphasia. *Brain* 134(10), 3083–3093.
- Stefanics, G., Hangya, B., Hernadi, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *Journal of Neuroscience* 30(41), 13578–13585.
- Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience* 2(2), 191–196.
- Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research* 261(1), 22–29.

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *Journal of Neuroscience* 31(10), 3805–3812.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science* 331(6022), 1279–1285.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion* 4(1), 46–64.

Tierney, A., & Kraus, N. (2014). Auditory-motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience* 8. Retrieved from <https://doi.org/10.3389/fnhum.2014.00949>

Tinbergen, N. (1963). On aims and methods of ethology. *Ethology* 20, 410–433.

Vallentin, D., Kosche, G., Lipkind, D., & Long, M. A. (2016). Inhibition protects acquired song segments during vocal learning in zebra finches. *Science* 351(6270), 267–271.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Jobard, G., Petit, L., Crivello, F., ... Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage* 54(1), 577–593.

Vuust, P., & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology* 5, 1111. Retrieved from <https://doi.org/10.3389/fpsyg.2014.01111>

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience* 10(4), 420–422.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences* 6(1), 37–46.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience* 8, 547–558.

Zuk, J., Ozernov-Palchik, O., Kim, H., Lakshminarayanan, K., Gabrieli, J. D. E., Tallal, P., & Gaab, N. (2013). Enhanced syllable discrimination thresholds in musicians. *PLoS ONE* 8(12), e80546.

Zumbansen, A., Peretz, I., & Hébert, S. (2014). Melodic intonation therapy: Back to basics for future research. *Frontiers in Neurology* 5. Retrieved from <https://doi.org/10.3389/fneur.2014.00007>

**Daniele Schön**

Daniele Schön, Institut de Neurosciences des Systèmes, Aix-Marseille Université & INSERM, Marseille, France

**Benjamin Morillon**

Benjamin Morillon, Institut de Neurosciences des Systèmes, Aix-Marseille Université & INSERM, Marseille, France

