Brain and Language 185 (2018) 30–37

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

The language of music: Common neural codes for structured sequences in music and natural language

Jeffrey N. Chiang[a,1], Matthew H. Rosenberg[a,1], Carolyn A. Bufford[a], Daniel Stephens[b], Antonio Lysy[b], Martin M. Monti[a,*]

[a] Department of Psychology, University of California Los Angeles, Los Angeles, CA, USA
[b] Department of Music, UCLA Herb Alpert School of Music, University of California Los Angeles, Los Angeles, CA, USA

ARTICLE INFO

Keywords:
Language
Music
Neuroimaging
Structured sequences
Multivariate Pattern Analysis (MVPA)

ABSTRACT

The ability to process structured sequences is a central feature of natural language but also characterizes many other domains of human cognition. In this fMRI study, we measured brain metabolic response in musicians as they generated structured and non-structured sequences in language and music. We employed a univariate and multivariate cross-classification approach to provide evidence that a common neural code underlies the production of structured sequences across the two domains. Crucially, the common substrate includes Broca's area, a region well known for processing structured sequences in language. These findings have several implications. First, they directly support the hypothesis that language and music share syntactic integration mechanisms. Second, they show that Broca's area is capable of operating supramodally across these two domains. Finally, these results dismiss the recent hypothesis that domain general processes of neighboring neural substrates explain the previously observed "overlap" between neuroimaging activations across the two domains.

1. Introduction

A central intuition in the study of human language as a cognitive phenomenon is the idea that, while listening to a linear signal such as speech, our minds spontaneously build abstract and structured hypotheses representing how discrete elements within a sequence relate to each other (Chomsky, 1957, 1965; Fitch and Martins, 2014; Jackendoff, 2002; Lashley, 1951; Monti, 2017). The use of such representations is most clearly displayed in natural language (Berwick, Friederici, Chomsky, & Bolhuis, 2013; Ding, Melloni, Zhang, Tian, & Poeppel, 2015), but also characterizes other aspects of human cognition, such as logical reasoning (Monti & Osherson, 2012; Osherson, 1975), algebraic cognition (Maruyama, Pallier, Jobert, Sigman, & Dehaene, 2012; Monti, Parsons, & Osherson, 2012; Varley, Klessinger, Romanowski, & Siegal, 2005), and music cognition (Katz & Pesetsky, 2011; Lerdahl, 2001; Patel, 2003), among others. The relationship between the syntactic operation of language and the syntax-like operations of other aspects of human cognition has thus been at the center of a long-standing debate concerning the degree to which human thought is embedded within, or enabled by, natural language (e.g., Lashley, 1951; Boeckx, 2010; Gleitman & Papafragou, 2013; Fitch & Martins, 2014; Fitch 2014; Monti, 2017).

Lashley (1951) commented on the prevalence of structured sequences across domains, noticing that they exhibited the following three properties: (1) connectedness; i.e. no node is isolated from the others, (2) a root element; i.e. "sentence" or "chord" that is superior to others and (3) acyclic structure; establishing order as a unique property (Fitch and Martins 2014; Lashley 1951). In the context of music cognition, the analogy with the structural aspects of language is particularly pronounced. As discussed elsewhere (e.g., Lerdahl & Jackendoff, 1985; Patel, 2003; Fadiga, Craighero, & D'Ausilio, 2009; Fitch 2014; Peretz, Vuvan, Lagrois, & Armony, 2015), music and language are both characterized by discrete elements (e.g., words, chords) which can be (recursively) combined, according to specific rules, to form organized structures (e.g., sentences, melodies) which are typically encoded within linear, time-dependent, signals.

Nonetheless, whether this analogy is substantial or merely superficial remains a debated issue (cf., Peretz et al., 2015). At one end of the spectrum, it has been proposed that language and music are governed by the very same syntactic processes applied to different building blocks (e.g., words vs. notes). According to this view, "[a]ll formal differences between language and music are a consequence of

* Corresponding author at: 7461E Franz Hall, Department of Psychology, University of California Los Angeles, Los Angeles, CA 90095, USA.
E-mail address: monti@ucla.edu (M.M. Monti).
[1] Equal contribution.

https://doi.org/10.1016/j.bandl.2018.07.003
Received 12 October 2017; Received in revised form 4 July 2018; Accepted 15 July 2018

differences in their fundamental building blocks[; i]n all other respects, language and music are identical" (Katz & Pesetsky, 2011). Along similar lines, it has been proposed that the common representations underlying the structure processing in language and music can be localized to the neural mechanisms encapsulated within the left inferior frontal gyrus (IFG; often referred to as Broca's Area), a region hypothesized to operate as a "supramodal hierarchical parser" (Fadiga et al., 2009; Tettamanti & Weniger, 2006). Consistent with this view, a rapidly growing neuroimaging literature has shown music processing to recruit cortical regions overlapping with areas known to be involved in syntactic and semantic aspects of natural language processing (Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Maess, Koelsch, Gunter, & Friederici, 2001; Koelsch, et al.. , 2002; Tillmann, Janata, & Bharucha, 2003; Koelsch et al., 2004; Koelsch, Fritz, Schulze, Alsop, & Schlaug, 2005; Brown, Martinez, & Parsons, 2006; see Rogalsky, Rong, Saberi, & Hickok, 2011, for a conflicting result). Nonetheless, while the observation of overlapping neural substrates is often taken to imply the presence of shared neurocognitive representations between language and music, this is not necessarily the case (Peretz et al., 2015) and indeed has never been shown to be true. This "missing link" in the neuroscientific literature leaves open the possibility that commonly recruited areas of the brain might, in fact, represent very different operations that do not translate, or align, across the two domains, or that are entirely unrelated to the processing of these relationships. In line with this observation, it has been suggested that language and music are in fact better thought of as modular and largely independent of each other (Marin & Perry, 1999; Peretz & Coltheart, 2003). In support of this view, a rich neuropsychological literature has described cases of individuals who exhibit amusia in the absence of aphasia, as well as aphasia in the absence of amusia (Luria et al., 1965; Peretz, 1993; Peretz et al., 1994; Ayotte, Peretz, Rousseau, Bard, & Bojanowski, 2000; Piccirilli, Sciarma, & Luzzi, 2000; Ayotte, Peretz, & Hyde, 2002).

The reason for the contradicting evidence is still a matter of debate. According to some, the fracture between neuropsychological and neuroimaging findings can be reconciled with a middle-ground solution in which language and music are viewed as partially overlapping systems (Patel, 2003; Patel, Iversen, Wassenaar, & Hagoort, 2008). Under this view, referred to as the shared syntactic integration resource hypothesis, language and music are characterized by both domain-specific (i.e., separate) and shared processes. The domain-specific processes relate to the particular features of each syntax, which are recognized as architecturally different, while shared processes provide neural resources for the activation of the relevant stored syntactic representations (Patel, 2012). According to others, the inconsistency between the two sets of findings might instead be due to experimental and neuroanatomical considerations (Fedorenko & Varley, 2016). Specifically, the overlap often reported, in neuroimaging studies, in left inferior frontal regions could be a reflection of task-general demands tied to the use of structural-violation paradigms (e.g., the P600 and the early left/right anterior negativity effects reported in electrophysiological studies; Janata, 1995; Maess et al., 2001; Koelsch et al., 2002, 2005; Steinbeis and Koelsch, 2008; Tillmann et al., 2003; and later localized to the inferior frontal gyri through neuroimaging; Musso et al., 2015; Kunert, Willems, Casasanto, Patel, & Hagoort, 2015). Deviant events are indeed likely to elicit ancillary processes including attentional capture, detection of violated expectations, or error correction, regardless of whether the violation applies to natural language, music, arithmetic, or motor sequences. Such processes are unrelated to the extracting or forging of structured sequences and are known to elicit activation in domain-general regions (proximal or partially overlapping with Broca's Area; see Fedorenko & Varley, 2016, for a detailed discussion).

In the present study, we address the relationship between the mechanisms of natural language and those of music in a 3 Tesla functional magnetic resonance imaging (fMRI) within-subjects design in which competent musicians generate structures in language (active/passive voice sentences versus repeating a verb) and music (root/second-inversion position ascending triads versus repeating a note; cf., Fig. 1 and Table 1). Crucially, we employ a (rarely explored) generation task to avoid the confound of salient events, and we use a multivariate cross-classification approach to resolve the interpretational ambiguity present in the previous neuroimaging literature (which has been specifically advocated for; see Peretz et al., 2015), thereby helping resolve the question of whether natural language and music share a common underlying neural code for representing structured sequences.

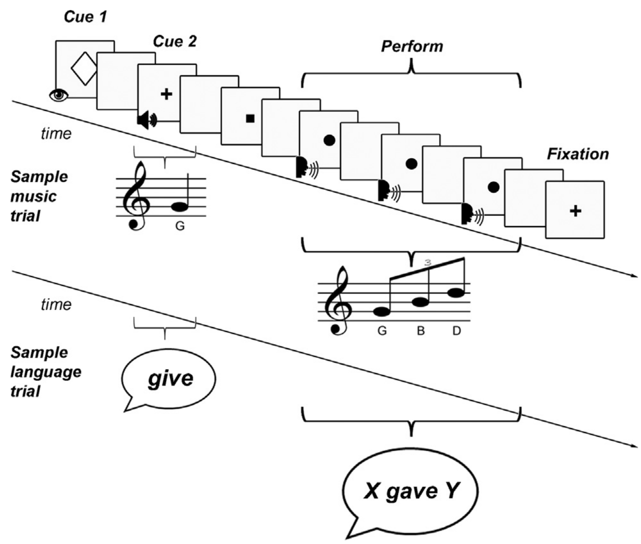

Fig. 1. Experimental design. Sample music and language trials timelines.

Table 1
Sample cues and stimuli.

Language	Structure Active	Structure Passive	Non-Structure Repeat
Cue 1	♦	♣	↻
Cue 2	"Pay"	"Give"	"Tell"
Response	"X paid Y"	"X was given Y"	"Tell, Tell, Tell"
Music	Root Position	IInd Inversion	Repeat
Cue 1	♦	♣	↻
Cue 2	"C"	"D"	"E"
Response	"C-E-G"	"D-G-B"	"E-E-E"

2. Methods

2.1. Participants

We recruited 21 total participants to reach the predetermined sample size (N = 20, 8 female participants) based on previous literature (Musso et al., 2015: N = 11; Kunert et al., 2015: N = 19; Koelsch et al., 2002: N = 20). An additional subject was recruited because the data from one of the participants exhibited excessive motion during the procedure (see below). Participants received $50 compensation for taking part in the experiment. All participants were native English speakers, right handed, and competent musicians currently enrolled in the UCLA Herb Alpert School of Music. Participants were only enrolled if they could demonstrate proficiency in singing/generating both a root position and IInd inversion ascending triad arpeggio. Participants with perfect pitch were excluded. Participants signed informed consent prior to taking part in the session, as per the procedures approved by the UCLA Institutional Review Board.