

## MUSIC, LANGUAGE AND MODULARITY FRAMED IN ACTION

Isabelle PERETZ

*BRAMS, University of Montreal, Canada*

Here, I examine to what extent music and speech share processing components by focusing on vocal production, that is, singing and speaking. In shaping my review, the modularity concept has been and continues to play a determinant role. Thus, I will first provide a brief background on the contemporary notion of modularity. Next, I will present evidence that musical abilities depend, in part, on modular processes. The evidence is coming mainly from neuropsychological dissociations. The relevance of findings of overlap in neuroimaging, of interference and domain-transfer effects between music and speech will also be addressed and discussed. Finally, I will contrast the modularity position with the resource-sharing framework proposed by Patel (2003, 2008a). This critical review should be viewed as an invitation to undertake future comparative research between music and language by focusing on the details of the functions that these mechanisms carry out, not only their specificity. Such comparative research is very important not only theoretically but also in practice because of their obvious clinical and educational implications.

### Introduction

We are a musical species as much as we are a linguistic one. By looking at cognition through music and language, we may get insight into the mechanisms that give humans its remarkable power to make sense of sound (Patel, 2008b). Such a comparative research between music and speech has been slow to emerge in cognitive (neuro)psychology. However, as José Morais and I have argued for 20 years, the divergences between music and speech are striking (e.g., Peretz, 2006; Peretz & Morais, 1989) and have crucial implications for the study of music in general, and its origins in particular. Thus, back to the future, fundamental questions regarding the nature of the mechanisms that might be shared between music and language may shape the neurocognitive study of music tomorrow.

For example, imagine you were searching for the genes that are respon-

---

Isabelle Peretz is affiliated to the Department of Psychology at the University of Montreal.

Preparation of this paper was supported by grants from Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research and from a Canada Research Chair. A prior version of this paper will appear as a chapter in *Language and music as cognitive systems*, edited by P. Rebuschat, Martin Rohrmeier, John Hawkins and Ian Cross, Oxford University Press.

Correspondence concerning this article should be addressed to Isabelle Peretz, BRAMS-Pavillon 1420 Mont-Royal, Université de Montréal, Montreal (Qc) Canada H2V 4P3. E-mail: [Isabelle.Peretz@umontreal.ca](mailto:Isabelle.Peretz@umontreal.ca)

sible for musicality. Finding the particular gene or genes for a behavioural trait is a challenging task for there are billion of possible loci for these genes in the genome. However, if indeed music and speech are very similar functions that have common origins, a good starting point would be to look for the genes that have already been identified for speech. One good candidate is the FOXP2 gene. The discovery of this gene as related to speech began with the study of the KE family of language-impaired individuals. The KE family has three generations, in which half the members suffer from a speech and language disorder (Hurst, Baraister, Auger, Graham, & Norell, 1990). Around half of the children of affected individuals have the disorder, whereas none of the children of unaffected individuals do. This inherited disorder has been linked to a small segment of chromosome 7 (Fisher, Vargha-Khadem, Watkins, Monaco, & Pembrey, 1998; Hurst et al., 1990). The chance discovery of an unrelated individual with a similar speech deficit has allowed the narrowing down of the disorder down to a mutation of a specific gene, named FOXP2 (Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001). This gene seems to play a causal role in the development of normal brain circuitry that underlies language and speech (Marcus & Fisher, 2003).

Interestingly, the speech disorder experienced by the KE family is not language-specific. It also affects oral movements. Hence, we may wonder if the mutation of the FOXP2 gene also affects vocal abilities such as singing. It does. Alcock, Passingham, Watkins, and Vargha-Khadem (2000a) tested nine affected members of the KE family and showed that they were impaired in rhythm production (and perception) while they performed as well as normal controls in melody (pitch-based) production (and perception). Hence, FOXP2 seems to participate to music rhythm. Hence, music and speech may have common origins after all.

However, pitch-based musical abilities seem governed by distinct genetic factors. The opposite pattern – preserved rhythm but impaired pitch – characterises amusic (or “tone-deaf”) individuals (e.g., Ayotte, Peretz, & Hyde, 2002). Individuals affected with congenital amusia are impaired on all tasks that require sequential organisation of pitch but do not necessarily have problems with time intervals (Hyde & Peretz, 2004). This pitch deficit is most apparent, and even diagnostic of their condition, when amusics are required to detect an anomalous (i.e., an out-of-key) note in a conventional melody (Ayotte et al., 2002). This musical pitch disorder is also hereditary (Peretz, Cummings, & Dubé, 2007). Indeed, the congenital amusic individuals identified to date have no speech disorder. Thus, the available data are compatible with the idea that there are two innate factors guiding the acquisition of the musical capacity, with one related to temporal sequencing (and possibly related to FOXP2) and the other, pitch sequencing (of which genes remain to be determined).

Thus, as illustrated here, the comparison between music and speech is highly valuable because it provides an entry-point into understanding the genetic factors that contribute to the potentially shared capacity for music and speech and the genetic factors that contribute to music alone. The latter factors, possibly related to pitch-based abilities, may, however, not be unique to music but be involved in speech prosody. This raises the question of what to compare in music and language and how to assess domain-specificity. These two questions are addressed in the present chapter.

More specifically, in this paper, I will expand the modularity position to action rather than to perception. Modularity in perception has been treated in several prior papers (e.g., Justus & Hutsler, 2005; McDermott & Hauser, 2005; Patel, 2003, 2008a; Peretz, 2001). By action, I mean singing and speaking. Here I will review the literature on these two major modes of vocal expression and discuss their respective modularity. First, I will provide a brief background on the contemporary notion of modularity. Next, I will review the evidence for modularity in speaking and singing as arising from four sources: 1) neuropsychological dissociation; 2) overlap in neuroimaging; 3) interference effects; and 4) domain-transfer effects. Finally, I will contrast the modularity position with the resource-sharing framework proposed by Patel (2003, 2008a).

### Modularity or domain-specificity

Modularity speaks directly to the nature of human evolved cognition. Above all, modularity is a useful framework for directing research and individuating cognitive systems. Unfortunately, the question of modularity has fuelled unresolved debates in the domain of language (Lieberman & Whalen, 2000) and of face processing (Gauthier & Curby, 2005). The seeds of this debate are also present in the music domain. Therefore, it is important to address the issue by distinguishing and clarifying some concepts that are often confused when questions of specialisation, domain-specificity, brain localisation and innateness are considered (see Peretz, 2006, for further discussion). These concepts were connected explicitly in Fodor's (1983) proposal on the modularity of mind, and they have been confounded in many subsequent discussions.

Since Fodor's seminal book, concepts have changed. Of all the characteristics, domain-specificity remains the most important (e.g., Peretz & Coltheart, 2003). A domain-specific operation is a distinct mechanism that deals with a particular aspect of the input and does this either exclusively or more effectively than any other mechanism. What individuates a module is its functional specialisation (Barrett & Kurzban, 2006). Most scientists today would probably agree that the mind involves distinct parts (e.g., one for

perception and one for motor control). The key notion is within these large systems, do we have functional specialisation for music and speech?

Functional specialisation is typically reserved for a whole faculty, such as the music faculty and the speech faculty. However, functional specialisation can be very narrow. As narrow as the operation it performs. As argued elsewhere (Coltheart, 1999), there is no theoretical reason for excluding the concept of domain-specificity at the level of processing components. A domain may be as broad and general as *auditory scene analysis* and as narrow and specific as *tonal encoding of pitch*. Both subsystems perform highly specific computations and hence are domain-specific. That is, both components deal with a particular aspect of music, and they do this either exclusively or more effectively than any other mechanisms. Yet, *auditory scene analysis* is supposed to intervene for all incoming sounds (Bregman, 1990), whereas *tonal encoding of pitch* is exclusive to music.

Thus, domain-specificity does not necessarily imply music-specificity or language-specificity. Rather music-specificity should be examined for each subsystem or processing component. In addition, domain-specificity does not necessarily require special-purpose learning mechanisms. Domain-specificity may either emerge from general learning processes or result from the nature of the input code. I will return to this point in section: The resource-sharing framework.

The question here is to what extent music (and language) processing relies on distinct or shared mechanisms. It remains possible that singing involves no music-specific component. In other words, singing may act as a musical form of speaking. For example, singing may engage the mechanisms for speech intonation. Music 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 in relying on the natural disposition for speech. Music may exaggerate particular speech features such as intonation and affective tone, that are so effective for bonding. In this perspective, the actual domain of the language modules is invaded (Sperber & Hirschfeld, 2004). Music could have stabilised in all cultures because music is so effective at co-opting one or several evolved modules. Multiple anchoring in several modules may even contribute to the ubiquity and power of music. Thus, domain-specificity (or modularity) for music and speech requires comparison tests.

### Tests of domain-specificity

Tests of domain specificity can be performed in at least four different ways: 1) by searching for neuropsychological dissociations between music

and speech in brain damaged patients or in developmental disorders; 2) by searching for distinct activation patterns elicited by music and speech in the normal brain; 3) by using interference paradigms in the normal brain; and 4) by studying the effects of transfer between musical abilities and speech abilities. Each method has been used in music and speech production tasks and will be reviewed here.

I will focus on production here because I have already addressed this issue in perception in prior papers (e.g., Peretz, 2001; Peretz & Coltheart, 2003) and because Patel (2003, 2008a) deals with perceptual studies as well. Here, I will review the literature that has compared singing and speaking and assess whether there is evidence for music-specificity.

Unlike speaking, the ability to sing is usually considered to be unevenly distributed in the general population. While fine singing is viewed as the privilege of a selected few who are widely prized for their skill, the vast majority would be deprived of singing skills. Such a belief fuels the notion that the musical capacity cannot be innately determined (Pinker, 1997). If genes were responsible for the human musical capacity, then everyone should be able to engage in musical activities. Everyone should be able to carry a tune, unless they are tone-deaf. Singing should be as natural as speaking is. Recently, we showed that contrary to common belief, singing proficiency is widespread. Occasional singers can match the singing abilities of professional singers (Dalla Bella, Giguère, & Peretz, 2007).

Singing appears as a natural disposition in humans. Singing is universal and found in all cultures. Moreover, singing is a group activity. Its participatory nature, requiring action coordination, is associated to a highly pleasurable experience. This is why singing is a fundamental human ability that is thought to promote group cohesion (Wallin et al., 2000). In support of the social importance of singing is the observation that mothers universally sing to their offspring and that, in turn, singing abilities emerge early and spontaneously during development. The first songs are produced at around one year of age and at 18 months, children start generating recognisable songs (e.g., Ostwald, 1973; for reviews, see Dowling, 1999). This initial proficiency finds echo in adult singing, which is remarkably consistent both within (Bergeson & Trehub, 2002; Halpern, 1989) and across subjects (Levitin, 1994; Levitin & Cook, 1996) in terms of starting pitch and tempo. Therefore, the adult population seems to possess the basic capacities to sing popular songs with proficiency.

With universality, early and spontaneous emergence, consistency and social function, singing abilities represent one of the richest sources of information regarding the nature and origins of music behaviour. Moreover, songs are a unique combination of speech and music. Yet, these are separable in many ways, for lyrics and melody rely on separate codes and are even often

composed by different persons. Yet, music and text are linked and are most often, if not always, heard and played in a combined form. Thus, the study of singing represents a very rich new area for understanding music cognition as it relates to language because it seems guided by largely unconscious processes (Loui, Guenther, Mathys, & Schlaug, 2008), is a natural alliance between music and speech, and is more natural than most perceptual situations.

### *Neuropsychological dissociations*

A module or domain-specific operation does not need to be neurally distinct or dissociable. It is possible that the neural substrate of a music module be intermingled with the networks devoted to a speech module. In that case, it will never be possible for brain damage to affect just the music modules whilst sparing the speech modules and vice versa, although the two types of modules are functionally distinct. In contrast, if the two modules are also neurally separable by involving different areas of the brain, it is possible to observe neuropsychological dissociations. These provide persuasive evidence for the existence of distinct modules. More generally, a mechanism must be neurally separable if the concept of modularity is to be of any use in cognitive neuroscience. The current evidence is consistent with the existence of neurally separable music and speech modules in singing and speaking.

Brain lesions can selectively interfere with speaking while singing remains essentially intact (Hébert, Racette, Gagnon, & Peretz, 2003; Peretz, Gagnon, Hébert, & Macoir, 2004; Racette, Bard, & Peretz, 2006; Schlaug, Marchina, & Norton, 2008; Wilson, Parsons, & Reutens, 2006). This corresponds to the common condition of aphasic patients who can no longer speak but sing. Most cases have preserved singing and prosody (Racette et al., 2006; Warren, Warren, Fox, & Warrington, 2003). Aphasic patients may remain able to sing familiar tunes and learn novel tunes; in contrast, they fail to produce intelligible lyrics in both singing and speaking (Hébert et al., 2003; Peretz et al., 2004; Warren et al., 2003). The results indicate that speech production, whether sung or spoken, is mediated by the same (impaired) language output system, and that this speech route is distinct from both the (spared) musical and prosodic route.

Conversely, brain damage can impair singing exclusively. Patients may lose the ability to sing familiar songs but retain the ability to recite the lyrics and speak with normal prosody (e.g., Peretz, Kolinsky, Tramo, Labrecque, Hublet, Demeurisse et al., 1994). The selectivity of the vocal deficit is not limited to nonmusicians. Schön, Lorber, Spacal, and Dementia (2004) reported the case of an opera singer who was no longer able to sing pitch intervals but who spoke with the correct intonation and expression. The

existence of a specific problem with singing alongside normal speaking, is consistent with damage to processing components that are both essential to the normal process of singing and specific to the musical domain.

A typical objection to this argument is that most people are amateurs at music but experts at speech. Hence, music may suffer more than speech in the case of brain insult. When damaged, amateur abilities (e.g., music) would be more impaired than expert abilities (e.g., speech). As demonstrated in aphasic patients, the expertise effect cannot account for the recurrent findings of brain-damaged patients who are able to sing effectively whilst being unable to speak. Moreover, in developmental disorders, evidence for a similar double dissociation can be found. Children with specific-language impairments can sing but not speak (e.g., Mogharbel, Sommer, Deutsch, Wenglorz, & Laufs, 2005-2006). Conversely, individuals with congenital amusia (or tone-deafness) cannot sing but speak normally (e.g., Ayotte et al., 2002; Dalla Bella, Giguère, & Peretz, 2009). In sum, the domain-specificity of music and language processing extends to production tasks.

Such neuropsychological cases constitute the best and most compelling evidence in favour of modularity for music and speech. The double dissociation implies the existence of anatomically and functionally segregated systems for music and speech in which one production system can function relatively independently of the other so that one system can be selectively impaired. Although this assumption remains unchallenged, sceptics have argued that double dissociations are not conclusive. A double dissociation can be simulated in an artificial network that is built with a unitary system. That is, lesioned connectionist systems are capable of generating double dissociations in the absence of clear separation of functions or modules (e.g., Plaut, 1995). However, there is as yet no plausible unitary explanation that can account for the pattern of selective impairment and sparing of musical abilities reported here. Thus, the evidence points to the existence of at least one distinct module for music and speech.

Could this distinct processing module for music be related to pitch production? Indeed, there is no need for all components that contribute to singing abilities to be specialised for music. Only one critical component, if damaged or absent, could account for all the manifestations of music-specificity. For example, all cases of congenital amusia whom we have studied so far seem to suffer from a dysfunction at this level (Peretz, 2008) and as a consequence may sing out-of-tune (Dalla Bella et al., 2009). Moreover, all amusic cases who suffer from a recognition or production disorder as a consequence of brain damage (Peretz, 2006) are systematically impaired on the pitch dimension, rarely on the time dimension.

In principle, an impairment on the time dimension, particularly in rhythmic synchronisation, should also be detrimental to musical activities.

Rhythm appears as the essence of music. Moreover, rhythm disorders can occur independently from pitch disorders (Alcock et al., 2000a; Alcock, Wade, Anslow, & Passingham, 2000b; Di Pietro, Laganaro, Leeman, & Schneider, 2004), arguing for the functional separability of rhythm and pitch-based processing of music. It remains to determine to what extent these rhythmic disorders affect musical abilities exclusively. Conversely, preserved rhythmic synchronisation may account for the observation that singing at unison (that is along with someone else) improves speech recovery in aphasics while speaking along does not (Racette et al., 2006). Rhythmic factors may also depend on the spoken language. We note that there are more often reports of preserved word articulation in singing in English speakers (e.g., Schlaug et al., 2008) than in French speakers. This might be due to the constraints in text-setting of a stress-language, such as English, to the temporal structure of the melody. More research is needed on the temporal dimension in both singing and speaking, and in chorus singing in particular.

Thus, the current evidence points to musical capacity as being the result of a confederation of functionally isolable modules. To date, however, only abilities related to pitch appear to be uniquely engaged in music. The music-specificity of many other modules remain to be examined (see Patel, 2008a; Peretz & Coltheart, 2003). Nevertheless, the current evidence, essentially based on pitch-related processes, argues against the view that the musical capacity has invaded the speech modules.

### *Overlap in neuroimaging*

Music processing, probably more than language processing, recruits a vast network of regions located in both the left and right hemispheres of the brain, with an overall right-sided asymmetry for pitch-based processing (Peretz & Zatorre, 2005). In this context, it is not surprising that functional neuroimaging of the normal brain reveals significant overlap in activation patterns between music and language tasks. This is also the case of the seven neuroimaging studies in which overt or covert singing and speaking have been compared (Brown, Martinez, & Parsons, 2006; Callan, Tsytsarev, Hanakawa, Callan, Katsuhara, Fukuyama et al., 2006; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Jeffries, Fritz, & Braun, 2003; Koelsch, Schulze, Sammler, Fritz, Müller, & Gruber, 2008; Özdemir, Norton, & Schlaug, 2006; Saito, Ishii, Yagi, Tatsumi, & Mizusawa, 2006). Significant overlap is to be expected. Speech and music not only recruit widely distributed networks of brain regions but also involve multiple processing systems that might be shared. The number of networks involved is particularly large in the case of production tasks since the output system also involves the perceptual systems for auditory monitoring. Many of these processing components might be shared

between music and speech, especially when singing contains lyrics.

In this context, the identification of distinct activation patterns for singing and speaking are more revealing than overlaps. All but one (Koelsch et al., 2008) of the published studies report distinct areas of activation for speaking and singing (but see below the study of Özdemir et al., 2006, in which increase of activation in a brain region is interpreted as a distinct neural correlate while it may simply reflect an increase in task difficulty). It is beyond the scope of the present paper to list all these potentially domain-specific brain regions. Instead, I will briefly summarise the findings of two studies because these are directly relevant to the work done with aphasics, as mentioned in section: *Neuropsychological dissociations*.

In one of these studies, Özdemir and collaborators (2006) used a vocal imitation task for spoken and sung bisyllabic words (e.g., “money” sung on a minor third). It is worth noting that words were pronounced at an abnormally low rate (e.g., one syllable per second), hence being more similar to chanting than speaking. Areas of activation common to all tasks included the inferior pre- and postcentral gyrus, the superior temporal gyrus (STG) and the superior temporal sulcus bilaterally. More interestingly, singing more than speaking revealed (additional) activation in the right STG and in the primary sensorymotor cortex. In addition, singing more than humming showed activation in the right STG, the operculum and inferior frontal gyrus. This is interpreted as possibly reflecting a distinct route for sung words, that is a route that might be used by non-fluent aphasics in singing. However, this might simply indicate that singing with words is a more difficult task than speaking alone or humming alone (Racette & Peretz, 2007). In sum, there was no compelling evidence that singing and speaking involved distinct neural networks.

In contrast, using a well-known song, not just a slowly sung interval as in Özdemir et al.’s study (2006), Saito and collaborators (2006) obtained evidence for a distinct neural network in singing (but not in speaking). They compared singing the lyrics and reciting the same lyrics both alone and along a pre-recorded voice. Both singing alone and at unison activated brain areas that were not involved in reciting the lyrics (but not vice versa: there were no distinct areas associated to speaking). The right inferior frontal gyrus, the right pre-motor cortex and the right anterior insula were found to be active in singing only (both alone and at unison). Since singing and speaking had the lyrics component in common but not the melody, one may interpret these brain areas as specifically related to melody production. However, there was no melody control condition. Interestingly, synchronised singing as compared to synchronised speaking activated the left anterior part of the inferior parietal lobe, the right posterior planum temporale, the right planum polare and the right middle insula. These specific areas may offer a neural account for the clinical observation that word intelligibility of non-

fluent aphasics is enhanced in synchronised singing, as mentioned earlier (Racette et al., 2006).

As an aside, it is noteworthy that there is also more activation in brain regions involved in reward (e.g., nucleus accumbens) in singing than in speaking (e.g., Callan et al., 2006), suggesting a greater emotional component to singing. This is consistent with the fact that singing, more than speaking, is experienced as a highly enjoyable activity in general and in aphasics, in particular. For aphasic patients, singing is often their only spared vocal mode of expression, thereby facilitating proper breathing and increase in volume. This positive experience often motivates them to participate in lengthy and laborious sessions of testing (Racette et al., 2006).

In sum, neuroimaging studies may provide interesting hypotheses regarding the similarities and differences between music and speech, especially when combined with lesion studies, as illustrated here in the case of speech disorders. However, neuroimaging data cannot rival neuropsychological data. This is because neural and functional dissociations have greater inferential power than overlap or associations. Neuroimaging studies are correlational. Moreover, each activated brain area is a vast region that can easily accommodate more than one distinct processing network. Higher resolution may reveal distinct areas. Thus, neuroimaging data alone can hardly be regarded as a challenge to domain-specificity for music (and for language). As attempts for neural separability fail, we should become increasingly skeptical regarding the isolation of music processing components from language processing. In this search for evidence of domain specificity, some tools are, however, more powerful than others. Transcranial Magnetic Stimulation (TMS), to which we now turn, is one of these more promising tools.

### *Interference effects*

TMS has become a widely used technique in cognitive neuroscience because it is the best available method that produces temporary interference with an ongoing neural process while neuroimaging is correlational by measuring a neural index of ongoing activity. This is important because whereas the other methods can indicate that a given neural response is associated with a behaviour of interest, the TMS method can be used to verify that it is essential, by interfering with it. The logic is similar to lesion studies. However, the TMS has three major advantages over lesion studies: 1) the interference is temporary (reversible); 2) the localisation of the “lesion” can be manipulated experimentally; 3) the local interference can be induced in a normal brain that has no co-morbidity due to the brain accident.

When applied in an inhibitory mode to the left inferior cortex in normal right-handed subjects, TMS can create speech arrest while the same

stimulation to the right homologous region does not interfere with either speech or singing (Epstein, Meador, Loring, Wright, Weisman, Sheppard et al., 1999; Stewart, Walsh, Frith, & Rothwell, 2001). On the other hand, singing interference is very difficult to obtain on either side (Epstein et al., 1999; Walsh, personal communication). When applied in a facilitatory mode to the hand motor cortices, speech and singing change the size of the TMS-induced motor evoked potentials of the right and left hand (with corticospinal projections from the left and right hemisphere, respectively; Lo, Fook-Chong, Lau, & Tan, 2003; Sparing, Meister, Wienemann, Buelte, Staedtgen, & Boroojerdi, 2007). During speech, the right-hand potentials are enhanced whereas the left-handed potentials are increased during singing and humming, relative to the articulation of meaningless syllables. Thus, TMS provides strong evidence for the existence of differentially lateralized mechanisms mediating music and language processing, including planning and execution of motor movements.

Similarly, it should be possible to obtain interference and facilitation effects between text and melody in “normal” singing. This is indeed what we found in a song learning task in both musicians and nonmusicians (Racette & Peretz, 2007). Singing both text and melody was more difficult than reciting the text or singing the melody on /la/. Singing both lyrics and tune of a novel song appears as a dual task in which melody and text compete for limited attentional or memory resources.

The interpretation of interference effects between music and language processing as evidence for the operation of distinct mechanisms may appear odd for “radical modularists”. Indeed, if music and language processing components were completely modular in Fodor’s (1983) sense, the processing of one domain, say music, should be encapsulated. That is, it should be immune to the parallel processing of speech. As illustrated above, the text and melody in songs interact with each other. It does not imply that melody and text are processed by a common core of mechanisms. On the contrary, the current evidence points to the existence of largely separable components that compete for general attention or memory. Thus, the observation of interference (or facilitation) does not challenge modularity, it only questions encapsulation. The use of information from multiple sources, especially in singing, is to be expected from an efficient system. However, integration of the information does not falsify a specialised use of information by dedicated music and speech systems.

### *Domain-transfer effects*

Recent research has examined transfer effects between musical and language abilities with the idea that such a transfer is mediated by shared

mechanisms. However, effects of music *training* on language are poorly understood (Schellenberg, 2006). Nonetheless, speculation abounds about the nature of the observed associations. For example, Patel and Iversen (2007) propose that musical training improves sensory tuning which in turn benefits the perception of speech. In principle, this proposal should extend to speaking and singing. Musicians should speak or learn a second language with more proficiency than nonmusicians. Such an association has been recently reported by Slevc and Miyake (2006) who showed that native Japanese speakers with high musical aptitude spoke English with a better pronunciation than their peers with less musical aptitude. Conversely, one would expect that speakers of tonal languages would be more musical than speakers of nontonal languages. Support for this prediction has been recently provided by Pfordresher and Brown (2009) who show that speakers of tonal languages are better able to reproduce musical pitch patterns in singing than English speakers. This domain-transfer effect from language to music supports the notion that tonal language acquisition might fine-tune pitch perception which in turn can be carried over to singing. Nevertheless, there are a number of shortcomings in current studies of domain-transfer effects, as we described in a recent paper (Schellenberg & Peretz, 2008).

First, musical aptitude, music lessons, and musicians are related but not identical concepts. Aptitude refers to “raw” (untutored) abilities, music lessons involve learning, whereas musicianship is likely to be a consequence of aptitude and training combined with other factors. Duration of music lessons predicts cognitive abilities – including language – among children and adults (Schellenberg, 2006). In contrast, comparisons of musicians and non-musicians yield null or inconsistent results (e.g., Helmbold, Rammsayer, & Altenmüller, 2005). Similarly problematic is the failure to account for musical training when studying aptitude (e.g., Slevc & Miyake, 2006), because musical training improves performance on tests of musical aptitude. In other words, observed associations could be either genetic or the consequence of music lessons.

A second and related issue concerns the nature and specificity of associations between musical experience and cognition. Discussion of “special links” with language (i.e., Slevc & Miyake, 2006) is misleading when associations between musical training and cognitive abilities are much more general, extending to working memory, mathematical and spatial abilities. Taking music lessons could be one learning experience that improves executive function, and, consequently, test-taking abilities in a variety of cognitive domains. Indeed, extended musical experience enhances executive control on both visual nonverbal and auditory tasks (Bialystok & Depape, 2009). Moreover, inferences of causation are unfounded in correlational studies of domain-transfer effects. Although isolated experimental evidence indicates

that music lessons have cognitive transfer effects (Moreno, Marques, Santos, Santos, Castro, & Besson, 2009; Schellenberg, 2004), additional studies with random assignment and appropriate control conditions are essential for identifying the nature of the association between music and language.

This leads to the final issue related to modularity for both music and language. Observed associations between music and language, as that reported by Slevc and Miyake (2006) and by Pfordresher and Brown (2009) could just be the product of executive function, domain-general attentional or corticofugal (Wong, Skoe, Russo, Dees, & Kraus, 2007) influences. In sum, even the optimal design, which would test for domain-transfer effects after training with random assignment, may not give insight into the nature of the shared mechanisms between music and speech. Yet, such studies are very important because they have clinical and educational implications.

### The resource-sharing framework

Patel (2003, 2008a) argues that domain-specificity only applies to representations or knowledge. The operations that operate upon these domain-specific representations can be shared or domain-general. Patel refers to these operations as shared neural resources. In other words, representational specificity is distinguished from processing specificity. In the modular view, domain-specificity refers to both the operation and its output representation. In principle, the resource-sharing framework and the modularity concept are amenable to empirical tests.

Much like an Excel program can be used with numbers or names but be independent of these codes, it should be possible to dissociate a processing component from its knowledge basis (and assess its domain-specificity). For example, the acquisition of tonal knowledge uses general principles, by extracting, for example, statistical regularities in the environment. This possibility has been considered for the acquisition of tonal knowledge (Krumhansl, 1990; Tillmann, Bharucha, & Bigand, 2000). Although *tonal encoding of pitch* is music-specific, it may be built on “listeners’ sensitivity to pitch distribution, [which is] an instance of general perceptual strategies to exploit regularities in the physical world” (Oram & Cuddy, 1995, p. 114). Thus, the input and output of the statistical computation may be domain-specific while the learning mechanism is not (Peretz, 2006; Saffran & Thiessen, 2006). Once acquired, the functioning of the system, say the *tonal encoding of pitch*, may be modular, by encoding musical pitch in terms of keys exclusively and automatically.

The same reasoning applies to *auditory scene analysis* and to *auditory grouping*. The fact that these two processing components organise incoming

sounds according to general Gestalt principles, such as pitch proximity, does not entail that their functioning is general-purpose and mediated by a single processing system. They need not be. For instance, it would be very surprising if visual and auditory scene analyses were mediated by the same system. Yet, both types of analyses obey to Gestalt principles. It is likely that the visual and auditory input codes adjust these mechanisms to their processing needs. Thus, the input codes may transform general-purpose mechanisms into highly specialised ones. The existence of multiple and highly specialised micro-systems, even if they function in a very similar way, is more likely, because modularization is more efficient (Marr, 1982).

Thus, it is possible that domain-specificity emerges from the operation of a general mechanism, or from shared neural resources as proposed by Patel (2003, 2008a). However, in practice, it may be very difficult to demonstrate it because the general or “shared” mechanism under study is likely to modularize with experience (Saffran & Thiessen, 2006).

A developmental perspective is likely to be useful in disentangling initial states from modularized end stage, in both typical and atypical developing populations. Developmental disorders could offer special insight into this debate. Advocates of a “domain-general” cognitive system may search for co-occurrence of impairments in music and language (and other spheres of cognition, such as spatial cognition). Such correlation may give cues as to the nature of the processes that are shared between music and language. It may turn out that domain-specificity depends on very few processing components relative to a largely shared common cognitive background. These key components must correspond to domain and human-specific adaptations, while the shared background is likely to be shared with animals. Developmental disorders are particularly well placed to yield insight into both parts of the debate: that which is unique to music and language, and that which is not. It follows that a great deal can be learned by comparing impaired and spared music and language and cognition in individuals both within and between disorders over the course of development.

### Concluding remarks

Although many questions about speech and music processing remain unresolved, there is evidence that musical abilities depend, in part, on modular processes. However, speaking and singing involve multiple processing components. The details of the functions that these mechanisms carry out, not only their specificity, should be the target of future empirical inquiry. As noted above, developmental perspective is likely to be critical in this debate. Neuroscientific tools such as optical imaging may also facilitate our

ability to assess whether distinct brain mechanisms subserve the acquisition of distinct domains of knowledge in infancy (e.g., Pena, Maki, Kovacic, Dehaene-Lambertz, Koizumi, Bouquet et al., 2003). It is clear that continued research, rather than rigid theoretical positions, is needed to make progress on the question of domain-specificity and domain-generality.

To conclude, the notion of modularity remains important in contemporary research. First, the modularity thesis informs empirical investigation by the search for specialisation. Second, modularity makes plausible candidates for evolved information-processing mechanisms and hence for genetically determined mechanisms. The modern concept of “modularity affords a useful conceptual framework in which productive debates surrounding cognitive systems can continue to be framed” (Barrett & Kurzban, 2006, p. 644).

### References

- Alcock, K.J., Passingham, R.E., Watkins, A.J., & Vargha-Khadem, F. (2000a). Pitch and timing abilities in inherited speech and language impairment. *Brain and Language*, 75, 34-46.
- Alcock, K.J., Wade, D., Anslow, P., & Passingham, R.E. (2000b). Pitch and timing abilities in adult left-hemisphere-dysphasic and right-hemisphere subjects. *Brain and Language*, 75, 47-65.
- Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, 125, 238-251.
- Barrett, H.C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, 113, 628-647.
- Bergeson, T.R., & Trehub, S.E. (2002). Absolute pitch and tempo in mother's sons to infants. *Psychological Science*, 13, 72-75.
- Bialystok, E., & Depape, A.M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception & Performance*, 35, 565-74.
- Bregman, A. (1990). *Auditory scene analysis. The perceptual organization of sound*. London: MIT press.
- Brown, S., Martinez, M.J., & Parsons, L.M. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, 23, 2791-2803.
- Callan, D.E., Tsytarev, V., Hanakawa, T., Callan, A.M., Katsuhara, M., Fukuyama, H., & Turner, R. (2006). Song and speech: Brain regions involved with perception and covert production. *Neuroimage*, 31, 1327-1342.
- Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Science*, 3, 115-120.
- Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *Journal of Acoustical Society of America*, 121, 1182-1189.
- Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2009). Singing in congenital amusia. *Journal of the Acoustical Society of America*, 126, 414-424.

- Di Pietro, M., Laganaro, M., Leeman, B., & Schnider, A. (2004). Receptive amusia: Temporal auditory processing deficit in a professional musician following a left temporo-parietal lesion. *Neuropsychologia*, 42, 868-877.
- Dowling, W.J. (1999). The development of music perception and cognition. In D. Deutsch (Ed.), *The psychology of music* (2nd edition ed., pp. 603-625). San Diego: Academic Press.
- Epstein, C.M., Meador, K.J., Loring, D.W., Wright, R.J., Weisman, J.D., Sheppard, S., Lah, J.J., Puhlovich, F., Gaitan, L., & Davey, K.R. (1999). Localization and characterization of speech arrest during transcranial magnetic stimulation. *Clinical Neurophysiology*, 110, 1073-1079.
- Fisher, S.E., Vargha-Khadem, F., Watkins, K.E., Monaco, A.P., & Pembrey, M.E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics*, 18, 168-170.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT press.
- Gauthier, I., & Curby, K.M. (2005). A perceptual traffic jam on highway N170. *Psychological Science*, 14, 30-32.
- Halpern, A.R. (1989). Memory for the absolute pitch of familiar songs. *Memory and Cognition*, 17, 572-581.
- Hébert, S., Racette, A., Gagnon, L., & Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain*, 126, 1838-1850.
- Helmbold, N., Rammsayer, T., & Altenmüller, E. (2005). Differences in primary mental abilities between musicians and nonmusicians. *Journal of Individual Differences*, 26, 74-85.
- Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area SPT. *Journal of Cognitive Neuroscience*, 15, 673-682.
- Hurst, J.A., Baraister, M., Auger, E., Graham, F., & Norell, S. (1990). An extended family with dominantly inherited speech disorder. *Developmental Medicine and Child Neurology*, 32, 352-355.
- Hyde, K.L., & Peretz, I. (2004). Brains that are out-of-tune but in-time. *Psychological Science*, 15, 356-360.
- Jeffries, K.J., Fritz, J.B., & Braun, A.R. (2003). Words in melody: An H2 15O PET study of brain activation during singing and speaking. *NeuroReport*, 14, 749-754.
- Justus, T., & Hutsler, J.J. (2005). Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain-specificity. *Music Perception*, 23, 1-27.
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2008). Functional architecture of verbal and tonal working memory: An fMRI study. *Human Brain Mapping*, 30, 859-873.
- Krumhansl, C.L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
- Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., & Monaco, A.P. (2001). A fork-head domain gene is mutated in a severe speech and language disorder. *Nature*, 413, 519-523.
- Levitin, D.J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics*, 56, 414-423.

- Levitin, D.J., & Cook, P.R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, *58*, 927-935.
- Liberman, A.M., & Whalen, D.H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*, 187-196.
- Lo, Y., Fook-Chong, S., Lau, D.P., & Tan, E.K. (2003). Cortical excitability changes associated with musical tasks: A transcranial magnetic stimulation study in humans. *Neuroscience Letters*, *252*, 85-88.
- Loui, P., Guenther, F., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology*, *18*, R331-R332.
- Marcus, G.F., & Fisher, S.E. (2003). FOXP2 in focus: What can genes tell us about speech and language? *Trends in Cognitive Sciences*, *7*, 257-262.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W.H. Freeman.
- McDermott, J., & Hauser, M. (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Perception*, *23*, 29-59.
- Mogharbel, C., Sommer, G., Deutsch, W., Wenglorz, M., & Laufs, I. (2005-2006). The vocal development of a girl who sings but does not speak. *Musicae Scientiae*, 235-258.
- 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*, 712-723.
- Oram, N., & Cuddy, L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research*, *57*, 103-118.
- Ostwald, P.F. (1973). Musical behavior in early childhood. *Developmental Medicine and Child Neurology*, *15*, 367-375.
- Özdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *NeuroImage*, *33*, 628-635.
- Patel, A. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*, 674-681.
- Patel, A. (2008a). *Music, language, and the brain*. Oxford: Oxford University Press.
- Patel, A. (2008b). Talk of the tone. *Nature*, *453*, 726-727.
- Patel, A., & Iversen, J. (2007). The linguistic benefits of musical abilities. *Trends in Cognitive Sciences*, *11*, 369-372.
- Pena, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences USA*, *100*, 11702-11705.
- Peretz, I. (2001). Music perception and recognition. In B. Rapp (Ed.), *The Handbook of Cognitive Neuropsychology* (pp. 519-540). Hove: Psychology Press.
- Peretz, I. (2006). The nature of music from a biological perspective. *Cognition*, *100*, 1-32.
- Peretz, I. (2008). Musical disorders: From behavior to genes. *Current Directions in Psychological Science*, *17*, 329-333.
- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*, 688-691.

- Peretz, I., Cummings, S., & Dubé, M.-P. (2007). The genetics of congenital amusia (or tone-deafness): A family aggregation study. *American Journal of Human Genetics*, *81*, 582-588.
- Peretz, I., Gagnon, L., Hébert, S., & Macoir, J. (2004). Singing in the brain: Insights from cognitive neuropsychology. *Music Perception*, *21*, 373-390.
- 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*, 1283-1301.
- Peretz, I., & Morais, J. (1989). Music and modularity. *Contemporary Music Review*, *4*, 277-291.
- Peretz, I., & Zatorre, R.J. (2005). Brain organization for music processing. *Annual Review of Psychology*, *56*, 89-114.
- Pfordresher, P., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception & Psychophysics*, *71*, 1385-1398.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Plaut, D.C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *17*, 291-321.
- Racette, A., Bard, C., & Peretz, I. (2006). Making non-fluent aphasics speak: Sing along! *Brain*, *129*, 2571-2584.
- Racette, A., & Peretz, I. (2007). Learning lyrics: To sing or not to sing? *Memory & Cognition*, *35*, 242-253.
- Saffran, J.R., & Thiessen, E.D. (2006). Domain-general learning capacities. In E. Hoff & M. Shatz (Eds.), *Handbook of language development* (pp. 68-86). Cambridge: Blackwell.
- Saito, Y., Ishii, K., Yagi, K., Tatsumi, I., & Mizusawa, H. (2006). Cerebral networks for spontaneous and synchronized singing and speaking. *NeuroReport*, *17*, 1893-1897.
- Schellenberg, E.G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*, 511-514.
- Schellenberg, E.G. (2006). Exposure to music: The truth about the consequences. In G.E. McPherson (Ed.), *The child as musician: A handbook of musical development* (pp. 111-134). Oxford, UK: Oxford University Press.
- Schellenberg, E.G., & Peretz, I. (2008). Music, language, and cognition: Unresolved issues. *Trends in Cognitive Sciences*, *12*, 45-46.
- Schlaug, G., Marchina, S., & Norton, A. (2008). From singing to speaking: Why singing may lead to recovery of expressive language function in patients with Broca's aphasia. *Music Perception*, *25*, 315-323.
- Schön, D., Lorber, B., Spacal, M., & Semenza, C. (2004). A selective deficit in the production of exact musical intervals following right-hemisphere damage. *Cognitive Neuropsychology*, *21*, 773-784.
- Slevc, L.R., & Miyake, A. (2006). Individual differences in second language proficiency: Does musical ability matter? *Psychological Science*, *17*, 675-681.
- Sparing, R., Meister, I.G., Wienemann, M., Buelte, D., Staedtgen, M., & Boroojerdi, B. (2007). Task-dependent modulation of functional connectivity between hand motor cortices and neuronal networks underlying language and music:

- a transcranial magnetic stimulation study in humans. *European Journal of Neuroscience*, 25, 319-323.
- Sperber, D., & Hirschfeld, L.A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences*, 8, 40-47.
- Stewart, L., Walsh, V., Frith, U., & Rothwell, J. (2001). Transcranial magnetic stimulation produces speech arrest but not song arrest. *Annals of the New York Academy of Sciences*, 930, 433-435.
- Tillmann, B., Bharucha, J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, 107, 885-913.
- Wallin, N., Merker, B., Brown, S., (Ed.). (2000). *The origins of music*. Cambridge, MA: MIT press.
- Warren, J.D., Warren, J.E., Fox, N.C., & Warrington, E.K. (2003). Nothing to say, something to sing: Primary progressive dynamic aphasia. *Neurocase*, 9, 140-155.
- Wilson, S., Parsons, K., & Reutens, D. (2006). Preserved singing in aphasia: A case study of the efficacy of melodic intonation therapy. *Music Perception*, 24, 23-36.
- Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10, 420-422.

Received July 28, 2009

Revision received December 30, 2009

Accepted January 2, 2010