



## Review

**Cite this article:** Peretz I, Vuvan D, Lagrois M-É, Armony JL. 2015 Neural overlap in processing music and speech. *Phil. Trans. R. Soc. B* **370**: 20140090.  
<http://dx.doi.org/10.1098/rstb.2014.0090>

One contribution of 12 to a theme issue 'Biology, cognition and origins of musicality'.

**Subject Areas:**  
neuroscience

**Keywords:**  
neural overlap, music, speech, fMRI

### Author for correspondence:

Isabelle Peretz  
e-mail: [isabelle.peretz@umontreal.ca](mailto:isabelle.peretz@umontreal.ca)

# Neural overlap in processing music and speech

Isabelle Peretz<sup>1,2</sup>, Dominique Vuvan<sup>1,2</sup>, Marie-Élaine Lagrois<sup>1,2</sup>  
and Jorge L. Armony<sup>1,3</sup>

<sup>1</sup>International Laboratory of Brain, Music and Sound Research (BRAMS), and Center for Research on Brain, Language and Music (CRBLM)

<sup>2</sup>Department of Psychology, University of Montreal, Montreal, Quebec, Canada

<sup>3</sup>Department of Psychiatry, McGill University and Douglas Mental Health University Institute, Montreal, Quebec, Canada

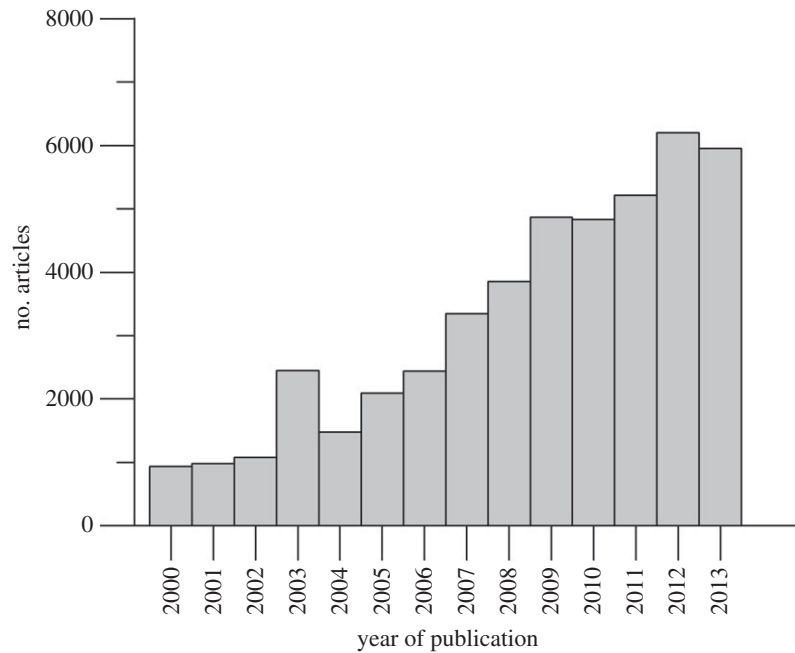
Neural overlap in processing music and speech, as measured by the co-activation of brain regions in neuroimaging studies, may suggest that parts of the neural circuitries established for language may have been recycled during evolution for musicality, or vice versa that musicality served as a springboard for language emergence. Such a perspective has important implications for several topics of general interest besides evolutionary origins. For instance, neural overlap is an important premise for the possibility of music training to influence language acquisition and literacy. However, neural overlap in processing music and speech does not entail sharing neural circuitries. Neural separability between music and speech may occur in overlapping brain regions. In this paper, we review the evidence and outline the issues faced in interpreting such neural data, and argue that converging evidence from several methodologies is needed before neural overlap is taken as evidence of sharing.

## 1. Introduction

Humans are born with the potential both to speak and make music. The relation between language and musicality<sup>1</sup> has been the topic of much debate. The debate is heated because it speaks directly to the nature of evolved human cognition. For some, musicality owes its efficacy to the natural disposition for speech. For example, music may exaggerate particular speech features such as intonation and affective tone that are so effective for bonding [2]. In other words, musicality may aim at the language system just as artistic masks target the face recognition system. We can stretch this argument further and envisage that music owes its efficacy by relying on the natural disposition for speech. From this perspective, the language modules are invaded [3]. Musicality could have emerged in all cultures because it is so effective at co-opting one or several evolved modules. For others [4], once we take away the tone of voice shared by speech and music, the specialization of language to convey conceptual information and of musicality to express affect is distinct. According to this view, musicality may have preceded language in evolution, and language may build on the natural disposition for musicality.

These divergent perspectives on the origins and functions of musicality are echoed in cognitive neuroscience. On the one hand, an increasing number of neuroimaging studies point to a large and significant neural overlap in the responses to vocal and musical stimuli, taken as evidence of neural sharing between music and speech processing (figure 1). On the other hand, a solid body of neuropsychological studies shows that musicality involves multiple processing components that can be selectively impaired without apparent effects on language (or any other cognitive ability [5,6]). These conflicting sets of data call for novel neurocomparative studies.

The neurocomparative study of music and speech processing is obviously limited in animals and must therefore rely on the use of sophisticated and



**Figure 1.** Results of a Google Scholar search. The search was performed on 11 July 2014 using the keywords '(neural) AND (overlap OR sharing) AND (music) AND (language OR speech)', with the number of search results plotted per year (non-cumulative).

non-invasive methods in humans. The most widely used technique today is functional magnetic resonance imaging (fMRI), which has been recently exploited to compare music and speech perception. As we explain below, co-activation of brain regions in fMRI is often interpreted as evidence of the sharing of the underlying neural circuitry. However, activation overlap does not provide us with sufficient evidence of neural sharing, for which rigorous and direct testing is required. In this review, we summarize contemporary views of brain organization for music and speech processing with a discussion of the challenges involved in concluding neural sharing (or separation) from evidence of overlap. We then review some methodological advances that aid in making more definitive conclusions regarding neural overlap. Finally, we discuss how evidence of neural sharing can impact current views of the neurobiology of musicality.

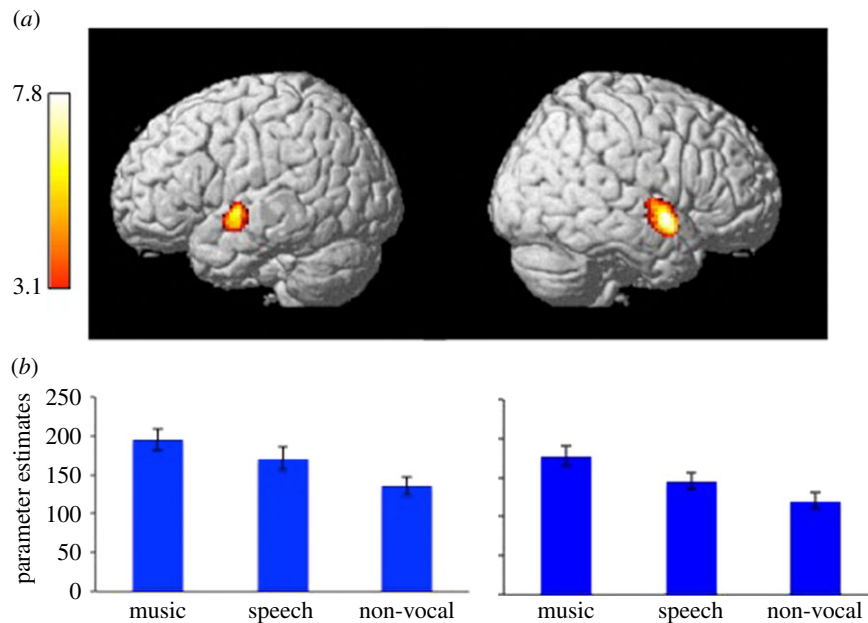
## 2. Brain specialization: from regions to networks

Research in cognitive neuroscience has been guided by the assumption that brain regions are specialized for a function. Each specialized function would be implemented in a relatively small neural space. For example, the superior temporal sulcus has been associated with voice processing [7]. This voice-preferred region responds, bilaterally, more strongly to human vocalizations, with or without linguistic content, than to non-vocal sounds or vocalizations produced by other animals [8]. Such a specialized neural system to process conspecific vocalizations has been observed in other species. For instance, the superior temporal plane of the macaque monkey responds preferentially to species-specific vocalizations over other vocalizations and sounds [9]. Other cortical areas, such as the caudal insular cortex in rhesus monkeys, also appear to be tuned to intraspecies vocalizations over a wide range of auditory stimuli such as environmental sounds and vocalizations from other animals [10].

Similarly, music processing may rely on a cortical area that is domain-specific and neurally separable [5]. For example, the

system that maps pitch onto musical keys, termed tonality or tonal encoding of pitch, may be music-selective. Current research points to the inferior frontal areas as critically involved [11–13]. However, this localization mostly corresponds to the processing of harmonic structure, a culture-specific elaboration of pitch that is quite recent in music history. Moreover, tonal encoding of pitch is likely to recruit a vast network because it involves multiple processes. For example, Jackendoff & Lerdahl [14] distinguish three different forms of elaboration of pitch hierarchies in a musical context, by considering different principles for pitch space, tonal reduction and tension/relaxation. Thus, it would not be surprising to discover that more than one brain region contribute to the musical interpretation of pitch. A major breakthrough would be to identify one of these brain regions as foundational for tonal encoding of pitch [5]. So far, such an essential component for musicality has not been localized in one specific region. Rather, the evidence points towards the connectivity between the right auditory cortex and inferior frontal gyrus (IFG) as being necessary for normal development of musical abilities [15,16].

The key question becomes to what extent parts of the musicality network can be shared both functionally and neurally with language. Logically, music and speech processing could share parts of their respective neural networks, such as the mechanisms for the acoustical analysis of pitch, and still be distinct, because musicality and language differ from one another in other respects, notably semantics. The idea that parts of the networks are shared is currently very popular. A Google Scholar search using the keywords '(neural) AND (overlap OR sharing) AND (music) AND (language OR speech)' reveals a linear increase of this notion in published articles over the past decade (figure 1). This explosion in research interest in music and language overlap is, however, often unsupported by comparative data. To cite a recent example, an fMRI study reporting stronger response to rhythmic musical structure in professional musicians, compared with non-musicians, in a region typically associated with processing of linguistic syntax (which was not examined in that study), led the authors to conclude that 'musical experts seem to rely on the same



**Figure 2.** (a) Statistical map showing voxels within the temporal lobe with stronger responses to musical stimuli than to human voice (including speech). (b) Left and right cluster-averaged parameter estimates, showing that responses to speech were smaller than to music but were significantly larger compared with human non-vocal stimuli (all pairwise comparisons significant,  $p < 0.005$ ). Data from Angulo-Perkins [22]. (Online version in colour.)

neural resources during the processing of syntactic violations in the rhythm domain that the brain usually uses for the evaluation of linguistic syntax' [17].

It is important to keep in mind that neural overlap does not necessarily entail neural sharing. The neural circuits established for musicality may be intermingled or adjacent to those used for a similar function in language and yet be neurally separable. For example, mirror neurons are interspersed among purely motor-related neurons in pre-motor regions of the macaque cortex [18]. Similarly, the neurons responsible for the computation of some musical feature may be interspersed among neurons involved in similar aspects in speech.

Moreover, most brain structures, such as Broca's area (occupying the left IFG), which is often the focus of interest in music and language comparisons, are relatively large and complex, and thus can easily accommodate more than one distinct processing network. Common localization of distinct networks or functions can be dictated by neural properties, like dense connectivity with other brain regions, and not dictated by sharing.

Recent network analyses have revealed highly connected network 'hubs', which may well be shared by music and speech processing. Hubs support efficient control or integration by facilitating the convergence of neuronal signals from different sensory modalities (e.g. auditory and motor) or cognitive domains (e.g. musicality and language) [19,20]. The hubs maintain anatomical and functional connections that span long distances and they tend to consume metabolic energy at a higher rate than non-hub regions [21]. Thus, hubs are not only centres of integration, but also points of increased haemodynamic responses and vulnerability. Accordingly, the co-activation of brain regions, as well as the sometimes observed co-occurrence of deficits in music and speech processing, may reflect the involvement of these integration centres rather than of distinct parts of their respective networks.

In this network perspective, how can we identify the parts of networks that would constitute evidence of 'brain specialization', as the subtitle of this section suggests? Several fMRI

studies have identified a region within the anterior superior temporal gyrus (STG) that responds more strongly to music than to human voice, including speech [22–25] (figure 2). Nonetheless, these studies also found large regions within the temporal lobes that responded more to both music and voice, compared with control conditions (e.g. non-vocal sounds or silence), with no significant differences between the two former categories. Moreover, even in regions that were categorized as 'music-preferred', significant activation in response to speech (compared, for instance, with non-vocal sounds) can be observed (figure 2b). Similarly, the so-called language-selective regions in the left posterior temporal lobe show a significant, albeit weaker, response to music [26]. Unfortunately, the approach employed in these studies does not allow researchers to determine whether the same neuronal populations respond to both music and speech (possibly with different strength), or if distinct, neighbouring groups of neurons were involved.

Fortunately, new methods have been developed to separate brain responses to different categories of stimuli in the same neural region. These techniques can be exploited to distinguish domain-specific neural activation at a finer-grained level than the standard use of fMRI can offer. These techniques will be introduced in §3.

### 3. Evidence of neural sharing

Most regions of the brain participate in multiple processes [27]. Moreover, music and speech processing share a large number of properties, from the acoustical analysis of the auditory input to the planning of motor output. Therefore, several brain areas are expected to overlap in the processing of music and speech [28]. However, as previously emphasized, neural overlap does not necessarily mean neural sharing. Here, we review the evidence for and against neural sharing; specifically, we summarize the studies in which the distinct contribution of neural populations to music and to speech processing has been examined in overlapping regions. This can be done by a

number of neuroimaging techniques, such as multi-voxel pattern analysis and fMRI adaptation, and more invasively by intracerebral recordings. The main results obtained with each technique are presented below.

Note that we chose not to cover lesion studies here because these have been reviewed relatively recently [5,6], and the use of this approach has declined over recent years. Nevertheless, it is worth mentioning here that lesion studies suggest neural segregation between musicality and language networks. Otherwise, brain damage could not affect just musical abilities while sparing language and other aspects of cognition [5,6,29]. The possibility of disrupting the operations of musicality without having any impact on language implies that some parts of its neural substrate are not shared with language.

### (a) Multi-voxel pattern analysis

This technique uses machine learning algorithms to categorize neuroimaging data. It differs from the standard approach that assesses, on a voxel-by-voxel basis, the mean difference in activity between conditions and identifies voxels that respond consistently more strongly to one condition (e.g. music) than to another (e.g. speech) [30]. Unlike such standard univariate analyses, multivariate decoding methods consider data from several voxels at once to identify patterns of activity associated with a particular stimulus, task or 'mental state'. The multivariate pattern analysis methods detect distributed representations (i.e. groups of voxels) whose combined activity discriminates between conditions of interest, even if the individual voxels do not exhibit statistically significant (i.e. in the context of the general linear model) differences between them [31]. One key advantage of this multivariate approach, which provides complementary information to that obtained through univariate methods [32,33], is its sensitivity to neural segregation in overlapping regions [34,35]. This technique has been used in several studies comparing music with speech.

As mentioned above, Rogalsky *et al.* [25] found large areas of activation in response to music and speech in overlapping portions of auditory cortex. However, multivariate pattern classification analyses indicated that within the regions of overlap, speech and music elicited distinguishable patterns of activation in the STGs. In addition, speech (jabberwocky sentences) elicited more ventrolateral activation, whereas music (novel melodies) elicited a more dorsomedial pattern extending into the parietal lobe. These findings highlight the existence of overlapping but distinct networks for music and speech within the same cortical areas. Similarly, Abrams *et al.* [36] used multivariate pattern analysis for natural and scrambled music and speech excerpts and also found distinct brain patterns of responses to the two categories of sounds in several regions within the temporal lobe and the inferior frontal cortex. Therefore, the pattern of neural activation was distinct between music and speech, although there was overlap in the areas activated by the two domains.

It is important to point out that, even if the stimuli are matched for emotional content, attention, memory, subjective interest, arousal and familiarity [36], any observed category differences in activation strengths and/or patterns could be owing to acoustical differences. In order to avoid this confound, at least to some extent, one can use sung melodies and spoken lyrics from songs. These are optimal stimuli because they are relatively complex and extend over several seconds. Furthermore, the coexistence of tunes and lyrics in

songs makes them very similar in terms of acoustical structure and familiarity. The main acoustic differences between song and speech are the more regular rhythm and pitch stability in each syllable of songs.

Comparing spoken lyrics, sung tunes and songs (corrected for rhythmic differences) with both multivariate and univariate analyses, Merrill *et al.* [37] observed a large overlap between song and speech in the bilateral STGs. The STGs were found to code for differences between words and pitch patterns whether these were embedded in a song or in speech. They also found that the left IFG coded for spoken words and showed predominance over the right IFG in prosody processing, whereas the right IFG was more active for processing the pitch pattern in songs. Interestingly, this result was only found when using the multivariate decoding method, demonstrating its higher sensitivity to the differential fine-scale coding of information. Another important result is the finding that the intraparietal sulcus shows sensitivity to discrete pitch relations in songs as opposed to the gliding pitches in speech prosody. Thus, as expected, the processing of lyrics, tunes and songs shares many features that are reflected in a fundamental similarity of brain areas involved in their perception. However, subtle differences between speech and music can lead to distinct patterns of brain activity.

In sum, music and speech stimuli seem to activate distinct neural populations in overlapping regions. However, most of the reported results could be owing, at least in part, to acoustical differences between categories. One way to circumvent this potential problem is to parametrically manipulate acoustical structure. For example, musical sounds can be 'morphed' into speech sounds gradually, so that the influence of acoustical changes on the brain responses can be measured systematically. So far, white noise has been morphed into a speech sound or a musical instrument sound separately [38], but the morphing technique can be extended to whole sentences and be used with other paradigms, like adaptation, to which we now turn.

### (b) Functional magnetic resonance imaging adaptation

Although the nature of the blood oxygen-level-dependent (BOLD) signal and its spatial resolution do not allow for directly identifying which neurons are active in response to a given stimulus, we can take advantage of the nonlinear dynamics of neural activity to indirectly address this question, by using the so-called fMRI adaptation paradigm [39]. This approach, based on the principle of neuronal adaptation/habituation, relies on the fact that the observed BOLD signal to successive stimuli depends on whether they stimulate the same or different neurons. That is, the activity associated with two stimuli will be smaller if they activate the same neuronal pool than if they stimulate different neurons. Although adaptation is strongest when repeating the same stimulus, it can also be observed when different exemplars from the same category are presented, and can thus be used to identify those brain regions in which different types of stimuli share a common neural representation.

Thus, in a region that responds to both speech and music, we would expect within-domain (i.e. speech–speech and music–music) adaptation. If the attenuation remains when a switch in domain is introduced (i.e. music–speech or speech–music), it would suggest that the same neuronal



population was responding to both categories. In contrast, if the attenuation disappears in the overlapping region after the switch in domains, this would be evidence that ‘fresh’ neurons from a distinct neural population were responsible for the processing of each of the two domains.

The paradigm of fMRI adaptation has been exploited twice in the neural comparison of music and speech. Sammler *et al.* [40] used this technique to induce neural adaptation to listening to lyrics, melodies and songs. Reductions of the BOLD response were observed along the superior temporal sulcus and gyrus (STS/STG) bilaterally. Within these regions, the left mid-STS showed an interaction of the adaptation effects for lyrics and tunes, suggesting shared processing of the two components. The degree of integration decayed towards more anterior regions of the left STS in which the stronger adaptation for lyrics than for tunes was suggestive of independent processing of lyrics. Evidence for an integrated representation of lyrics and tunes was also found in the left pre-motor cortex, possibly related to the build-up of a vocal code for singing.

However, in a subsidiary analysis of a recent study [41], we showed music–music, but no speech–music adaptation, in the ‘music-preferred’ area in the anterior STG [42]. This preliminary result suggests that distinct neural populations underlie the activations observed in this region for speech and music.

Altogether, results show both a neural dissociation as well as a high degree of neural sharing between music and speech, which could arise from the involvement of voice-specific areas. Further studies exploiting this paradigm represent an opportunity for characterizing the nature of the shared mechanisms.

One optimal avenue for future adaptation studies is provided by the song illusion discovered by Deutsch [43]. This illusion is created by the repetition of a phrase that sounds initially like speech and through repetition, as if it were sung. One possible account of this illusion is that the neurons underlying speech perception get adapted through repetition, whereas the neural population underlying music perception does not. The robustness of music perception to neural attenuation may originate from the fact that repetition is a characterizing feature of music and not of speech [44].

The only fMRI study that has tested the neural correlates of the song illusion used different phrases, albeit produced by the same speaker and matched for syllable length, in the spoken and sung condition [45]. Using such stimuli, it was found that BOLD responses were larger for speech perceived as sung than spoken in multiple brain regions, including the anterior STG and the right midposterior STG. There was no area more responsive to speech than to song. Although these results are compatible with a distinct musical contribution to the illusion, the use of an adaptation paradigm would provide more compelling evidence.

The use of adaptation procedures with fMRI should be pursued not only in passive listening tasks like the studies described above but also in active tasks. However, like any paradigm, it has its limits. Reduced activity can result from adaptation, but also from practice or expectation. Similarly, if adaptation disappears after a change of condition, say from speech to music, this could be because speech-induced neural changes interfered with music performance (see [46] for a discussion of various ways to interpret brain response adaptation).

### (c) Intracranial recordings

The implantation of electrodes for pre-surgical evaluation of temporal lobe epilepsy represents a rare chance to distinguish neural responses to music and speech with excellent temporal and spatial resolution that largely exceeds that achieved with non-invasive methods. This high spatio-temporal resolution can address the question of shared versus distinct neural populations by revealing the degree to which the time course of neural responses differ in regions in which music and speech processing overlap. So far, depth electrodes have not been used in the comparison of music and speech.

Electrical activity has been recorded intracranially through subdural electrodes located above the left or right perisylvian region [47]. This method presents the advantage of recording temporal activity with high precision. However, the method still faces the difficulty of inferring the location of the sources on the basis of brain surface recordings, especially in overlapping regions.

Using this method to examine the contribution of the STGs to the linguistic and musical syntax, Sammler *et al.* [47] compared the early negativities evoked by violations of structure in sentences and chord sequences in five patients. The results showed considerable overlap in the bilateral STG, but also differences in the hemispheric timing and relative involvement of the frontal and temporal brain structures. While the combined data lend support for a co-localization of early musical and linguistic syntax processing in the temporal lobe, the mechanisms involved seem to depend on the (music or language) domain considered.

### (d) Future directions

The existing evidence points towards substantial neural overlap between music and speech processing. A many-to-one mapping between cognitive functions and brain structures seems to characterize the human brain [27]. Therefore, it is more likely to find evidence of overlap than segregation. Nevertheless, there is converging evidence for music-specific responses along the neural pathways. The evidence is still scarce but strengthened by the diversity of the neuroimaging approaches used so far. Therefore, the question of overlap between music and speech processing must still be considered as an open question for the field. In this paper, we have reviewed technological advances that will allow us to tackle this issue more rigorously than in the past.

While using the novel fMRI techniques for cross-domain comparison, it may be useful to consider the following recommendations:

- (1) Make comparisons in native (subject-specific) brain space, rather in a common (normalized) one. Averaging across individually variable anatomies blurs brain activations and can create an artificial overlap of closely neighbouring but non-overlapping responses. For example, location of pitch maps in Heschl’s gyri and planum temporale varies widely across individuals [48,49]. Such variable organization and localization of cortical maps call for individually determined regions of interest.
- (2) Consider connectivity. The human brain is a highly connected and interactive structure. Domain specificity at high levels can impact low-level processing. That is, lower-level brain areas, such as the brain stem and the

primary auditory cortex, are influenced by higher-level areas via efferent connections [50].

- (3) Manipulate stimulus/task parameters. Cognitive demands made by supposedly analogous operations in processing music and speech may widely differ. In order to avoid differences in brain responses driven by task difficulty rather than by domain specificity, one can manipulate a common factor (e.g. speed) or use an interference paradigm as is often the case in behavioural studies. In doing so, the perceptual demands should always be carefully controlled in order to avoid differences in brain responses driven by purely acoustic differences. As mentioned in §3b, Deutsch's song illusion presents an ideal example of this type of control, because the input remains the same, whereas the percept changes through repetition.

## 4. Implications and conclusion

Neural sharing is a key concept for explaining transfer effects between music and language. Patel [28,51] has introduced the OPERA framework to explain why musical training may lead to enhanced speech processing. An essential condition of the OPERA hypothesis is neural overlap, a term used by Patel to refer to 'neural sharing'. That is, in order for musical training to influence the neural processing of speech, a shared characteristic in both domains must be processed by a population of neurons shared by the musicality and language brain networks.

The original OPERA hypothesis [51] focuses on *acoustic* features (e.g. waveform periodicity) rather than on *cognitive* demands (e.g. auditory working memory). This early focus on acoustical features converges with the evidence reviewed here suggesting that the processing of music and speech overlaps in posterior auditory cortex (basic acoustic processing) and becomes more differentiated anteriorly (domain-specific representation) [24,40]. However, the expanded OPERA hypothesis [28] incorporates the idea of shared *cognitive* processing into the discussion of neural overlap, based on the proposals that musical training enhances auditory attention and working memory [52,53]. Indeed, there is abundant evidence that attention sharpens sensory encoding in primary brain areas (see [54] for a meta-analysis of attention-related modulations in the auditory cortex). However, little is known about the neural specificity of these top-down effects. Backward propagation along neural pathways may be diffuse and affect several distinct sensory systems, not just speech. In that case, neural sharing is not necessary for transfer between musicality and language. Therefore, as research guided by the OPERA hypothesis proceeds, it is important to consider methods that are sufficiently

sophisticated to make justifiable claims regarding the neural sharing between music and speech processing.

The systematic search for neural sharing between music and speech is an important avenue not only for clinical and education purposes, but also for the understanding of the neurobiological origins of musicality. It is common for neural circuits established for one purpose to be re-used or recycled during evolution. For example, in the zebra finch, some song nuclei may participate in the learning of non-vocal tasks such as food avoidance. These findings suggest that the specialized forebrain pre-motor nuclei controlling song evolved from circuits involved in behaviours related to feeding [55]. In humans, we have examined the possibility that musicality recycles emotion circuits that have evolved for emotional vocalizations [56].

Dehaene & Cohen [57] have proposed an interesting 'recycling' proposal because it entails neural and functional constraints on the newly acquired function. The recycling may only occur if a network of neural structures already has (most of) the structures necessary to support the novel set of cognitive and physical procedures that characterize the new function. As a result, the neural manifestations of novel abilities should have some common characteristics and share some possibilities for learning with non-human primates. This theory makes clear some of the limits and costs of neuronal recycling. The greater the distance between the function(s) and the existing cortical structure, the harder the learning process will be, and the more likely that the learning process will disrupt the other functions that the common neural circuitry supports. Therefore, if some core component of musicality can be found to share a brain region or network involved in language, it may reveal a novel pathway by which humans may have achieved their highly sophisticated use of sound.

**Acknowledgements.** The authors thank the anonymous reviewers for their insightful comments.

**Author contributions.** I.P. conceived the idea and wrote the article. J.A., M.L. and D.V. contributed to the drafting of the manuscript.

**Funding statement.** I.P. is supported by a Canada Research Chair in neurocognition of music.

**Competing interests.** The authors report no competing interests.

## Endnote

<sup>1</sup>According to the proposal introduced in reference [1], we are distinguishing between musicality and music, as follows. Musicality is defined as a spontaneously developing set of traits based on, and constrained by, our cognitive and biological system. *Music* is defined as a social and cultural construct based on that musicality.

## References

- Honing H, ten Cate C, Peretz I, Trehub SE. 2015 Without it no music: cognition, biology and evolution of musicality. *Phil. Trans. R. Soc. B* **370**, 20140088. (doi:10.1098/rstb.2014.0088)
- Juslin PN. 2001 Communication emotion in music performance: a review and a theoretical framework. In *Music and emotion: theory and research* (eds PN Juslin, JA Sloboda), pp. 309–337. New York, NY: Oxford University Press.
- Sperber D, Hirschfeld LA. 2004 The cognitive foundations of cultural stability and diversity. *Trends Cogn. Sci.* **8**, 40–46. (doi:10.1016/j.tics.2003.11.002)
- Jackendoff R. 2009 Parallels and nonparallels between language and music. *Music Percept.* **26**, 195–204. (doi:10.1525/mp.2009.26.3.195)
- Peretz I. 2006 The nature of music from a biological perspective. *Cognition* **100**, 1–32. (doi:10.1016/j.cognition.2005.11.004)
- Peretz I. 2009 Music, language and modularity framed in action. *Psychol. Belg.* **49**, 157–175.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. 2000 Voice-selective areas in human auditory cortex. *Nature* **403**, 309–312. (doi:10.1038/35002078)
- Fecteau S, Armony JL, Joannette Y, Belin P. 2004 Is voice processing species-specific in human auditory cortex? An fMRI study. *NeuroImage* **23**, 840–848. (doi:10.1016/j.neuroimage.2004.09.019)

9. Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK. 2008 A voice region in the monkey brain. *Nat. Neurosci.* **11**, 367–374. (doi:10.1038/nn2043)
10. Remedios R, Logothetis NK, Kayser C. 2009 An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. *J. Neurosci.* **29**, 1034–1045. (doi:10.1523/jneurosci.4089-08.2009)
11. Koelsch S, Gunter TC, Zysset S, Lohmann G, Friederici AD. 2002 Bach speaks: a cortical 'language-network' serves the processing of music. *Neuroimage* **17**, 956–966. (doi:10.1006/nimg.2002.1154)
12. Tillmann B, Janata P, Bharucha JJ. 2003 Activation of the inferior frontal cortex in musical priming. *Cogn. Brain Res.* **16**, 155–161. (doi:10.1016/S0926-6410(02)00245-8)
13. Tillmann B, Koelsch S, Escoffier N, Bigand E, Lalitte P, Friederici A, Von Cramon D. 2006 Cognitive priming in sung and instrumental music: activation of inferior frontal cortex. *Neuroimage* **31**, 1771–1782. (doi:10.1016/j.neuroimage.2006.02.028)
14. Jackendoff R, Lerdahl F. 2006 The capacity for music: what is it, and what's special about it? *Cognition* **100**, 33–72. (doi:10.1016/j.cognition.2005.11.005)
15. Peretz I. 2013 The biological foundations of music: insights from congenital amusia. In *The psychology of music* (ed. D Deutsch), pp. 551–564. Waltham, MA: Academic Press.
16. Albouy P *et al.* 2013 Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain* **136**, 1639–1661. (doi:10.1093/brain/awt082)
17. Herdener M, Humbel T, Esposito F, Habermeyer B, Cattapan-Ludewig K, Seifritz E. 2014 Jazz drummers recruit language-specific areas for the processing of rhythmic structure. *Cereb. Cortex* **24**, 836–843. (doi:10.1093/cercor/bhs367)
18. Rizzolatti G, Craighero L. 2004 The mirror-neuron system. *Annu. Rev. Neurosci.* **27**, 169–192. (doi:10.1146/annurev.neuro.27.070203.144230)
19. van den Heuvel MP, Sporns O. 2013 Network hubs in the human brain. *Trends Cogn. Sci.* **17**, 683–696. (doi:10.1016/j.tics.2013.09.012)
20. Yan C, He Y. 2011 Driving and driven architectures of directed small-world human brain functional networks. *PLoS ONE* **6**, e23460. (doi:10.1371/journal.pone.0023460)
21. Collin G, Sporns O, Mandl RCW, van den Heuvel MP. 2013 Structural and functional aspects relating to cost and benefit of rich club organization in the human cerebral cortex. *Cereb. Cortex* **24**, 2258–2267. (doi:10.1093/cercor/bht064)
22. Angulo-Perkins A, Aubé W, Peretz I, Barrios F, Armony JL, Concha L. 2014 Music listening engages specific cortical regions within the temporal lobes: differences between musicians and nonmusicians. *Cortex* **59**, 126–137. (doi:10.1016/j.cortex.2014.07.013)
23. Fedorenko E, McDermott JH, Norman-Haignere S, Kanwisher N. 2012 Sensitivity to musical structure in the human brain. *J. Neurophysiol.* **108**, 3289–3300. (doi:10.1152/jn.00209.2012)
24. Leaver AM, Rauschecker JP. 2010 Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* **30**, 7604–7612. (doi:10.1523/jneurosci.0296-10.2010)
25. Rogalsky C, Rong F, Saberi K, Hickok G. 2011 Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J. Neurosci.* **31**, 3843–3852. (doi:10.1523/Jneurosci.4515-10.2011)
26. Fedorenko E, Behr MK, Kanwisher N. 2011 Functional specificity for high-level linguistic processing in the human brain. *Proc. Natl Acad. Sci. USA* **108**, 16 428–16 433. (doi:10.1073/pnas.1112937108)
27. Anderson ML. 2010 Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* **33**, 245–266. (doi:10.1017/S0140525X10000853)
28. Patel AD. 2014 Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear. Res.* **308**, 98–108. (doi:10.1016/j.heares.2013.08.011)
29. Peretz I, Coltheart M. 2003 Modularity of music processing. *Nat. Neurosci.* **6**, 688–691. (doi:10.1038/nn1083)
30. Friston KJ, Tononi G, Sporns O, Edelman G. 1995 Characterising the complexity of neuronal interactions. *Hum. Brain Mapp.* **3**, 312–324. (doi:10.1002/hbm.460030405)
31. Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006 Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* **10**, 424–430. (doi:10.1016/j.tics.2006.07.005)
32. Davis T, Poldrack RA. 2013 Measuring neural representations with fMRI: practices and pitfalls. *Ann. N.Y. Acad. Sci.* **1296**, 108–134. (doi:10.1111/nyas.12156)
33. Jimura K, Poldrack RA. 2012 Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia* **50**, 544–552. (doi:10.1016/j.neuropsychologia.2011.11.007)
34. Peelen MV, Downing PE. 2007 Using multi-voxel pattern analysis of fMRI data to interpret overlapping functional activations. *Trends Cogn. Sci.* **11**, 4–5. (doi:10.1016/j.tics.2006.10.009)
35. Todd MT, Nystrom LE, Cohen JD. 2013 Confounds in multivariate pattern analysis: theory and rule representation case study. *Neuroimage* **77**, 157–165. (doi:10.1016/j.neuroimage.2013.03.039)
36. Abrams DA, Bhatara A, Ryali S, Balaban E, Levitin DJ, Menon V. 2011 Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cereb. Cortex* **21**, 1507–1518. (doi:10.1093/cercor/bhq198)
37. Merrill J, Sammler D, Bangert M, Goldhahn D, Lohmann G, Turner R, Friederici AD. 2012 Perception of words and pitch patterns in song and speech. *Front. Psychol.* **3**, 76. (doi:10.3389/fpsyg.2012.00076)
38. Specht K, Osnes B, Hugdahl K. 2009 Detection of differential speech-specific processes in the temporal lobe using fMRI and a dynamic 'sound morphing' technique. *Hum. Brain Mapp.* **30**, 3436–3444. (doi:10.1002/hbm.20768)
39. Grill-Spector K, Malach R. 2001 fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol.* **107**, 293–321. (doi:10.1016/S0001-6918(01)00019-1)
40. Sammler D, Baird A, Valabrègue R, Clément S, Dupont S, Belin P, Samson S. 2010 The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. *J. Neurosci.* **30**, 3572–3578. (doi:10.1523/JNEUROSCI.2751-09.2010)
41. Aubé W, Angulo-Perkins A, Peretz I, Concha L, Armony JL. 2014 Fear across the senses: brain responses to music, vocalizations and facial expressions. *Soc. Cogn. Affect. Neurosci.* **9**, nsu067 (doi:10.1093/scan/nsu067)
42. Armony JL, Aubé W, Angulo-Perkins A, Peretz I, Concha L. Submitted. The specificity of neural responses to music and their relation to voice processing: an fMRI adaptation study. *Neurosci. Lett.*
43. Deutsch D, Henthorn T, Lapidis R. 2011 Illusory transformation from speech to song. *J. Acoust. Soc. Am.* **129**, 2245–2252. (doi:10.1121/1.3562174)
44. Margulis E. 2013 Repetition and emotive communication in music versus speech. *Front. Psychol.* **4**, 167. (doi:10.3389/fpsyg.2013.00167)
45. Tierney A, Dick F, Deutsch D, Sereno M. 2013 Speech versus song: multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cereb. Cortex* **23**, 249–254. (doi:10.1093/cercor/bhs003)
46. Grill-Spector K, Henson R, Martin A. 2006 Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* **10**, 14–23. (doi:10.1016/j.tics.2005.11.006)
47. Sammler D *et al.* 2013 Co-localizing linguistic and musical syntax with intracranial EEG. *Neuroimage* **64**, 134–146. (doi:10.1016/j.neuroimage.2012.09.035)
48. Schönwiesner M, Zatorre RJ. 2008 Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. *Exp. Brain Res.* **187**, 97–105. (doi:10.1007/s00221-008-1286-z)
49. Westbury C, Zatorre R, Evans A. 1999 Quantifying variability in the planum temporale: a probability map. *Cereb. Cortex* **9**, 392–405. (doi:10.1093/cercor/9.4.392)
50. Skoe E, Kraus N. 2012 Human subcortical auditory function provides a new conceptual framework for considering modularity. In *Language and music as cognitive systems* (eds P Rebuschat, M Rohrmeier, J Hawkins, I Cross), pp. 269–282. New York, NY: Oxford University Press.
51. Patel AD. 2011 Why would musical training benefit the neural encoding of speech? The OPERA

- hypothesis. *Front. Psychol.* **2**, 142. (doi:10.3389/fpsyg.2011.00142)
52. Strait D, Kraus N. 2011 Playing music for a smarter ear: cognitive, perceptual and neurobiological evidence. *Music Percept.* **29**, 133–146. (doi:10.1525/MP.2011.29.2.133)
53. Besson M, Chobert J, Marie C. 2011 Transfer of training between music and speech: common processing, attention, and memory. *Front. Psychol.* **2**, 94. (doi:10.3389/fpsyg.2011.00094)
54. Alho K, Rinne T, Herron TJ, Woods DL. 2014 Stimulus-dependent activations and attention-related modulations in the auditory cortex: a meta-analysis of fMRI studies. *Hear. Res.* **307**, 29–41. (doi:10.1016/j.heares.2013.08.001)
55. Tokarev K, Tiunova A, Scharff C, Anokhin K. 2011 Food for song: expression of c-Fos and ZENK in the zebra finch song nuclei during food aversion learning. *PLoS ONE* **6**, e21157. (doi:10.1371/journal.pone.0021157)
56. Peretz I, Aubé W, Armony J. 2013 Toward a neurobiology of musical emotions. In *The evolution of emotional communication: from sounds in nonhuman mammals to speech and music in man* (eds E Altenmüller, S Schmidt, E Zimmermann), pp. 277–299. New York, NY: Oxford University Press.
57. Dehaene S, Cohen L. 2007 Cultural recycling of cortical maps. *Neuron* **56**, 384–398. (doi:10.1016/j.neuron.2007.10.004)