17

Music and Language in the Brain

Balancing Domain-Specific and Domain-General Mechanisms

Stephen C. Van Hedger and Urvi Maheshwari

17.1 Introduction

Music is the universal language of mankind

Henry Wadsworth Longfellow

Over the years, the relationship between music and language has been explored by philosophers, poets, musicians, evolutionary biologists, and, most pertinently for the present chapter, cognitive neuroscientists. A common theme emerging from these comparisons is that music and language share remarkable similarities, with music perhaps even representing a kind of universal language (a sentiment captured by the opening quotes to this chapter).

Indeed, the observable links between music and language are hard to ignore. Music and spoken language both make use of complex sound patterns that unfold over time. Both systems are shaped by culture (Cross, 2001), are human activities present in every known culture (McDermott & Hauser, 2005; Mehr et al., 2019), are learned over the course of development (McMullen & Saffran, 2004), and convey meaning to listeners (Slevc & Patel, 2011). Archaeological evidence of bone flutes dating back 50,000 years (Higham et al., 2012; Turk et al., 2020; though see Diedrich, 2015), suggests that music making was an integral part of early cultural development. The existence of music for tens of thousands of years has led some to theorize that music and language were part of the same early communicative system, making use of prosodic changes (i.e., patterns of intonation and stress) to convey affective state (Brown, 2017; Mithen, 2005). Although the evolutionary function of music has been the subject of considerable debate (e.g., Honing et al., 2015; Pinker, 1999), the perspective offered by Brown (2017) suggests that music and language share contemporary similarities because both stem from a common communicative system.

Yet, reductionist comparisons of music and language highlight some of the tenuous links that bridge both domains. For example, despite both music and spoken language consisting of structured sequences of sounds that convey meaning to listeners, the syntactic organization and means of conveying meaning differ considerably across domains. Differences between music and language are perhaps most salient in the domain of semantics. Whereas language can convey concrete, referential meaning through (mostly) arbitrary patterns of sound (e.g., "The tired professor grades the term papers"), music cannot convey a similarly specific referential meaning, even though it might be able to represent a subjective psychological state of the tired professor somewhat adeptly (e.g., via slow, quiet, repetitive sound patterns). This example represents a rather extreme difference in how music and language operate, but it highlights the need for researchers exploring music-language links to consider appropriate levels of analysis to make fruitful cross-domain comparisons (cf. Asano & Boeckx, 2015). This is critical to gaining traction on the question of how distinct or overlapping neural representations of these systems are. Moreover, understanding the relative distinctiveness of both systems has broader implications. For example, it provides insight into whether speech is a special cognitive ability, or whether it might be explained under a broader auditory framework that includes music (Peretz & Hyde, 2003; Pinker & Jackendoff, 2005).

In the present chapter, we compare music and language from a cognitive neuroscience lens. We begin by highlighting the apparent similarities in how these domains are organized (Section 17.2), in terms of acoustic (Section 17.2.1), syntactic (Section 17.2.2), semantic (Section 17.2.3) features. However, throughout this discussion, we highlight common pitfalls within this area of research, which involve attempts to apply the lens of one domain too myopically to the other domain. We then turn away from the specific features of both music and language to examine broader cognitive mechanisms involved in acquiring and maintaining these systems of knowledge (Section 17.3). We focus on the developmental constraints in learning music and language (Section 17.3.1), the role of implicit learning mechanisms in pattern extraction across domains (Section 17.3.2), and the conditions under which training in one domain generalizes to the other domain (Section 17.3.3). We conclude (Section 17.4) by summarizing the current state of the scientific literature and commenting on how the comparison of music and language lends insight into fundamental questions within cognitive neuroscience.

17.2 Featural Similarities between Music and Language

The featural similarities between music and language have become the subject of considerable research in psychology and cognitive neuroscience. Understanding the featural elements of both music and language is a necessary step in assessing the degree to which these domains have distinct or shared processing in the brain. In this section, we deconstruct these featural similarities across music and language, from a description of the sound patterns that comprise both systems (Section 17.2.1), to the ways in which sounds are organized hierarchically (Section 17.2.2), and how these organized sound patterns convey meaning to listeners (Section 17.2.3).

17.2.1 Acoustic Features of Music and Language

Both music and spoken language are acoustic signals that unfold over time. Music and speech signals interact with the same auditory organs and are not inherently differentiated at the level of the cochlear partition. Yet, music and speech are (generally) not confusable signals to most listeners, despite following the same sensory pathway and aggregating as a single waveform if heard simultaneously. How, then, do listeners clearly separate these signals to appreciate both as separable auditory objects?

Although a comprehensive description of the acoustic similarities and differences between music and speech is beyond the scope of this chapter, both spectral (related to the frequency components of sounds) and temporal (related to how sounds unfold over time) factors differentiate music from speech. Music typically contains discrete (stable) pitches and isochronous rhythmic patterns that allow for the extraction of a "beat" (e.g., Patel, 2003). By contrast, speech typically contains larger, more continuous pitch changes and rhythmic patterns that are not as strongly periodic (Peelle & Davis, 2012; Zatorre & Baum, 2012). Music and speech also differ in their reliance on fine-grained (spectral) information versus temporal (amplitude envelope) information (Smith et al., 2002). These findings can be contextualized by considering the demands of music and speech comprehension. In speech, temporal differences on the order of tens of milliseconds can entirely change the meaning of the signal – for example, in differentiating stop consonants like /ba/ and /pa/ (e.g., Miller & Volaitis, 1989). Music unfolds more slowly (Ding et al., 2017) and places greater demands on spectral resolution. For example, most listeners can readily detect when elements of a musical performance are "out of tune" relative to one another, even when these differences are under one semitone, corresponding to about a 5.95 percent change in frequency (Larrouy-Maestri et al., 2019). By contrast, pitch changes in spoken language are on the order of half an octave (i.e., 50 percent changes in pitch), with precise tuning or intervallic information not playing a critical role in understanding (Patel, Peretz, et al., 1998)

These acoustic differences between music and speech might partly underlie the observed differences in brain lateralization across domains. Although music and speech are not well differentiated early in the auditory pathway, including in the primary auditory cortex where tonotopic organization is largely preserved (e.g., Humphries et al., 2010; Zatorre et al., 1992), a consistent finding over decades of neuroimaging research has found that language processing tends to be left lateralized, whereas music processing tends to be right lateralized (Tervaniemi & Hugdahl, 2003). Given that the left hemisphere has been found to be more temporally sensitive (extracting information over short timescales), whereas the right hemisphere has been found to be more spectrally sensitive (extracting information over longer timescales), the lateralization of music and speech may in part be due to acoustic differences (Zatorre et al., 2002).

One approach trying to examine the extent to which observed neural differences between music and speech might be acoustically driven focuses on how *song*, relative to non-sung speech, is represented in the brain (e.g., Norman-Haignere et al., 2015). Although this approach reduces the spectro-temporal richness that characterizes music to a single instrument (the voice), it addresses the role of acoustic features in differentiating music and speech representations, as both signals are produced by the same source. To highlight the leverage gained by using song to assess acoustic differences in speech and music processing, Albouy and colleagues (2020) played participants a cappella songs (with lyrics) using a functional magnetic resonance imaging (fMRI) paradigm. When songs were degraded in the temporal domain, speech perception was impaired, yet music perception was preserved; and when songs were degraded in the spectral domain, music perception was impaired, and speech perception was preserved. Critically, the processing of these degraded sounds in terms of song or speech was highly lateralized: the neural decoding of speech was dependent on activity patterns in the left auditory cortex, and the neural decoding of music was dependent on activity patterns in the right auditory cortex. By degrading the same sounds to selectively impair speech and music perception, the researchers provide strong evidence that the hemispheric asymmetry found between speech and music can be explained by specialization of left and right auditory cortex for temporal and spectral analysis, respectively.

Given that both song and speech make use of the same vocal and articulatory mechanisms, this opens the possibility that these signals might not be clearly differentiable at an acoustic level. For example, the "speech-to-song" illusion is an effect where an acoustic signal, clearly heard as speech initially, transforms to song as a function of repetition (Deutsch et al., 2008). This illusion provides a promising window for distinguishing acoustic factors from music and speech representations in the brain, as the same acoustic stimulus can serve as both speech and music. A significant amount of research has been conducted to examine the acoustic factors that make some speech segments more likely to transform into song points to acoustic factors, such as the pitch salience and pitch stability of the utterance, as well as the extent to which repetition leads to perceptual grouping of sounds that accentuates metrical information, (e.g., stressed and unstressed beats; Falk et al., 2014). This work examining speech and song has found that speech – given its greater variability in terms of pitch changes and rhythmic variation – sometimes appears to incidentally contain sufficient pitch stability and sufficient metrical patterns to transform to song, blurring the distinction between music and speech at an acoustic level.

The speech-to-song illusion thus offers a valuable window into the neural processing of both music and spoken language, as it allows for the manipulation of domain (music versus speech) while controlling for acoustic factors. Using fMRI, Tierney and colleagues (2013) played participants excerpts from an audiobook. Some of these excerpts transformed into songs when repeated (i.e., the speech-to-song illusion), whereas others did not (control stimuli). This design allowed the researchers to examine the changes in blood-oxygenation-level-dependent (BOLD) responses in the brain as a function of hearing acoustically similar stimuli as either speech or song. The authors found several additional areas that increased in BOLD activation in response to hearing song (versus speech), including areas implicated in pitch processing (e.g., anterior and posterior superior temporal gyrus), and areas implicated in audio-motor integration (e.g., supramarginal gyrus and inferior frontal gyrus). The authors conclude that these differences in "song" versus "speech" stimuli are unlikely to be the result of acoustic factors, but rather reflect the increased demands that song perception places on pitch processing and motor processing. Interestingly, the findings from this study were entirely asymmetrical – perception of song led to increased BOLD responses in multiple pitch- and motor-sensitive areas, but the control stimuli that did not transform did not lead to increased BOLD responses in any area. These findings suggest that perceiving a stimulus as music might place greater processing demands in a myriad of brain areas, which has been proposed as one of the ways in which musical training might lead to benefits in speech processing (e.g., see Patel, 2011 and also Section 17.3.3).

Music and spoken language are quite similar at an acoustic level. Both contain complex harmonic information, a perceptible fundamental frequency, and periodic and quasi-periodic rhythmic patterns as they unfold over time. Yet, musical sounds tend to have more stable pitches and more regular rhythms that allow for the perception of a metrical beat. In earlier parts of the auditory pathway (including the earliest cortical processing of sounds in the primary auditory cortex), music and speech are not clearly distinguishable. However, methodological advances and clever experimental paradigms (e.g., using the speech-to-song illusion) have demonstrated differentiation of music and speech signals in areas implicated in pitch and motor processing. The existence of speech selective and music selective areas of the brain, even when controlling for acoustic factors, suggests that these systems have domain-specific representations, even if they share substantial overlap in earlier auditory processing areas.

17.2.2 Syntactic Features of Music and Language

The acoustic signatures of both music and language are not static, but rather unfold over time. An important consideration when comparing how music and language are represented in the brain is to examine the ways in which these sounds are structured. Both music and language are governed by rules that constrain how smaller units (e.g., notes, words) are combined into larger structural units or sequences (e.g., phrases, sentences). Both music and language have *syntax*, broadly defined as the set of rules that dictate how discrete elements of a system are combined hierarchically into larger units or sequences (e.g., Adger, 2015).

Even though both language and music are governed by syntactic rules, a pitfall in this area of research is to examine these relationships using a myopic lens (e.g., see Asano & Boeckx, 2015 for a similar argument). Although music and language have a hierarchical structure, trying to identify surface-level parallels across domains is problematic for several reasons (Lerdahl, 2013). First, hierarchical organization can occur at multiple levels of analysis within each domain (e.g., the phonological, morphological, and sentential level in language; the harmonic and rhythmic levels in music), and it is unclear how the "units" of each domain should be compared. Second, syntax is intricately tied to meaning, particularly within music (Asano & Boeckx, 2015). For example, *harmonic syntax* (the rules governing how tonal sequences unfold over time) generates strong expectations in listeners, and violating these expectations is one of the ways in which music can represent affective meaning (e.g., Steinbeis et al., 2006). Similarly, in language, syntax is used to determine and convey meaning, perhaps most saliently in a process called syntactic bootstrapping. On these accounts, language learners might acquire semantic knowledge (e.g., verb meanings) from syntactic knowledge, such as using grammatical cues (e.g., tense markings, morphology) and discourse structure (Brown, 1957; Gleitman, 1990; Jakobson, 1965; Landau & Gleitman, 2009; Naigles, 1990; Naigles & Swensen, 2007) to narrow their hypothesis space about the meanings of words. Given the radically different ways in which music and language convey meaning (see Section 17.2.3), this intertwining of syntax and semantics complicates direct and literal comparisons of musical and linguistic syntax. Consequently, much of the research examining syntactic relationships between music and language focuses on the *processing* of syntax in both domains, rather than on literal translations between musical and linguistic syntactic units.

How is syntax processed in the brain? In language, syntactic processing has been dissociated from semantic processing using a variety of methodologies, including electroencephalography (EEG; Neville et al., 1991) and fMRI (Friederici, 2003). Early neuroimaging work in this area found evidence for the involvement of "Broca's Area" in the processing of syntax in language (Embick et al., 2000). However, the specific role of Broca's Area in the context of the classic Broca–Wernicke–Lichtheim–Geschwind model of language (e.g., Geschwind, 1970) has been more recently reexamined, in part because it is not clearly marked anatomically (Tremblay & Dick, 2016) and appears to dynamically change functionality depending on coactivation of other related networks (Hagoort, 2014). As such, it is more appropriate to characterize syntax processing as being subserved by a broader network than Broca's Area, with the left inferior frontal gyrus and anterior superior temporal gyrus critically implicated in syntactic processing (Grodzinsky & Friederici, 2006).

Intriguingly, musical syntax has been argued to have substantial shared processing overlap with linguistic syntax (Patel, 2003). Converging evidence from several neuroscientific methodologies, including EEG (Patel, Gibson, et al., 1998), magnetoencephalography (MEG; Maess et al., 2001), and fMRI (e.g., Tillmann et al., 2003) highlight the similarities in both the time course of processing musical and linguistic syntax, as well as the presumed activated networks (including the inferior frontal gyri and “Broca’s Area”). These findings have been incorporated into the shared syntactic integration resource hypothesis (SSIRH; Patel, 2003), which posits that the online processing of both music and language syntax draws upon a common syntactic processor in the brain. Although the SSIRH might initially appear to be at odds with double dissociations found between music and language – specifically, amusia without aphasia (Peretz et al., 1994) and aphasia without amusia (Luria et al., 1965), the hypothesis states that both domains are representationally distinct (i.e., have domain-specificity), and only draw upon common processing resources when integrating elements into larger syntactic structures.

One prediction that stems from the SSIRH is that syntactic processing demands in music should interfere with processing demands in language (or vice versa) – a finding that has been supported in the literature (e.g., Fedorenko et al., 2009; Slevc et al., 2009). Neuroimaging (fMRI) investigations of these interference effects have found that the processing of musical and linguistic syntax results in interactive effects in the left inferior frontal gyrus – consistent with the SSIRH and the view that the left inferior frontal gyrus is critically involved in the online processing of syntax, regardless of domain (Kunert et al., 2015). In further support of the SSIRH, individuals with Broca’s aphasia were found to exhibit syntactic processing difficulties for music, with these difficulties not easily attributable to low-level deficits in pitch perception or short-term memory (Patel et al., 2008).

The SSIRH has received considerable support from converging methodologies. However, the primary challengers to this hypothesis have questioned whether the behavioral and neuroscientific results are most parsimoniously explained under a “syntactic processor” framework, as opposed to a more general cognitive framework invoking attentional and memory processes. This has been shown using “garden path” sentences (i.e., sentences that are grammatically correct, but bias an individual toward the wrong interpretation and thus require online reevaluation and reanalysis). Work supporting the SSIRH found that syntactic garden path sentences (e.g., “The old man the boat”) interfere with the processing of musical syntax (Slevc et al., 2009). However, follow-up work found that this interference is not specific to syntactic processing, as *semantic* garden path sentences (e.g., “The old man went to the bank to withdraw the net”) result in the same patterns of interference on musical syntax (Perruchet & Poulin-Charronnat, 2013). These results suggest that it is the nature of garden path sentences – which broadly place demands on attention and working memory to reevaluate and correctly interpret – rather than syntactic processing demands specifically, that might be most parsimoniously explaining shared processing between musical and linguistic structure. This challenge to the SSIRH can be integrated into a broader research literature suggesting that Broca's Area might be better explained in terms of broader (working) memory processes in sequence understanding (Fiebach et al., 2005; Stromswold et al., 1996). In addition, musical syntax has also been explained in terms of auditory short-term memory processes (Bigand et al., 2014). These findings do not challenge the empirical findings used in support of the SSIRH; rather, they suggest that an alternative explanation of the results might involve broader auditory memory and attentional processes over specific syntactic processing.

The relationship between musical and linguistic syntax is thus complex and actively debated. At first glance, the similarities between the two domains appear numerous (e.g., both involve the sequencing of smaller units into hierarchical structures), which has undoubtedly contributed to the notion of music as a kind of language. However, the presumed similarities between musical and linguistic syntax depends on the level of analysis which one applies to syntactic structures in both domains. The most promising approach in this area of research has been to focus on the *processing* of syntactic structures in both domains, rather than on deriving a common taxonomy for describing syntactic units. To this end, the SSIRH has become a highly influential framework for understanding syntactic processing in both music and language. Although the SSIRH has been supported in many contexts, using converging methodologies to show overlapping neural networks (e.g., inferior frontal gyrus) and behavioral interference effects in processing musical and linguistic syntax, one outstanding question in this area of research is whether these findings are best explained via broader auditory attention and memory mechanisms.

17.2.3 Semantic Features of Music and Language

Both music and language are communicative systems that convey meaning to listeners. As such, they can be considered under the framework of semantics (e.g., Crystal, 1981). In this section, we first highlight the key differences in how musical and linguistic meaning operate. We then turn to some promising approaches to assess the extent to which semantic processing might overlap in musical and linguistic domains.

Musical meaning is largely derived from its intramusical structure (e.g., Lerdahl & Jackendoff, 1983; Meyer, 1956; Narmour, 1990), which establishes (and sometimes violates) listeners' expectations in service of aesthetic or emotional meaning (Steinbeis et al., 2006). Under this framework, musical meaning is *self-referential* – that is, determined by musical structure, which blurs the distinction between musical syntax and musical meaning (e.g., Asano & Boeckx, 2015). Although music can certainly take on meaning beyond its own internal structure (for example, externally associating the repeating two note leitmotif from the "Jaws" soundtrack, composed by John Williams, with a shark or with an ominous event more broadly), these "extra-musical" associations have been treated as a more ancillary source of meaning compared to meaning generated from the musical structure itself (Meyer, 1956).

In contrast, linguistic meaning is derived largely from its ability to symbolically reference ideas outside of the internal structure of language.[1] In describing the "design features" (key properties) of language, Hockett (1958, 1960) describes how language is differentiated from other forms of communication (e.g., nonhuman animal communication) and details several elements of language that other forms of communication do not possess (although see Wacewicz & Żywiczyński, 2015 for a modern critique; also see Andrews, 2014, 2019 for a perspective from cognitive neuroscience). Of particular importance for the present chapter is the design feature of *arbitrariness* – that is, the referential units in language (e.g., spoken or written forms of the word "dog") are arbitrarily associated with the concept (in this case, the concept of a furry, four-legged animal). The arbitrariness of language is thought to be a critical feature of the generativity of linguistic meaning, as well as expressing abstract concepts (e.g., liberty), as there could not be a sound or symbol that would "naturally" map onto this construct (Lupyan & Winter, 2018).

However, the notion of *arbitrariness* as a design feature of language has been challenged for decades (e.g., see Bolinger, 1949; Jakobson, 1995). Specifically, there exist contexts in which language-meaning mappings are *iconic*, in which the symbol (e.g., the sound of a word) maps onto its referent in a nonarbitrary manner. Examples of these nonarbitrary mapping include onomatopoeia (e.g., the words *buzz* or *hiss* mimic acoustic features of their referents), as well as less obvious iconic relationships between words and their meanings, such as the association between *sl-* sounds and negative properties, such as *slime, sludge, slum,* and *slur* (Perniss et al., 2010). One of the most prominent examples of nonarbitrary mappings in language is the "maluma/takete" effect (also referred to as the "bouba/kiki" effect), in which sounds like "maluma" or "bouba" map onto rounded shapes, whereas sounds like "takete" or "kiki" map onto to spiky shapes (Köhler, 1929; Ramachandran & Hubbard, 2001). The bouba/kiki effect replicates across cultures and writing systems (Ćwiek et al., 2022). Moreover, this iconic mapping – also called *sound symbolism* (e.g., Nygaard et al., 2009) – occurs in real-world instances (i.e., is not limited to pseudowords like "bouba" or "kiki") (Sidhu et al., 2021), more frequently than would be expected by chance, and facilitates word learning in children (Monaghan et al., 2014).

These iconic mappings between linguistic sounds and their referents open up the possibility that language and music might share more fundamental ways of conveying meaning, especially given the case that spoken language and music have a number of acoustic similarities (e.g., see Section 17.2.1). The foundations for these sound symbolic mappings have been found in prelinguistic infants, which suggests that the foundation for this sound-meaning mapping might be derived from broader statistical properties of the environment or even innate synesthetic correspondences (Imai & Kita, 2014; Walker et al., 2010). Moreover, individuals have been shown to spontaneously use this analog acoustic channel of expression, independent from the propositional structure of language, to convey referential information about an object. For example, individuals modulate their speaking rate to convey information about an object's speed and modulate their fundamental frequency to convey information about an object's vertical movement (Shintel et al., 2006). Subsequent work demonstrated that underscoring of spoken sentences with either accelerating or decelerating musical motifs influenced individuals' perceptual interpretations of the sentence content (e.g., picturing a galloping horse if the music was accelerating and a resting horse if the music was decelerating), suggesting that these nonarbitrary mappings interact across domains, in a conceptually similar manner to the SSIRH but for semantic processing (Hedger et al., 2013).

How might these nonarbitrary associations between sound and meaning in both music and language be similarly represented in terms of neural activity? In EEG paradigms, semantic violations have reliably elicited a specific ERP component – the N400 – making this a promising methodology and biomarker for addressing this question. The N400 is a negative peaking electrophysio logical event peaking about 400 milliseconds after a stimulus onset and is consistently implicated in semantic – but not syntactic – violations (Kutas & Hillyard, 1980). For example, if one were presented with the written sentence "She takes her coffee with cream and floor," one word at a time, one would expect an N400 ERP following the word "floor" as it represents a semantic violation given the preceding context.

Subsequent research on the N400 has shown that it is elicited by semantic violations that extend beyond language (see Kutas & Federmeier, 2011 for a review), including the processing of pictures (Ganis et al., 1996), mathematical symbols (Niedeggen & Rösler, 1999), and nonlinguistic environmental sounds (Uddin et al., 2018). These findings, which associate the N400 with violations of meaning more broadly, suggest that music might also exhibit N400 responses. Although early work in this area did not find any evidence that violations in music elicited N400s, musical violations were marked by different ERP signatures such as the P300 (Besson & Macar, 1987). Critically, however, this work operationalized musical expectancy violations via intramusical violations (altering notes within a scale or melody), not meaning derived from sound symbolic mappings.

Koelsch and colleagues (2004) tested whether music might elicit an N400 by examining semantic processing in music more in line with the ideas of sound symbolism. In this paradigm, individuals were presented with written words (e.g., "wide"), which could be preceded by a related or an unrelated prime, in either language or music. In language, the related prime was a spoken sentence that was semantically related to the target word (e.g., "The gaze wandered into the distance"), whereas the unrelated prime was a spoken sentence incongruent with the target word (e.g., "The manacles allow only little movement"). In music, the related prime was an excerpt that was selected to reference the target word through nonarbitrary acoustic-meaning mappings (e.g., playing an excerpt from Opus 54 by Richard Strauss, which depicts wideness through instrumental spacing set in open position), whereas the unrelated prime was an excerpt incongruent with the target word (e.g., playing an excerpt from *E-minor piece for accordion* by Heikki Valpola, which depicts narrowness through closely spaced chords). The critical questions of this research were (1) does a musical prime result in an N400 if incongruent with the subsequent word, and (2) if so, how do the electrophysiological components of this ERP differ between linguistic and musical primes? The results demonstrated that, in this capacity, music elicited an N400 and moreover, the signature of the N400 (e.g., its latency, its magnitude) was statistically comparable to the N400 elicited by linguistic primes. The latter point is particularly important given that N400 amplitudes can index the strength of semantic priming and violation (Kutas & Federmeier, 2011). Although the findings of Koelsch and colleagues (2004) are important in demonstrating that music can convey sufficient specificity in referential meaning to elicit N400 responses, the findings should not be interpreted as evidence that music represents meaning in the same manner as language (e.g., see Slevc & Patel, 2011). In other words, even if music conveys meaning through sound symbolic mappings, this is inherently coarser in terms of communicative specificity compared to language, which can make use of both sound symbolic and arbitrary sound-to-meaning mappings.

Although sound symbolic mappings provide one means of comparing meaning across language and music, paralinguistic (nonlexical) cues in speech provide a particularly strong point of comparison across domains. In speech, speakers manipulate prosody – the stress, timing, and intonation of speech – to convey information to listeners. Prosody is often discussed in two realms: affective (emotional) prosody, which conveys the emotional state of the speaker (Pichon & Kell, 2013), and linguistic prosody, which provides structural disambiguation, such as differentiating a question from a statement through pitch (Bolinger, 1983). Prosody is often described as the "music" of speech (Cole, 2015), with affective prosody providing some of the clearest links between meaning in music and language, given that both signals are considered to convey emotional information.

One question that arises in comparing prosody and music is whether both systems make use of similar acoustic manipulations to convey emotional state. In other words, for a given emotional state (e.g., sadness), do the acoustic signatures in speech and music resemble one another? Answering this question is complicated by the fact that both music and language are culturally constrained; thus, the answer to the question may depend on how language and music are being defined in a particular sociocultural context. The extent to which emotional categories are universally represented or recognized across musical and linguistic systems is actively debated (e.g., Fritz et al., 2009; Paulmann & Uskul, 2014; Smit et al., 2022; see also Nettl, 1999 for an alternative account on the universality of meaning in music). Nevertheless, within a Western music framework, research has addressed whether there might be consistent acoustic cues that are used in both systems to convey emotion.

In Western music, "basic" emotions such as happiness and sadness have been associated with consistent psychoacoustic properties. Happiness is often represented in terms of major scale tonality, a relatively fast tempo, and relatively loud dynamics, whereas sadness is often represented in terms of minor scale tonality, a relatively slow tempo, and relatively soft dynamics (e.g., Juslin, 2000). In speech, these general acoustic properties also reflect happiness and sadness in emotional prosody, possibly because these cues covary with arousal (e.g., Curtis & Bharucha, 2010; Scherer, 1995), which has been used to further bolster the conjecture that both music and language evolutionarily stemmed from a protolinguistic, affective communicative system (Mithen, 2005).

Associations between emotional prosody and music perception have also been used to investigate and ultimately challenge the domain-specificity of musical and linguistic representations. As mentioned previously, a double dissociation between music and language has been reported (aphasia without amusia and amusia without aphasia). Yet, given the substantial overlap in how emotion is conveyed in both music and emotional prosody, it is possible that individuals with *amusia* might show impairments in emotional prosody perception, highlighting shared processing across domains. This was verified empirically by Thompson and colleagues (2012). In their design, participants with amusia were compared to control participants without amusia in terms of decoding semantically neutral sentences (e.g., "The broom is in the closet and the book is on the desk") spoken with six different intended emotional categories (happy, sad, tender, irritated, afraid, and no emotion). Although the two groups performed comparably for the no-emotion condition, the listeners with amusia were significantly impaired relative to the control group in identifying the other emotion categories. These findings are consistent with neuroimaging studies, which demonstrate substantial overlap in the processing of emotion in speech and music (e.g., Escoffier et al., 2013), and suggest that music and language convey emotional information via shared networks.

Comparing how music and language convey meaning is inextricably tied to how meaning is defined in both systems. In language, the arbitrary association between symbols and their referents allows for the composition of limitless and novel meanings. In contrast, musical meaning is often determined through intramusical structure, blurring the line between semantic and syntactic representations. However, there are several promising means of comparing meaning in both music and language. First, both music and language can convey meaning using sound symbolic processes, with music showing comparable electrophysiological signatures of semantic violations as language under these circumstances. Second, emotional prosody in language has clear acoustic parallels to music, and both systems appear to convey emotional meaning via similar psychoacoustic properties. These findings suggest that although semantic representations in music and language might be distinct under many circumstances, the two systems exhibit overlap in processing when meaning is inextricably tied to the properties of the sounds (i.e., when sound-to-meaning mappings are nonarbitrary).

17.3 Development of Language and Music

A common theme throughout this chapter is that language and music both draw on the auditory system in complex and sometimes overlapping ways. However, the interaction between these domains is not limited to featural parallels along acoustic, syntactic, and semantic dimensions. Rather, the cross-modal exchange between music and language is intertwined in development across the lifespan, drawing on both domain-specific mechanisms and domain-general developmental processes. In this section, we review the role of critical periods in the development of music and language (Section 17.3.1), the implicit mechanisms that underpin learning processes within these domains (Section 17.3.2), and the ways in which training in one domain can influence processing in the other domain (Sections 17.3.3 and 17.3.4). Throughout this section, we highlight the ways in which auditory experience is shaped in the brain and represented in the world through music and language.

17.3.1 Early Development: Critical Periods

The notion of a critical period for language acquisition – that is, a language is acquired most naturally and accurately within a clearly defined developmental window – is widely accepted (see Lenneberg, 1967; Penfield, 1965). The term "critical period" is often used interchangeably with the term "sensitive period" but conveys an important distinction: whereas critical period suggests that windows of acquisition are sharply defined, sensitive period suggests that humans experience gradual shifts in sensitivity to stimuli in their environment. Here, we use the term critical period for consistency, but also suggest that windows for language and music acquisition are flexible.

Several aspects of language are acquired within the first few months of life (e.g., Johnson & Newport, 1989; Jusczyck & Bertoncini, 1988; Mehler et al., 1988; see for review, Dahaene-Lambertz et al., 2008). In fact, the critical period might even begin before birth since newborn infants show preference for languages they hear *in-utero* (Mehler et al., 1988). Within the first few days of birth, infants can discriminate between languages with different rhythmic structure (Dehaene-Lambertz et al. 2006; Nazzi et al., 1998; Querleu et al. 1988), identify and categorize abstract phonemes (Dehaene-Lambertz & Pena, 2001), and by four months of age, even distinguish their native language from rhythmically similar languages (Nazzi and Ramus, 2003). However, at around four to six years old, children's proficiency in detecting such differences declines rapidly, plateauing around adulthood (Johnson & Newport, 1989, 1991; Newport, 1990; see for review: Meisel, 2013). Although the specific onset and offset times of the critical period for language acquisition continues to be debated, one view is held consistent: proficiency in a given language is strongly related to the age of exposure to that language (e.g., Newport, 1990; Johnson & Newport, 1989, 1991; Oyama, 1976).

By contrast, evidence for a critical period for the acquisition of musical abilities is less widespread. The strongest claims for a critical period in music are grounded in the phenomenon of absolute pitch (AP), which is the ability to name or produce a musical note without a reference (e.g., Deutsch, 2013). Critical periods in AP are supported in principle by several converging findings, including (1) associating the relative incidence of AP with the age at which individuals began musical instruction (Bachem, 1955), (2) training studies that demonstrate superior AP learning among children compared to adults (Crozier, 1997), and (3) pharmacological interventions, suggested to "reopen" critical periods in animal models, leading to enhanced AP learning among adults (Gervain et al., 2013).

Another line of evidence to support critical periods in AP involves the intersection of musical and linguistic processing. Specifically, one theory of AP acquisition is that early experience attending to pitch in a linguistic context (e.g., in the case of tonal languages, such as Mandarin, Cantonese, and Vietnamese) facilitates the development of AP if musical lessons are begun relatively early, when a presumed critical period is still open (Deutsch et al., 2004). For example, Deutsch and colleagues (2006) compared AP prevalence in two groups of first-year music conservatory students – those who either spoke a tonal language (Mandarin) or a nontonal language (English) – and found a greater prevalence of AP in Mandarin- compared to English-speaking students. From these findings, it seems that we might acquire aspects of music in the same way as learning a first language. In fact, a global study conducted by Liu and colleagues (2023) involving upward of 30,000 adult native speakers across nineteen tonal languages and nearly half a million native speakers of other, atonal languages found that tonal language speakers were better at discriminating melodies, but worse at beat perception relative to atonal language speakers. The authors concluded that linguistic experience might shape the perception of music. However, the notion of critical periods in AP has been recently challenged by claims that some adults can acquire AP with sufficient training (Van Hedger et al., 2019; Wong et al., 2020). These findings parallel recent claims in language research – for example, suggesting that developing a native-like accent in a second language is possible under specific conditions beyond a critical period (Dollmann et al., 2020). Overall, these findings may still be consistent with a critical period framework, in which plasticity to acquire new skills is higher within a specified period of development but may persist to some degree across the lifespan (Newport, 2006).

Such parallels between language and music might prompt us to think of the ways in which these domains draw on similar cognitive capacities and experiences early in development. Several aspects of language learning, including accent development, language production, comprehension of morphology and syntax, as well as processing speed and accuracy are facilitated within critical periods during infancy and childhood. Some evidence for this comes from second language learning. For example, Korean- and Chinese-speaking adults make significant errors in grammatical judgments of English with age. Moreover, age of exposure also impacts the way language is represented in the brain. Additionally, ERP and fMRI studies show strong left hemisphere activation for their native language in bilingual adults (Dehaene et al., 1997; Perani et al., 1996; Yetkin et al., 1996). When a second language is learned, patterns of brain activation nearly overlap with those for the first or native language. For late language learners however, neural organization is less lateralized and more variable across individuals (Dehaene et al., 1997; Kim et al., 1997; Yetkin et al., 1996). Several cross-linguistic studies show domain-specific contrasts in first and second language learning (e.g., Neville et al., 1997; Weber-Fox & Neville, 1996) providing evidence for a critical period for acquiring phonological and grammatical patterns of language learning.

A closer look at the role of speech perception in language learning shows that a language-specific pattern of listening accelerates language acquisition over the first two-and-a-half years of life. This is evidenced by a phenomenon of *perceptual narrowing* with increased exposure to the native, over other languages between six and twelve months of age – the critical period for phonetic learning (Oyama, 1976). Werker and Tees (1984) demonstrated this phenomenon of perceptual narrowing in a study investigating infants' ability to discriminate English from other (i.e., Hindi and Salish) consonants. English-speaking infants showed a reduced ability to discriminate between Hindi and Salish consonant contrasts at twelve compared to six months of age, even though native Hindi- and Salish-learning infants continued to effectively discriminate between them. Subsequent evidence from behavioral and neurological studies supports the connection between speech perception and language acquisition. Kuhl and colleagues (2005; 2006) used both behavioral and ERP measures to show that infants' ability to discriminate native versus non-native speech at seven months of age differentially predicts later language outcomes at two years of age: infants better at discriminating native speech showed faster advancements in language acquisition. These results were later extended to show discrimination between English and Spanish (Rivera-Gaxiola et al., 2005, as well as Finnish and a non-native Russian contrast (Silvén et al., 2014), among other languages.

Corresponding evidence of perceptual narrowing is also found in music, as musical experience influences children's perception of musical scales. Lynch and colleagues (1990) tested the ability to notice mistunings in melodies based on Western and non-Western (i.e., Javanese pelog) scales in English-speaking infants and adults, finding that six-month-old infants were similarly able to perceive Western and non-Western scales, whereas adults were better at perceiving Western scales. Follow-up experiments conducted by Lynch and

Eilers (1992) found that six-month-old infants were less perceptually acculturated than one-year-olds, suggesting that young infants have an equipotentiality for detecting differences between musical scales, but show a perceptual narrowing with age.

Given these similarities in learning language and music during early childhood, it might be the case that increased overall auditory experience facilitates perception. Under this view, perceptual narrowing observed during this critical period for phonetic learning is malleable. One example of this malleability comes from research examining infants' ability to discriminate lexical tones, that is, pitch patterns used to contrast word meaning in some languages. Infants learning nontonal languages find it harder to discriminate lexical tones by nine months of age than their tonal language learning counterparts (Mattock et al., 2008; Mattock & Burnham, 2006), but these differences can be shifted through additional auditory experience. For example, a study by Kuhl and colleagues (2003) showed that nine-month-old English-learning infants' ability to discriminate speech contrasts in Mandarin increased after greater exposure to a Mandarin speaker. English-learning infants exposed to Mandarin showed an ability to discriminate speech contrasts on a par with Mandarin-learning infants. Similar research using EEG measures also demonstrated neural sensitivity to speech contrasts (Conboy and Kuhl, 2011; Lytle et al., 2018). An intervention using music, rather than a tonal language, showed a similar shift in children's neural processing of non-native speech contrasts and musical rhythm (Zhao & Kuhl, 2016, 2022). Zhao and colleagues (2022) explored whether music intervention reversed the observed decline of early sensory encoding of non-native speech with gains in native linguistic experience. Seven-to-ten-month-old English-learning infants received a music intervention (i.e., increased auditory experience) and their frequency-following response (FFR) to Mandarin lexical tones was measured longitudinally. The no-intervention group showed a marked decline in FFR pitch-tracking accuracy to Mandarin lexical tones, whereas the music-intervention group showed no such decline. From these findings, it seems that language and music both influence infants' speech encoding within a presumed critical period of development.

17.3.2 Domain-General Learning Mechanisms: Statistical Learning

Newborn infants receive various types of information from their physical and social environments. To make sense of the world, these young learners must keep track of different regularities in their environment and learn complex rules that organize their environment, without prior knowledge about which information is important. One way in which infants do this is through statistical learning, a domain-general learning mechanism by which infants become sensitive to different patterns in their environment implicitly, without external feedback or reinforcement (Saffran et al., 1996; Saffran & Kirkham, 2018).

Initially, statistical learning was limited in scope, conceptualized solely as a computational process of tracking syllable patterns to support segmentation

(Saffran et al., 1996a, 1996b). However, subsequent research has shown that statistical learning is domain-general, extending beyond speech to recognizing regularities within other environmental inputs. Infants can detect statistical patterns within streams of continuous speech by the time they are eight months old.

Similar to speech, listeners are sensitive to musical input from their environments (McMullin & Saffran, 2004). Young learners discover higher-order structures to organize individual units of language (e.g., words and syntax) and music (e.g., melodic and harmonic sequences) in an implicit manner (Fiser & Aslin, 2001; Tillmannet al., 2000). Not only can infants detect regularities in speech, but they also can do the same for musical tones by the time they are eight months old (Saffran et al., 1999). Eventually, listeners can implicitly learn several musical features, including timbre and melodic pitch relationships, from their listening environments (see Ettlinger et al., 2011; Rohrmeier & Rebuschat, 2012). Under the statistical learning hypothesis, the brain not only calculates and organizes the transitional probabilities (TPs) of sequential phenomena (words, tones, etc.) from the environment, but also updates this acquired statistical knowledge in line with new environmental inputs through the lifespan (Daikoku et al., 2017). Differences in brain responses as a result of statistical learning may then depend on the previously acquired knowledge (e.g., Furl et al., 2011; Yumoto et al., 2005).

Evidence for experience and expectancy-related differences in statistical learning can be found in both language and music. In language, continuous speech has both universal regularities and language-specific statistical patterns. For example, adjacent units (syllables) in language have a high transitional probability of co-occurrence across languages (Saffran et al., 1996a, 1996b; Swingley, 2005) that can be implicitly segmented are found across languages, and therefore be leveraged by infant and adult language learners. By contrast, linguistic properties like lexical stress are language-specific, and therefore, not readily available for all language learners (e.g., Sohail & Johnson, 2016). Adult learners might bootstrap language-specific patterns (lexical stress, phonotactics, etc.) onto universal elements such as TPs to identify word boundaries in continuous speech (e.g., Benitez & Saffran, 2021). For example, English-speaking adults given brief, unguided exposure to a foreign language (Italian) in the form of podcasts demonstrate initial word-form learning, compared to adults with no exposure to the foreign language (Alexander et al., 2022), which suggests that second-language learners might be sensitive to the characteristics of different word features (TPs, phonetic patterns, etc.) of a novel language after some exposure to that language. In music, some evidence of expectancy-related learning comes from research conducted with musically trained individuals. For example, adult musicians perform better on statistical learning tasks and are generally better at learning the implicit statistical structures of sung language compared to nonmusicians (François and Schön, 2011, 2014; Vasuki et al., 2016). In an ERP study, musicians showed a larger N1 familiarity effect compared to their nonmusician counterparts, which suggested that they are better at recognizing statistical patterns. Moreover, ERPs recorded during the familiarization of a sung language showed that the N400 learning curve saturates earlier in musicians, which suggested that they are also faster at detecting statistical regularities in sung language compared to their nonmusician counterparts (François and Schön, 2014). Similarly, a longitudinal study conducted with eight-year-old children found that children learning music were better at identifying statistical regularities in auditory streams compared to children who were not formally trained in music (François et al., 2014). Moreover, subsequent research by Vasuki and colleagues (2017) showed that musically trained children were better at melody discrimination, rhythm discrimination, frequency discrimination, in addition to auditory statistical learning, compared to their nontrained counterparts; and the triplet onset stimulus used in their statistical learning paradigm elicited larger responses in musically trained compared to untrained children.

Into adulthood, both language and music demonstrate effects of *entrenchment*, wherein the efficacy of learning the statistical regularities of novel language and music sequences depends on prior representations developed within both domains. In language, variance in performance on a novel statistical learning task could be partly explained by an individual's previous language background – specifically, the extent to which the novel sequences aligned with the individual's native language (Siegelman et al., 2018). In music, tonal sequences that did not adhere to a conventional tuning standard were judged to sound "correct" if they incidentally contained a greater proportion of common intervals found in Western music (perfect fourths and fifths), even though the presence of these intervals was not indicative of the structure of the novel sequences (Van Hedger et al., 2022). These parallels across language and music suggest that as listeners develop increasingly rich representations in both domains, subsequent implicit learning efficacy interacts with these previously developed representations.

Language and music both rely at least partially on statistical learning, a domain-general mechanism for learning and processing novel information. Learners might be using this mechanism for acquiring and maintaining properties of both language (Peretz et al., 2012; Romberg & Saffran, 2010; Saffran, 2001; Saffran, 2003) as well as music (Furl et al., 2011; Saffran, 2003; Saffran et al., 1999). The literature on statistical learning shows us that language and music are not simply parallel domains with surface level similarities, but rather, draw on shared domain-general processes that shape auditory experience, with developed representations influencing the efficacy of statistical learning into adulthood.

17.3.3 Training and Transfer between Domains

So far, we have suggested language and music are distinct modalities, comparable not just by surface-level features, but also related by a common auditory sensory experience and maintained by shared cognitive learning mechanisms. We can now turn our attention to the ways in which the modalities interact with each other in cross-modal exchange. One view holds that the prevalence of higher-order structural similarities indicates that music draws from language in unequal measure (Pinker, 1997) or perhaps, even that language is a product of musical experience (Darwin, 2007/1874). Another view challenges the idea that these domains are linked at a structural level (e.g., Federenko, Behr, et al. 2011; Federenko, McDermott, et al., 2012). While music and language have differentiable representations at the neural level, overlap in the sensory and cognitive systems employed by these modalities allows room for the possibility that experience and training in one domain might facilitate performance in another domain.

Evidence for training and transfer effects from language to music come from research conducted with linguistically and musically trained samples. In one study, Bidelman and colleagues (2013) compared adult Cantonese-speaking nonmusicians to English-speaking trained musicians and English-speaking controls on pitch discrimination and pitch memory tasks. They found that Cantonese-speaking participants outperformed English-speaking controls on all aspects of pitch perception and discrimination. By contrast, Cantonese speakers performed similarly to their English-speaking musically trained counterparts, suggesting that pitch expertise, arising from either musical training or tonal language experience, is associated with lower sensitivity to pitch discrimination, higher tonal memory and melodic discrimination. Moreover, the authors found a significant relationship between behavioral measures of pitch ability and length of exposure to the tonal language or musical training. In another study, Elmer and colleagues (2011) conducted an event-related fMRI study to examine whether long-term language training facilitates the discrimination of nonverbal stimulus attributes in a nonverbal auditory discrimination task, and addressed whether functional transfer effects might support behavioral performance by comparing professional simultaneous interpreters to controls. Their results suggested that discrimination of target stimuli was associated with differential BOLD responses in the frontoparietal regions of the two groups, even though behavioral results showed no significant effects. Language training seemed to modulate brain activity in regions associated with top-down regulation of auditory functions, such as auditory attention and categorization. These findings are in line with the idea that overall auditory experience plays a crucial role in language processing.

Parallel evidence for music to language transfer comes from different levels of processing: multiple studies have examined acoustic features (e.g., Chartrand & Belin, 2006; Bidelman & Krishnan, 2010; Weiss & Bidelman, 2015), cognitive factors (e.g., Anvari et al., 2002; Moreno et al., 2009), and executive function (e.g., Bialystok & DePape, 2009). Moreover, the impact of musical training on language acquisition and development has been explored in the context of neural encoding of speech in noise (Strait et al., 2009; Strait & Kraus, 2011), phonological awareness (e.g., Tsang & Conrad, 2011), as well as vocabulary and reading comprehension (e.g., Corrigal & Trainor, 2011;

Moreno, Bialystok et al., 2011). Musical training studies corresponding to the ones discussed above reliably show that musical training facilitates linguistic performance. For example, Magne and colleagues (2006) found that musically trained children detected incongruities in music and language better than their musically untrained counterparts. Musically trained children showed early negative components in music and late positive components in language, whereas no components were present in their untrained counterparts. These results are in line with the finding that musically trained individuals show enhanced neural encoding of speech sounds in the auditory brainstem compared to their untrained counterparts (Kraus & Chanrasekaran, 2010).

Similar to language-to-music transfer, enhanced performance in language is also correlated with length and degree of exposure to musical training (e.g., Besson et al., 2011; Musacchia et al., 2007; Strait et al., 2009). Although these studies relied on expert populations, longitudinal training experiments have yielded similar results. For example, Moreno and colleagues (2009, 2011a, 2011b) conducted a series of longitudinal studies to explore whether musical training facilitated language processing in young children. Eight-year-old children given musical training showed improvements in EEG correlates of speech-related pitch processing after receiving six months of training, compared to matched controls. Behaviorally, four-to-six-year-old children showed similar enhancements in verbal intelligence after participating in an intensified version of the musical training program conducted over twenty days (Moreno, Bialystok et al., 2011). These findings illustrate that a degree of plasticity may be acquired through exposure to auditory activities in the form of language or music.

17.3.4 Mechanisms Underlying Transfer: Insights from Cognitive Neuroscience

A central question emerging from examining the developmental learning trajectories of music and language, including how training in one domain can transfer to the other domain, is: *what exactly is being learned*? Understanding the perceptual, cognitive, and neural processes by which the brain generalizes and transfers information from one domain (e.g., music) to another (e.g., language) requires us to conceptualize the two domains as relying on shared auditory and attentional systems. Moreover, given recent evidence that lifelong musicianship is associated with preserved white matter tract integrity in individuals, which may promote healthy aging (Andrews et al., 2021), understanding the proposed mechanisms of transfer has implications for designing interventions to improve psychological and neurobiological health across the lifespan.

One explanation for the cross-modal interaction between language and music comes from the OPERA hypothesis (Patel, 2011), which states that, neural areas recruited in music and language overlap (O) when partaking in a learning process that involves precision (P), emotional engagement (E), repetition (R), and attentional focus (A). According to this theory, cross-modal transfer occurs because largely overlapping cortical and subcortical brain regions are recruited when the acoustic features of speech and music are encoded. Increased precision of neural processing through the components of this model (i.e., emotional engagement, repetition, and attention) make the acoustic information more salient to the listener. Cross-modal transfer is possible when acoustic features of music or language are encoded with a high degree of precision with increased exposure (e.g., tonal language) and training (e.g., formal musical training). Training allows learners to give their attention to the acoustic features of music and language, and this enhanced attention in turn mediates the overlapping neural networks involved in music and language processing.

Building on the OPERA hypothesis, Moreno and Bidelman (2014) proposed a different model of cross-modal transfer. Under this view, the neural networks determining the transfer of auditory information across music and language are affected by two dimensions. First, is a sensory-cognitive dimension that distinguishes low-level auditory-sensory information from high-level domain-general processes that support language. Whereas some training might lead to experience-dependent plasticity, other training might impact cortical plasticity, leading to overarching changes in language processing and executive function. A second dimension along which training and transfer occurs is the near-far dimension. A near-transfer involves seeing the development of more precise listening skills (relative pitch discrimination, improvements in pattern detection, etc.) because of repeated exposure to auditory patterns. By contrast, a far-transfer occurs when the music training benefits auditory sensory encoding in speech production and language acquisition. The benefit of musical training is not just a product of length and intensity of training, but also the degree to which the training targets higher-order cognitive skills.

Both these accounts shed some light on the ways in which underlying neural and cognitive mechanisms underlying music and language might be facilitating a cross-modal transfer between these domains. The developmental trajectory of these cognitive domains, discussed in earlier parts of the chapter, also lends insights into the possibilities and restraints on transfers between them (cf. discussion of sensory periods: White et al., 2013). Some skills (AP, detecting unfamiliar sounds, etc.) emerge early, and may be restricted by a developmental timeline. However, it also seems true that both abilities interact beyond these critical periods, and might contribute to learning and development (e.g., second language acquisition, musical pattern detection) within the other domain.

A third account of transfer across domains is that training plays a relatively small role, with preexisting differences accounting for most of the shared variance in language and musical abilities (e.g., Swaminathan & Schellenberg, 2019; see also Figure 17.1). For example, Swaminathan and Schellenberg (2020) found that speech abilities were positively correlated with musical abilities and IQ among six-to-nine-year-old children; however, musical training

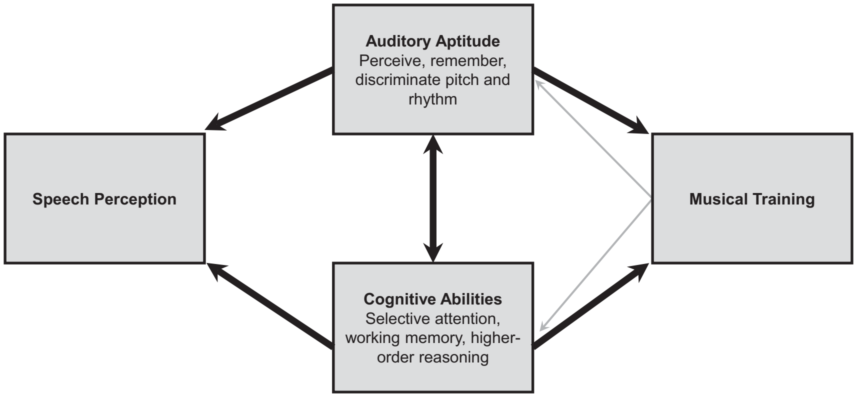

Figure 17.1 Potential paths explaining observed associations between musical training and spoken language perception.
Note: The width and shading of the arrows represent pathway strength, with thicker and darker arrows representing stronger connections. The figure outlines two primary ways through which musical training could be associated with speech perception. The first, non-causal path is due to preexisting auditory and cognitive abilities (middle), which then influence an individual's tendency to continue with musical training and also relate to an individual's spoken language abilities. The second, causal path shows how musical training could improve both auditory and cognitive abilities, which then in turn could improve speech perception.
Figure modified and expanded upon from Swaminathan and Schellenberg (2019).
Thaut, M. H., & Hodges, D. A. (Eds.). The Oxford Handbook of Music and the Brain. Reproduced with permission of the Licensor through PLSclear.

was not a significant predictor of language abilities. Among adults, recent work has reported cases of "musical sleepers" – individuals without formal musical training who nevertheless demonstrate perceptual acuity and electrophysio logical signatures of sound encoding (e.g., FFRs) that are comparable to individuals with extensive musical training (Mankel & Bidelman, 2018). In line with the preexisting difference account, research has also demonstrated that personality factors (e.g., openness) are associated in both children and adults with the likelihood of taking music lessons and persisting with training (Corrigall et al., 2013). Although these findings do not inherently reject the notion that training in one domain may transfer to another domain, they strongly caution against the use of quasi-experimental designs (e.g., comparing trained musicians to nonmusicians) in making causal claims about the nature of training and transfer.

Through a discussion of (1) the developmental constraints on acquiring music and language, (2) the domain-general learning mechanisms involved in learning music and language, and (3) the mechanisms underlying cross-modal transfer between music and language in this section, we suggest that learning in these domains is both an implicit, bottom-up process that relies on auditory-sensory input to engage overlapping neural networks, and an explicit, top-down process that requires precise encoding of information to facilitate cross-modal transfer between the two domains. To better understand the underlying mechanisms that allow for training and transfer, longitudinal studies with random assignment to condition are critical, given the reports that preexisting auditory processing and personality factors tend to covary with musical training.

17.4 Conclusion

Comparisons between music and language have been the focus of research in psychology and neuroscience for several decades. Throughout this chapter, we demonstrate how music and language overlap in terms of psychoacoustic properties, recruit shared neural networks in processing syntactically organized sequences, and can use nonarbitrary cues to convey meaning. Given these surface-level similarities between language and music, it seems tempting, and even practical to think of the two domains as related and parallel domains. However, applying too fine-grained a level of analysis, such as focusing solely on featural cues, restricts the way we think about the interactions between these domains. For example, language and music are organized hierarchically at multiple levels of analysis, ranging from the phonological to morphological, and sentential level in language, and from harmonic to rhythmic levels in music. Seeing these solely as hierarchical processes limits us from comparing the units (morphology, rhythm, etc.) in a meaningful way. Conversely, looking for functional equivalence between the two domains might lend itself to miscalibrating the true underlying purpose and function of these skills. For example, meaning is conveyed differently in both music and language: it differs not just in transparency, but also content. It is difficult to compare, and therefore, can be dangerous to equate across modalities. Music and language are not just parallel domains that share surface-level features and should be treated and compared at an appropriate level of analysis.

Despite these pitfalls that inevitably arise from featural comparisons of music and language, studying the cognitive and perceptual processes underlying these domains can provide insight into the ways in which music and language interact and draw on each other. Both modalities are maintained by domain-general learning mechanisms, with similar constraints on developmental trajectories of learning (e.g., critical periods) and shared neurocognitive systems for supporting implicit learning via statistical regularities in the environment. Moreover, training in one domain often leads to enhanced skills in the other domain, showing evidence for a complex training and transfer mechanism. Applying an appropriate level of analysis demonstrates that music and language interact in sophisticated ways and can provide insight into shared (domain-general) versus distinct (domain-specific) neural and cognitive mechanisms.

Notes

1 Several related accounts have proposed alternative and complementary sources of linguistic meaning, such as sensorimotor systems (e.g., Gallese & Lakoff,

2005), embodied perception and action (e.g., Gibbs, 2003), metaphoric images and schemas (e.g., Glucksberg & Keysar, 1990; Murphy, 1996), as well as nonreferential sources of musical meaning including embodiment (e.g., Cox, 2001), and culture (e.g., Shepard & Wicke, 1997). For the purposes of our discussion, however, we are focusing on the similarities and differences in language and music from the human ability to derive meaning by mapping sounds to words and words to their referents (Section 17.2.1) and bootstrapping our hypotheses about the meanings of words, sentences, and musical sequences from syntax (Section 17.2.2).

References

Adger, D. (2015). Syntax: Syntax. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(2), 131–147. https://doi.org/10.1002/wcs.1332

Albouy, P., Benjamin, L., Morillon, B., & Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*, *367*(6481), 1043–1047. https://doi.org/10.1126/science.aaz3468

Alexander, E., Van Hedger, S. C., & Batterink, L. J. (2023). Learning words without trying: Daily second language podcasts support word-form learning in adults. *Psychonomic Bulletin & Review*, 30(2), 751–762. https://doi.org/10.3758/s13423-022-02190-1

Andrews, E. (2014). *Neuroscience and Multilingualism*. Cambridge University Press.

Andrews, E. (2019). Cognitive neuroscience and multilingualism. In J. W. Schwieter & M. Paradis (Eds.), *The Handbook of the Neuroscience of Multilingualism* (1st ed., pp. 19–47). Wiley. https://doi.org/10.1002/9781119387725.ch2

Andrews, E., Eierud, C., Banks, D., Harshbarger, T., Michael, A., & Rammell, C. (2021). Effects of lifelong musicianship on white matter integrity and cognitive brain reserve. *Brain Sciences*, *11*(1), 67. https://doi.org/10.3390/brainsci11010067

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*, 83(2), 111–130. https://doi.org/10.1016/S0022-0965(02)00124-8

Asano, R., & Boeckx, C. (2015). Syntax in language and music: What is the right level of comparison? *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00942

Bachem, A. (1955). Absolute pitch. *The Journal of the Acoustical Society of America*, *27*(6), 1180–1185. https://doi.org/10.1121/1.1908155

Benitez, V. L., & Saffran, J. R. (2021). Two for the price of one: Concurrent learning of words and phonotactic regularities from continuous speech. *PLoS ONE*, 16(6), e0253039. https://doi.org/10.1371/journal.pone.0253039

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology*, 2, 94. https://doi.org/10.3389/fpsyg.2011.00094

Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology*, *24*(1), 14–25. https://doi.org/10.1111/j.1469-8986.1987.tb01853.x

Bialystok, E., & DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 565. https://doi.org/10.1037/a0012735

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLoS ONE*, 8(4), e60676. https://doi.org/10.1371/journal.pone.0060676

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*, *1355*, 112–125. https://doi.org/10.1016/j.brainres.2010.07.100

Bidelman, G. M., & Yoo, J. (2020). Musicians show improved speech segregation in competitive, multi-talker cocktail party scenarios. *Frontiers in Psychology, 11*. www.frontiersin.org/articles/10.3389/fpsyg.2020.01927

Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., & Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory. *Frontiers in Systems Neuroscience*, *8*. https://doi.org/10.3389/fnsys.2014.00094

Bolinger, D. (1949). The sign is not arbitrary. *Boletín del Instituto Caro y Cuervo (= Thesaurus)*, *5*, 52–62.

Bolinger, D. (1983). Intonation and gesture. *American Speech*, *58*(2), 156. https://doi.org/10.2307/455326

Brown, R. (1957). Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology*, *55*(1), 1–5. https://doi.org/10.1037/h0041199

Brown, R. (1973). *A First Language: The Early Stages*. Harvard University Press.

Brown, S. (2017). A joint prosodic origin of language and music. *Frontiers in Psychology*, *8*, 1894. https://doi.org/10.3389/fpsyg.2017.01894

Chartrand, J. P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, 405(3), 164–167. https://doi.org/10.1016/j.neulet.2006.06.053

Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, *30*(1–2), 1–31. https://doi.org/10.1080/23273798.2014.963130

Conboy, B. T., & Kuhl, P. K. (2011). Impact of second-language experience in infancy: Brain measures of first-and second-language speech perception. *Developmental Science*, 14(2), 242–248. https://doi.org/10.1111/j.1467-7687.2010.00973.x

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00222

Corrigall, K. A., & Trainor, L. J. (2011). Associations between length of music training and reading skills in children. *Music Perception*, 29(2), 147–155. https://doi.org/10.1525/mp.2011.29.2.147

Cox, A. (2001). The mimetic hypothesis and embodied musical meaning. *Musicae Scientiae*, *5*(2), 195–212. https://doi.org/10.1177/102986490100500204

Cross, I. (2001). Music, cognition, culture, and evolution. *Annals of the New York Academy of Sciences*, *930*(1), 28–42. https://doi.org/10.1111/j.1749-6632.2001.tb05723.x

Crozier, J. B. (1997). Absolute pitch: Practice makes perfect, the earlier the better. *Psychology of Music*, *25*(2), 110–119. https://doi.org/10.1177/0305735697252002

Crystal, D. (1981). Semantics. In G. E. Arnold, F. Winckel, & B. D. Wyke (Eds.), *Clinical Linguistics* (Vol. 3, pp. 131–191). Springer Vienna. https://doi.org/10.1007/978-3-7091-4001-7_5

Curtis, M. E., & Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion*, *10*(3), 335–348. https://doi.org/10.1037/a0017928

Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., ... & Winter, B. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1841), 20200390. https://doi.org/10.1098/rstb.2020.0390

Darwin, C. (2007/1874). *The Descent of Man: The Concise Edition*. Penguin.

Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., ... & Le Bihan, D. (1997). Anatomical variability in the cortical representation of first and second language. *Neuroreport*, 8(17), 3809–3815.

Dehaene-Lambertz, G., Hertz-Pannier, L., & Dubois, J. (2006). Nature and nurture in language acquisition: Anatomical and functional brain-imaging studies in infants. *Trends in Neurosciences*, 29(7), 367–373. https://doi.org/10.1016/j.tins.2006.05.011

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., & Dehaene, S. (2008). How does early brain organization promote language acquisition in humans? *European Review*, 16(4), 399–411. doi:10.1017/S1062798708000513

Dehaene-Lambertz, G., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport*, 12(14), 3155–3158.

Deutsch, D. (2013). Absolute pitch. In *The Psychology of Music* (pp. 141–182). Elsevier. https://doi.org/10.1016/B978-0-12-381460-9.00005-5

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception*, *21*(3), 339–356. https://doi.org/10.1525/mp.2004.21.3.339

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *The Journal of the Acoustical Society of America*, 119(2), 719–722. https://doi.org/10.1121/1.2151799

Deutsch, D., Lapidis, R., & Henthorn, T. (2008). The speech-to-song illusion. *The Journal of the Acoustical Society of America*, *124*(4), 2471–2471. https://doi.org/10.1121/1.4808987

Diedrich, C. G. (2015). "Neanderthal bone flutes": Simply products of Ice Age spotted hyena scavenging activities on cave bear cubs in European cave bear dens. *Royal Society Open Science*, *2*(4), 140022. https://doi.org/10.1098/rsos.140022

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, *81*, 181–187. https://doi.org/10.1016/j.neubiorev.2017.02.011

Dollmann, J., Kogan, I., & Wei?mann, M. (2020). Speaking accent-free in L2 beyond the critical period: The compensatory role of individual abilities and opportunity structures. *Applied Linguistics*, *41*(5), 787–809. https://doi.org/10.1093/applin/amz029

Elmer, S., Meyer, M., Marrama, L., & Jäncke, L. (2011). Intensive language training and attention modulate the involvement of fronto-parietal regions during a non-verbal auditory discrimination task. *European Journal of Neuroscience*, *34*(1), 165–175. https://doi.org/10.1111/j.1460-9568.2011.07728.x

Embick, D., Marantz, A., Miyashita, Y., O'Neil, W., & Sakai, K. L. (2000). A syntactic specialization for Broca's area. *Proceedings of the National Academy of Sciences*, *97*(11), 6150–6154. https://doi.org/10.1073/pnas.100098897

Escoffier, N., Zhong, J., Schirmer, A., & Qiu, A. (2013). Emotional expressions in voice and music: Same code, same effect? *Human Brain Mapping*, *34*(8), 1796–1810. https://doi.org/10.1002/hbm.22029

Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1491–1506. https://doi.org/10.1037/a0036858

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, 108(39), 16428–16433. https://doi.org/10.1073/pnas.1112937108

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. https://doi.org/10.1152/jn.00209.2012

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. https://doi.org/10.3758/MC.37.1.1

Fiebach, C. J., Schlesewsky, M., Lohmann, G., von Cramon, D. Y., & Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: Syntactic integration versus syntactic working memory. *Human Brain Mapping*, *24*(2), 79–91. https://doi.org/10.1002/hbm.20070

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12(6), 499–504. https://doi.org/10.1111/1467-9280.003

François, C., Jaillet, F., Takerkart, S., & Schön, D. (2014). Faster sound stream segmentation in musicians than in nonmusicians. *PLoS ONE*, 9(7), e101340. https://doi.org/10.1371/journal.pone.0101340

François, C., & Schön, D. (2014). Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: The role of musical practice. *Hearing Research*, 308, 122–128. https://doi.org/10.1016/j.heares.2013.08.018

Francois, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex*, 21(10), 2357–2365. https://doi.org/10.1093/cercor/bhr022

Friederici, A. D. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex*, *13*(2), 170–177. https://doi.org/10.1093/cercor/13.2.170

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, *19*(7), 573–576. https://doi.org/10.1016/j.cub.2009.02.058

Furl, N., Kumar, S., Alter, K., Durrant, S., Shawe-Taylor, J., & Griffiths, T. D. (2011). Neural prediction of higher-order auditory sequence statistics. *Neuroimage*, 54(3), 2267–2277. https://doi.org/10.1016/j.neuroimage.2010.10.038

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3–4), 455–479. https://doi.org/10.1080/02643290442000310

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*(2), 89–106. https://doi.org/10.1162/jocn.1996.8.2.89

Gervain, J., Vines, B. W., Chen, L. M., Seo, R. J., Hensch, T. K., Werker, J. F., & Young, A. H. (2013). Valproate reopens critical-period learning of absolute pitch. *Frontiers in Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00102

Geschwind, N. (1970). The organization of language and the brain: Language disorders after brain damage help in elucidating the neural basis of verbal behavior. *Science*, *170*(3961), 940–944. https://doi.org/10.1126/science.170.3961.940

Gibbs, R. W. (2003). Embodied experience and linguistic meaning. *Brain and Language*, *84*(1), 1–15. https://doi.org/10.1016/S0093–934X(02)00517-5

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review*, *97*(1), 3–18. https://doi.org/10.1037/0033-295X.97.1.3

Grodzinsky, Y., & Friederici, A. D. (2006). Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology*, *16*(2), 240–246. https://doi.org/10.1016/j.conb.2006.03.007

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current Opinion in Neurobiology*, *28*, 136–141. https://doi.org/10.1016/j.conb.2014.07.013

Hedger, S. C., Nusbaum, H. C., & Hoeckner, B. (2013). Conveying movement in music and prosody. *PLoS ONE*, *8*(10), e76744. https://doi.org/10.1371/journal.pone.0076744

Higham, T., Basell, L., Jacobi, R., Wood, R., Ramsey, C. B., & Conard, N. J. (2012). Testing models for the beginnings of the Aurignacian and the advent of figurative art and music: The radiocarbon chronology of Gei?enklösterle. *Journal of Human Evolution*, *62*(6), 664–676. https://doi.org/10.1016/j.jhevol.2012.03.003

Hockett, C. F. (1958). A course in modern linguistics. *Language Learning*, *8*(3–4), 73–75. https://doi.org/10.1111/j.1467-1770.1958.tb00870.x

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*(3), 88–96. https://doi.org/10.1038/scientificamerican0960–88

Honing, H., ten Cate, C., Peretz, I., & Trehub, S. E. (2015). Without it no music: Cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140088. https://doi.org/10.1098/rstb.2014.0088

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *NeuroImage*, *50*(3), 1202–1211. https://doi.org/10.1016/j.neuroimage.2010.01.046

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130298. https://doi.org/10.1098/rstb.2013.0298

Jakobson, R. (1965). Quest for the essence of language. *Diogenes*, 13(51), 21–37. https://doi.org/10.1177/039219216501305103

Jakobson, R. (1995). *On Language* (ed. L. R. Waugh). Harvard University Press.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Jusczyk, P. W., & Bertoncini, J. (1988). Viewing the development of speech perception as an innately guided learning process. *Language and Speech*, 31(3), 217–238. https://doi.org/10.1177/002383098803100301

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1797–1812. https://doi.org/10.1037/0096-1523.26.6.1797

Kim, K. H., Relkin, N. R., Lee, K. M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, 388(6638), 171–174. https://doi.org/10.1038/40623

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), 302–307. https://doi.org/10.1038/nn1197

Köhler, W. (1929). *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. Liveright.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8), 599–605. https://doi.org/10.1038/nrn2882

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period". *Language Learning and Development*, 1(3–4), 237–264. https://doi.org/10.1080/15475441.2005.9671948

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, 100(15), 9096–9101. https://doi.org/10.1073/pnas.1532872100

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. https://doi.org/10.1371/journal.pone.0141069

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Landau, B., & Gleitman, L. R. (2009/1985). *Language and Experience: Evidence from the Blind Child* (Vol. 8). Harvard University Press.

Landau, B., Gleitman, L. R., & Landau, B. (2009). *Language and Experience: Evidence from the Blind Child* (Vol. 8). Harvard University Press.

Larrouy-Maestri, P., Harrison, P. M. C., & Müllensiefen, D. (2019). The mistuning perception test: A new measurement instrument. *Behavior Research Methods*, *51*(2), 663–675. https://doi.org/10.3758/s13428-019-01225-1

Lenneberg, E. H. (1967). *Biological Foundations of Language*. John Wiley & Sons.

Lerdahl, F. (2013). Musical syntax and its relation to linguistic syntax. In M. A. Arbib (Ed.), *Language, Music, and the Brain* (pp. 257–272). The MIT Press. https://doi.org/10.7551/mitpress/9780262018104.003.0010

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music* (Repr.). MIT Press.

Liu, J., Hilton, C. B., Bergelson, E., & Mehr, S. A. (2023). Language experience predicts music processing in a half-million speakers of fifty-four languages. *Current Biology*, *33*(10), 1916–1925.

Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic? *Philosophical Transactions of the Royal*

Society B: Biological Sciences, *373*(1752), 20170137. https://doi.org/10.1098/rstb.2017.0137

Luria, A. R., Tsvetkova, L. S., & Futer, D. S. (1965). Aphasia in a composer. *Journal of the Neurological Sciences*, *2*(3), 288–292. https://doi.org/10.1016/0022-510X(65)90113-9

Lynch, M. P., & Eilers, R. E. (1992). A study of perceptual development for musical tuning. *Perception & Psychophysics*, 52(6), 599–608. https://doi.org/10.3758/BF03211696

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science*, 1(4), 272–276. https://doi.org/10.1111/j.1467-9280.1990.tb00213.x

Lytle, S. R., Garcia-Sierra, A., & Kuhl, P. K. (2018). Two are better than one: Infant language learning from video improves in the presence of peers. *Proceedings of the National Academy of Sciences*, 115(40), 9859–9866. https://doi.org/10.1073/pnas.1611621115

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. https://doi.org/10.1038/87502

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. https://doi.org/10.1162/jocn.2006.18.2.199

Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences*, *115*(51), 13129–13134. https://doi.org/10.1073/pnas.1811793115

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*, 10(3), 241–265. https://doi.org/10.1207/s15327078in1003_3

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, 106(3), 1367–1381. https://doi.org/10.1016/j.cognition.2007.07.002

McDermott, J., & Hauser, M. (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Perception*, *23*(1), 29–59. https://doi.org/10.1525/mp.2005.23.1.29

McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, *21*(3), 289–311. https://doi.org/10.1525/mp.2004.21.3.289

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29(2), 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., ... & Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*(6468), eaax0868. https://doi.org/10.1126/science.aax0868

Meisel, J. M. (2013). *Language Acquisition and Change: A Morphosyntactic Perspective*. Edinburgh University Press.

Meyer, L. B. (1956). *Emotion and Meaning in Music* (Paperback ed., [Nachdr.]). University of Chicago Press.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, *46*(6), 505–512. https://doi.org/10.3758/BF03208147

Mithen, S. J. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. Weidenfeld & Nicolson.

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130299. https://doi.org/10.1098/rstb.2013.0299

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011a). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, 22(11), 1425–1433. https://doi.org/10.1177/0956797611416999

Moreno, S., & Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing Research*, 308, 84–97. https://doi.org/10.1016/j.heares.2013.09.012A

Moreno, S., Friesen, D., & Bialystok, E. (2011b). Effect of music training on promoting preliteracy skills: Preliminary causal evidence. *Music Perception*, 29(2), 165–172. https://doi.org/10.1525/mp.2011.29.2.165

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, 19(3), 712–723. https://doi.org/10.1093/cercor/bhn120A

Murphy, G. L. (1996). On metaphoric representation. *Cognition*, *60*(2), 173–204. https://doi.org/10.1016/0010-0277(96)00711-1

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. https://doi.org/10.1073/pnas.0701498104

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*(2), 357–374.

Naigles, L. R., & Swensen, L. D. (2007). Syntactic supports for word learning. In E. Hoff & M. Shatz (Eds.), *Blackwell Handbook of Language Development* (pp. 212–231). Blackwell Publishing. https://doi.org/10.1002/9780470757833.ch11

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. University of Chicago Press.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 756.

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, 41(1), 233–243. https://doi.org/10.1016/S0167-6393(02)00106-1

Nettl, B. (1999). An ethnomusicologist contemplates universals in musical sound and musical culture. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 463–472). The MIT Press. https://doi.org/10.7551/mitpress/5190.003.0032

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*(2), 151–165. https://doi.org/10.1162/jocn.1991.3.2.151

Neville, H. J., Coffey, S. A., Lawson, D. S., Fischer, A., Emmorey, K., & Bellugi, U. (1997). Neural systems mediating American Sign Language: Effects of sensory experience and age of acquisition. *Brain and Language*, *57*(3), 285–308. https://doi.org/10.1006/brln.1997.1739

Newport, E. (2006). Language development, critical periods in. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science* (p. s00506). John Wiley & Sons, Ltd.https://doi.org/10.1002/0470018860.s00506

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, 14(1), 11–28. https://doi.org/10.1016/0364-0213(90)90024-Q

Niedeggen, M., & Rösler, F. (1999). N400 effects reflect activation spread during retrieval of arithmetic facts. *Psychological Science*, *10*(3), 271–276. https://doi.org/10.1111/1467-9280.00149

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*(6), 1281–1296. https://doi.org/10.1016/j.neuron.2015.11.035

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, *112*(1), 181–186. https://doi.org/10.1016/j.cognition.2009.04.001

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*, 261–283. https://doi.org/10.1007/BF01067377

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681. https://doi.org/10.1038/nn1082

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00142

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*(6), 717–733. https://doi.org/10.1162/089892998563121

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. https://doi.org/10.1080/02687030701803804

Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language*, *61*(1), 123–144. https://doi.org/10.1006/brln.1997.1862

Paulmann, S., & Uskul, A. K. (2014). Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition and Emotion*, *28*(2), 230–244. https://doi.org/10.1080/02699931.2013.812033

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00320

Penfield, W. (1965). Conditioning the uncommitted cortex for language learning. *Brain*, 88(4), 787–798.

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., ... & Mehler, J. (1996). Brain processing of native and foreign languages. *NeuroReport – International Journal for Rapid Communications of Research in Neuroscience*, 7(15), 2439–2444.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, *7*(8), 362–367. https://doi.org/10.1016/S1364-6613(03)00150-5

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, *117*(6), 1283–1301. https://doi.org/10.1093/brain/117.6.1283

Peretz, I., Saffran, J., Schön, D., & Gosselin, N. (2012). Statistical learning of speech, not music, in congenital amusia. *Annals of the New York Academy of Sciences*, 1252(1), 361–366. https://doi.org/10.1111/j.1749-6632.2011.06429.x

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00227

Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychonomic Bulletin & Review*, *20*(2), 310–317. https://doi.org/10.3758/s13423-012-0344-5

Pichon, S., & Kell, C. A. (2013). Affective and sensorimotor components of emotional prosody generation. *The Journal of Neuroscience*, *33*(4), 1640–1650. https://doi.org/10.1523/JNEUROSCI.3530-12.2013

Pinker, S. (1997). Words and rules in the human brain. *Nature*, 387(6633), 547–548. https://doi.org/10.1038/42347

Pinker, S. (1999). *How the Mind Works* (1. publ). Norton.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, *95*(2), 201–236. https://doi.org/10.1016/j.cognition.2004.08.004

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crèpin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 28(3), 191–212. https://doi.org/10.1016/0028-2243(88)90030-5

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3–34.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, 8(2), 162–172. https://doi.org/10.1111/j.1467-7687.2005.00403.x

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906–914. https://doi.org/10.1002/wcs.78

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, 81(2), 149–169. https://doi.org/10.1016/S0010-0277(01)00132-9

Saffran, J. R. (2003). Musical learning and language development. *Annals of the New York Academy of Sciences*, 999(1), 397–401. https://doi.org/10.1196/annals.1284.050

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

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. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, 69, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*(3), 235–248. https://doi.org/10.1016/S0892-1997(05)80231-0

Shepherd, J., & Wicke, P. (1997). *Music and Cultural Theory*. Polity Press.

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, *55*(2), 167–177. https://doi.org/10.1016/j.jml.2006.03.002

Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin & Review*, *28*(4), 1390–1398. https://doi.org/10.3758/s13423-021-01883-3

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

Silvén, M., Voeten, M., Kouvo, A., & Lundén, M. (2014). Speech perception and vocabulary growth: A longitudinal study of Finnish-Russian bilinguals and Finnish monolinguals from infancy to three years. *International Journal of Behavioral Development*, 38(4), 323–332. https://doi.org/10.1177/0165025414533748

Slevc, L. R., & Patel, A. D. (2011). Meaning in music and language: Three key differences. *Physics of Life Reviews*, *8*(2), 110–111. https://doi.org/10.1016/j.plrev.2011.05.003

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. https://doi.org/10.3758/16.2.374

Smit, E. A., Milne, A. J., Sarvasy, H. S., & Dean, R. T. (2022). Emotional responses in Papua New Guinea show negligible evidence for a universal effect of major versus minor music. *PLoS ONE*, *17*(6), e0269597. https://doi.org/10.1371/journal.pone.0269597

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*(6876), 87–90. https://doi.org/10.1038/416087a

Sohail, J., & Johnson, E. K. (2016). How transitional probabilities and the edge effect contribute to listeners' phonological bootstrapping success. *Language Learning and Development*, *12*(2), 105–115. https://doi.org/10.1080/15475441.2015.1073153

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, *18*(8), 1380–1393. https://doi.org/10.1162/jocn.2006.18.8.1380

Strait, D. L., & Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Frontiers in Psychology*, 2, 113. https://doi.org/10.3389/fpsyg.2011.00113

Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency – effects of training on subcortical processing of vocal expressions of emotion. *European Journal of Neuroscience*, 29(3), 661–668. https://doi.org/10.1111/j.1460-9568.2009.06617.x

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language*, *52*(3), 452–473. https://doi.org/10.1006/brln.1996.0024

Swaminathan, S., & Schellenberg, E. G. (2019). Music training and cognitive abilities: Associations, causes, and consequences. In M. H. Thaut & D. A. Hodges (Eds.), *The Oxford Handbook of Music and the Brain* (pp. 644–670). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198804123.013.26

Swaminathan, S., & Schellenberg, E. G. (2020). Musical ability, music training, and language ability in childhood. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(12), 2340–2348. https://doi.org/10.1037/xlm0000798

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, 50(1), 86–132. https://doi.org/10.1016/j.cogpsych.2004.06.001

Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Reviews*, *43*(3), 231–246. https://doi.org/10.1016/j.brainresrev.2003.08.004

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, *109*(46), 19027–19032. https://doi.org/10.1073/pnas.1210344109

Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2013). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex*, *23*(2), 249–254. https://doi.org/10.1093/cercor/bhs003

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, 107(4), 885. https://doi.org/10.1037/0033-295X.107.4.885

Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research*, *16*(2), 145–161. https://doi.org/10.1016/S0926-6410(02)00245-8

Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language*, *162*, 60–71. https://doi.org/10.1016/j.bandl.2016.08.004

Tsang, C. D., & Conrad, N. J. (2011). Music training and reading readiness. *Music Perception*, 29(2), 157–163. https://doi.org/10.1525/mp.2011.29.2.157

Turk, M., Turk, I., & Otte, M. (2020). The neanderthal musical instrument from Divje Babe I Cave (Slovenia): A critical review of the discussion. *Applied Sciences*, *10*(4), 1226. https://doi.org/10.3390/app10041226

Uddin, S., Heald, S. L. M., Van Hedger, S. C., & Nusbaum, H. C. (2018). Hearing sounds as words: Neural responses to environmental sounds in the context of fluent speech. *Brain and Language*, *179*, 51–61. https://doi.org/10.1016/j.bandl.2018.02.004

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2019). Absolute pitch can be learned by some adults. *PLoS ONE*, *14*(9), e0223047. https://doi.org/10.1371/journal.pone.0223047

Van Hedger, S. C., Johnsrude, I. S., & Batterink, L. J. (2022). Musical instrument familiarity affects statistical learning of tone sequences. *Cognition*, *218*, 104949. https://doi.org/10.1016/j.cognition.2021.104949

Vasuki, P. R. M., Sharma, M., Demuth, K., & Arciuli, J. (2016). Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research*, 342, 112–123. https://doi.org/10.1016/j.heares.2016.10.008

Vasuki, P. R. M., Sharma, M., Ibrahim, R., & Arciuli, J. (2017). Statistical learning and auditory processing in children with music training: An ERP study. *Clinical Neurophysiology*, 128(7), 1270–1281. https://doi.org/10.1016/j.clinph.2017.04.010

Wacewicz, S., & Żywiczyński, P. (2015). Language evolution: Why Hockett's design features are a non-starter. *Biosemiotics*, *8*(1), 29–46. https://doi.org/10.1007/s12304-014-9203-2

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, *21*(1), 21–25. https://doi.org/10.1177/0956797609354734

Weiss, M. W., & Bidelman, G. M. (2015). Listening to the brainstem: Musicianship enhances intelligibility of subcortical representations for speech. *Journal of Neuroscience*, 35(4), 1687–1691. https://doi.org/10.1523/JNEUROSCI.3680-14.2015

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

White, E. J., Hutka, S. A., Williams, L. J., & Moreno, S. (2013). Learning, neural plasticity and sensitive periods: Implications for language acquisition, music training and transfer across the lifespan. *Frontiers in Systems Neuroscience*, 7, 90. https://doi.org/10.3389/fnsys.2013.00090

Wong, Y. K., Lui, K. F. H., Yip, K. H. M., & Wong, A. C.-N. (2020). Is it impossible to acquire absolute pitch in adulthood? *Attention, Perception, & Psychophysics*, *82*(3), 1407–1430. https://doi.org/10.3758/s13414-019-01869-3

Yetkin, O., Yetkin, F. Z., Haughton, V. M., & Cox, R. W. (1996). Use of functional MR to map language in multilingual volunteers. *American Journal of Neuroradiology*, 17(3), 473–477.

Yumoto, M., Uno, A., Itoh, K., Karino, S., Saitoh, O., Kaneko, Y., … & Kaga, K. (2005). Audiovisual phonological mismatch produces early negativity in auditory cortex. *Neuroreport*, 16(8), 803–806.

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLOS Biology*, *10*(7), e1001372. https://doi.org/10.1371/journal.pbio.1001372

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. https://doi.org/10.1016/S1364-6613(00)01816-7

Zatorre, R. J., Evans, A., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*(5058), 846–849. https://doi.org/10.1126/science.1589767

Zhao, T. C., & Kuhl, P. K. (2016). Musical intervention enhances infants' neural processing of temporal structure in music and speech. *Proceedings of the National Academy of Sciences*, 113(19), 5212–5217. https://doi.org/10.1073/pnas.1603984113

Zhao, T. C., Llanos, F., Chandrasekaran, B., & Kuhl, P. K. (2022). Language experience during the sensitive period narrows infants' sensory encoding of lexical tones: Music intervention reverses it. *Frontiers in Human Neuroscience*, 16, 941853. https://doi.org/10.3389/fnhum.2022.941853