The human language system does not support music perception

Authors

Xuanyi Chen*[1,2,3], Josef Affourtit[2,3], Rachel Ryskin[2,3,4], Tamar I. Regev[2,3], Samuel Norman-Haignere[5], Olessia Jouravlev[2,3,6], Saima Malik-Moraleda[2,3,7], Hope Kean[2,3], Rosemary Varley†[8], and Evelina Fedorenko*†[2,3,7]

† Co-senior authors
* Corresponding Authors
xc24@rice.edu and evelina9@mit.edu; 43 Vassar Street, Room 46-3037, Cambridge, MA, 02139

Affiliations

[1]Department of Cognitive Sciences, Rice University, TX 77005, USA
[2]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, USA
[3]McGovern Institute for Brain Research, MIT, Cambridge, MA 02139, USA
[4]Department of Cognitive & Information Sciences, University of California, Merced, Merced, CA 95343, USA
[5]Zuckerman Mind, Brain, Behavior Institute, Columbia University, New York, NY 10027, USA
[6]Department of Cognitive Science, Carleton University, Ottawa, Ontario, Canada
[7]The Program in Speech and Hearing Bioscience and Technology, Harvard University, Cambridge, MA 02138, USA
[8]Psychology & Language Sciences, UCL, London, WCN1 1PF, UK

Author contributions:

	XC*	JA*	RR	TR	SNH	OJ	SMM	HK	RV†	EF†
Conceptualization					☑				☑	☑
Design and materials creation	☑			☑	☑	☑		☑	☑	☑
Experimental script creation	☑	☑								
fMRI data collection	☑	☑				☑				
fMRI data preprocessing and analysis	☑	☑				☑	☑	☑		
Behavioral data collection	☑		☑						☑	
Behavioral data analysis			☑						☑	☑
Formal statistical analysis	☑	☑	☑							☑
Figures	☑		☑	☑						
Writing	☑		☑						☑	☑
Editing + comments		☑		☑	☑	☑	☑	☑		
Overall supervision										☑

Conflict of interest

The authors declare no competing financial interests.

Abstract

Language and music are two human-unique capacities whose relationship remains debated. Some argue for overlap in processing mechanisms, especially for structure processing, but others fail to find overlap. Using fMRI, we examined the responses of language brain regions to diverse music stimuli, and also probed the musical abilities of individuals with severe aphasia. Across four experiments, we obtained a clear answer: music perception does not recruit nor requires the language system. The language regions' responses to music are generally low and do not consistently exceed responses elicited by non-music auditory conditions, like animal sounds. Further, the language regions are not sensitive to music structure: they show low responses to both intact and scrambled music, and to melodies with vs. without structural violations. Finally, individuals with aphasia who cannot judge sentence grammaticality perform well on melody well-formedness judgments. Thus the mechanisms that process structure in language do not appear to process music, including musical syntax.

Introduction

To interpret language or appreciate music, we must understand how different elements—words in language, notes and chords in music—relate to each other. Parallels between the structural properties of language and music have been drawn for over a century (e.g., Riemann, 1877, as cited in Swain, 1995; Lindblom & Sundberg, 1969; Fay, 1971; Boiles, 1973; Cooper, 1973; Bernstein, 1976; Sundberg & Lindblom, 1976; Lerdahl & Jackendoff, 1977, 1983; Roads & Wieneke, 1979; Krumhansl & Keil, 1982; Baroni et al., 1983; Swain, 1995; cf. Jackendoff, 2009). However, the question of whether music processing relies on the same mechanisms as those that support language processing continues to spark debate.

The empirical landscape is complex. A large number of studies have argued for overlap in structural processing based on behavioral (e.g., Fedorenko et al., 2009; Slevc et al., 2009; Hoch et al., 2011; Van de Cavey & Hartsuiker, 2016; Kunert et al., 2016), ERP (e.g., Janata, 1995; Patel et al., 1998; Koelsch et al., 2000), MEG (e.g., Maess et al., 2001), fMRI (e.g., Koelsch et al., 2002; Levitin & Menon, 2003; Tillmann et al., 2003; Koelsch, 2006; Kunert et al., 2015; Musso et al., 2015), and ECoG (e.g., Sammler et al., 2009, 2013) evidence (see Tillman, 2012; Kunert & Slevc, 2015; LaCroix et al., 2016, for reviews). However, we would argue that no prior study has compellingly established reliance on shared syntactic processing mechanisms in language and music.

First, evidence from behavioral, ERP, and, to a large extent, MEG studies is indirect because they do not allow to unambiguously determine where neural responses originate (in ERP and MEG, this is due to the 'inverse problem'; Tarantola, 2004; Baillet et al., 2014).

Second, the bulk of the evidence comes from structure-violation paradigms. In such paradigms, responses to the critical condition—which contains an element that violates the rules of tonal music—are contrasted with responses to the control condition, where stimuli obey the rules of tonal music. Because structural violations (across domains) constitute unexpected events, the observed overlap may—and has been argued by some to—reflect domain-general processes, like attention or error detection (e.g., Bigand et al., 2001; Poulin-Charronat et al., 2005; Tillmann et al., 2006; Hoch et al., 2011; Perruchet & Poulin-Charronnat, 2013). Indeed, at least in some studies, unexpected *non-structural* events in music, like a timbre change, have been found to lead to similar neural responses in fMRI (e.g., Koelsch et al., 2002; cf. some differences in EEG effects – e.g., Koelsch et al., 2001), putting into question the interpretation in terms of shared syntactic mechanisms. Relatedly, meta-analyses of neural responses to unexpected events (e.g., Corbetta & Shulman, 2002; Fouragnan et al., 2018; Corlett et al., 2021) have identified regions grossly resembling those reported in studies of music structure violations (see Fedorenko & Varley, 2016 for discussion). It is also important to note that a brain region responsible for processing structure should respond strongly to well-formed stimuli (in addition to being sensitive to deviations from well-formedness)—something that is rarely established (see the fifth point below).

115　*Third*, most prior fMRI (and MEG) investigations have relied on comparisons of group-level
116　activation maps. Such analyses suffer from low functional resolution (e.g., Nieto-Castañón &
117　Fedorenko, 2012; Fedorenko, 2021), especially in cases where the precise locations of functional
118　regions vary across individuals, as in the association cortex (Fischl et al., 2008; Frost & Goebel,
119　2012; Tahmasebi et al., 2012; Vazquez-Rodriguez et al., 2019). Thus, observing activation
120　overlap at the group level does not unequivocally support shared mechanisms. Indeed, studies
121　that used individual-subjects analyses have reported a low or no response to music in the
122　language-responsive regions (Fedorenko et al., 2011; Rogalsky et al., 2011; Deen et al., 2015).
123
124　*Fourth*, the interpretation of some of the observed effects has relied on the so-called 'reverse
125　inference' (Poldrack, 2006, 2011), where function is inferred from a coarse anatomical location:
126　for example, some music-structure-related effects observed in or around 'Broca's area' have
127　been interpreted as reflecting the engagement of linguistic-structure-processing mechanisms
128　(e.g., Maess et al., 2001; Koelsch et al., 2002) given the long-standing association between
129　'Broca's area' and language, including syntactic processing specifically (e.g., Caramazza &
130　Zurif, 1976; Friederici et al., 2006). However, this reasoning is not valid: Broca's area is a
131　heterogeneous region, which houses components of at least two functionally distinct brain
132　networks (Fedorenko et al., 2012; Fedorenko & Blank, 2020): the language-selective network,
133　which responds during language processing, visual or auditory, but does not respond to diverse
134　non-linguistic stimuli (Fedorenko et al., 2011; Monti et al., 2009, 2012; see Fedorenko & Varley,
135　2016 for a review) and the domain-general executive control or 'multiple demand (MD)'
136　network, which responds to any demanding cognitive task and is robustly modulated by task
137　difficulty (Duncan, 2010, 2013; Fedorenko et al., 2013; Assem et al., 2020). As a result, here and
138　more generally, functional interpretation based on coarse anatomical localization is not justified.
139
140　*Fifth*, many prior fMRI investigations have not reported the magnitudes of response to the
141　relevant conditions and only examined statistical maps for the contrast of interest (e.g., a whole
142　brain map showing voxels that respond reliably more strongly to melodies with vs. without a
143　structural violation, and to sentences with vs. without a structural violation). Response
144　magnitudes of experimental conditions relative to a low-level baseline and to each other are
145　critical for interpreting a functional profile of a brain region (see e.g., Chen et al., 2017, for
146　discussion). For example, a reliable *violation > no violation* effect in music (similar arguments
147　apply to language) could be observed when both conditions elicit above-baseline responses, and
148　the violation condition elicits a stronger response (**Figure 1A** left bar graph)—a reasonable
149　profile for a brain region that supports music processing and is sensitive to the target structural
150　manipulation. However, a reliable *violation > no violation* effect could also be observed when
151　both conditions elicit below-baseline responses, and the violation condition elicits a less negative
152　response (**Figure 1A** middle bar graph), or when both conditions elicit low responses—in the
153　presence of a strong response to stimuli in other domains—and the between-condition difference
154　is small (**Figure 1A** right bar graph; note that with sufficient power even very small effects can
155　be highly reliable, but this does not make them meaningful; e.g., Cumming, 2012; Sullivan &
156　Feinn, 2012). The two latter profiles, where a brain region is more active during silence than

157 when listening to music, or when the response is overall low and the effect of interest is
158 minuscule, would be harder to reconcile with a role of this brain region in music processing.
159
160 Similarly, with respect to the music-language overlap question, a reliable *violation > no*
161 *violation* effect for both language and music could be observed in a brain region where sentences
162 with violations and melodies with violations elicit similarly strong responses, and those without
163 violations elicit lower responses (**Figure 1B** left bar graph); but it could also arise in a brain
164 region where sentences with violations elicit a strong response, sentences without violations
165 elicit a lower response, but melodies elicit an overall low response, with the violation condition
166 eliciting a higher response than the no-violation condition (**Figure 1B** right bar graph). Whereas
167 in the first case, it may be reasonable to argue that the brain region in question supports some
168 computation that is necessary to process structure violations in both domains, such interpretation
169 would not be straightforward in the second case. In particular, given the large main effect of
170 language > music, any account of possible computations supported by such a brain region would
171 need to explain this difference instead of simply focusing on the presence of a reliable effect of
172 violation in both domains. In summary, without examining the magnitudes of response, it is not
173 possible to distinguish among many, potentially very different, functional profiles, without which
174 hypothesizing about a brain region's computations is precarious.
175

176
177 **Figure 1**: Illustration of the importance of examining the magnitudes of neural response to the
178 experimental conditions rather than only the statistical significance maps for the contrast(s) of interest. A
179 significant *violation > no violation* effect (A), and overlap between a significant *violation > no violation*
180 effect in language vs. in music (B) are each compatible with multiple distinct functional profiles, only one
181 of which (on the left in each case) supports the typically proposed interpretation (a region that processes
182 structure in some domain of interest in A, and a region that processes structure in both language and
183 music in B).
184
185 Aside from the limitations above, to the best of our knowledge, all prior brain imaging studies
186 have used a single manipulation in one set of materials and one set of participants. To
187 compellingly argue that a brain region supports (some aspects of) structural processing in both
188 language and music, it is important to establish both the *robustness* of the key effect by
189 replicating it with a new set of experimental materials and/or in a new group of participants, and

6

190 its *generalizability* to other contrasts between conditions that engage the hypothesized
191 computation and ones that do not. For example, to argue that a brain region houses a core
192 syntactic mechanism needed to process hierarchical relations and/or recursion in both language
193 and music (e.g., Patel, 2003; Fadiga et al., 2009; Roberts, 2012; Koelsch et al., 2013; Fitch &
194 Martins, 2014), one would need to demonstrate that this region i) responds robustly to diverse
195 structured linguistic and musical stimuli (which all invoke the hypothesized shared computation),
196 ii) is sensitive to more than a single manipulation targeting the hypothesized computations
197 specifically (structured vs. unstructured stimuli, stimuli with vs. without structural violations,
198 stimuli that are more vs. less structurally complex (e.g., with long-distance vs. local
199 dependencies), adaptation to structure vs. some other aspect of the stimulus, etc.) in order to rule
200 out paradigm-/task-specific accounts, and iii) replicates across materials and participants.

202 Finally, the neuropsychological patient evidence is at odds with the idea of shared mechanisms
203 for processing language and music. If language and music relied on the same syntactic
204 processing mechanism, individuals impaired in their processing of linguistic syntax should also
205 exhibit impairments in musical syntax. Although some prior studies report subtle musical deficits
206 in patients with aphasia (Patel et al., 2008; Sammler et al., 2011), the evidence is equivocal, and
207 many aphasic patients appear to have little or no difficulties with music, including the processing
208 of music structure (Luria et al., 1965; Brust, 1980; Marin, 1982; Basso & Capitani, 1985; Polk &
209 Kertesz, 1993; Slevc et al., 2016; Chiapetta et al., 2022). Similarly, children with Specific
210 Language Impairment—a developmental disorder that affects several aspects of linguistic and
211 cognitive processing, including syntactic processing (e.g., Bortolini et al., 1998; Bishop &
212 Norbury, 2002)—show no impairments in musical processing (Fancourt, 2013). In an attempt to
213 reconcile the evidence from acquired and developmental disorders with claims about structure-
214 processing overlap based on behavioral and neural evidence from neurotypical participants, Patel
215 (2003, 2008, 2012; see Slevc & Okada, 2015 and Asano et al., 2021 for related proposals) put
216 forward a hypothesis whereby the representations mediating language and music are stored in
217 distinct brain areas, but the mechanisms that perform online computations on those
218 representations are partially overlapping. We return to this idea in the Discussion.

220 To bring clarity to this ongoing debate, we conducted three fMRI experiments with neurotypical
221 adults, and a behavioral study with individuals with severe aphasia. In each fMRI experiment,
222 we used a well-established language 'localizer' task based on the reading of sentences and
223 nonword sequences (Fedorenko et al., 2010; see Scott et al., 2017 and Malik-Moraleda, Ayyash
224 et al., 2021 for evidence that this localizer is modality-independent) to identify language-
225 responsive areas in each participant individually. These areas have been shown, across dozens of
226 brain imaging studies, to be robustly sensitive to linguistic syntactic processing demands in
227 diverse manipulations (e.g., Keller et al., 2001; Röder et al., 2002; Friederici, 2011; Pallier et al.,
228 2011; Bautista & Wilson, 2016, among many others)—including when defined with the same
229 localizer as the one used here (e.g., Fedorenko et al., 2010, 2012a, 2020; Blank et al., 2016;
230 Mollica et al., 2020; Shain, Blank et al., 2020; Shain et al., 2021a)—and their damage leads to
231 linguistic, including syntactic, deficits (e.g., Caplan et al., 1996; Dick et al., 2001; Wilson &
232 Saygin, 2004; Tyler et al., 2011; Wilson et al., 2012; Mesulam et al., 2014; Ding et al., 2020;

233   Matchin & Hickok, 2020, among many others). We then examined the responses of these

234   language areas to music, and their necessity for processing music structure. In Experiment 1, we

235   included diverse music stimuli including orchestral music, single-instrument music, synthetic

236   drum music, and synthetic melodies, a minimal comparison between songs and spoken lyrics,

237   and a set of non-music auditory control conditions. We additionally examined sensitivity to

238   structure in music across two structure-scrambling manipulations. In Experiment 2, we further

239   probed sensitivity to structure in music using the most common manipulation, contrasting

240   responses to well-formed melodies vs. melodies containing a note that does not obey the

241   constraints of Western tonal music. And in Experiment 3, we examined the ability to

242   discriminate between well-formed melodies and melodies containing a structural violation in

243   three profoundly aphasic individuals across two tasks. Finally, in Experiment 4, we examined the

244   responses of the language regions to yet another set of music stimuli in a new set of participants.

245   Further, the participants were all native speakers of Mandarin, a tonal language, which allowed

246   us to evaluate the hypothesis that language regions may play a greater role in music processing in

247   individuals with higher sensitivity to linguistic pitch (e.g., Deutsch et al., 2006, 2009; Bidelman

248   et al., 2011; Creel et al., 2018; Ngo et al., 2016).

249

250   **Results**

251

252   *Does music elicit a response in the language network?*

253

254   To test whether language regions respond to music, we used four contrasts using data from

255   Experiments 1 and 2. First, we compared the responses to each of the music conditions

256   (orchestral music, single instrument music, synthetic drum music, and synthetic melodies in

257   Experiment 1; well-formed melodies in Experiment 2) against the fixation baseline. Second, we

258   compared the responses to the music conditions against the response to the nonword strings

259   condition—an unstructured and meaningless linguistic stimulus (in Experiment 1, we used the

260   auditory nonwords condition, and in Experiment 2, we used the visual nonwords condition from

261   the language localizer). Third, in Experiment 1, we additionally compared the responses to the

262   music conditions against the response to non-linguistic, non-music stimuli (animal and

263   environmental sounds). A brain region that supports music processing should respond more

264   strongly to music than the fixation baseline and the nonwords condition (our baseline for the

265   language regions); further, if the response is selective, it should be stronger than the response

266   elicited by non-music auditory stimuli. And finally, in Experiment 1, we also directly compared

267   the responses to songs vs. lyrics. A brain region that responds to music should respond more

268   strongly to songs given that they contain a melodic contour in addition to the linguistic content.

269

270   None of the music conditions elicited a strong response in the language network (**Figure 2**;

271   **Table 1**). The responses to music (i) fell at or below the fixation baseline (except for the well-

272   formed melodies condition in Experiment 2, which elicited a small but above-baseline response),

273   (ii) were lower than the response elicited by auditory nonwords (except for the LMFG language

274   fROI, where the responses to music and nonwords were similarly low), and (iii) did not

275   significantly differ from the responses elicited by non-linguistic, non-music conditions. Finally,

the response to songs, which contain both linguistic content and a melodic contour, was not significantly higher than the response elicited by the linguistic content alone (lyrics); in fact, at the network level, the response to songs was reliably lower than to lyrics.
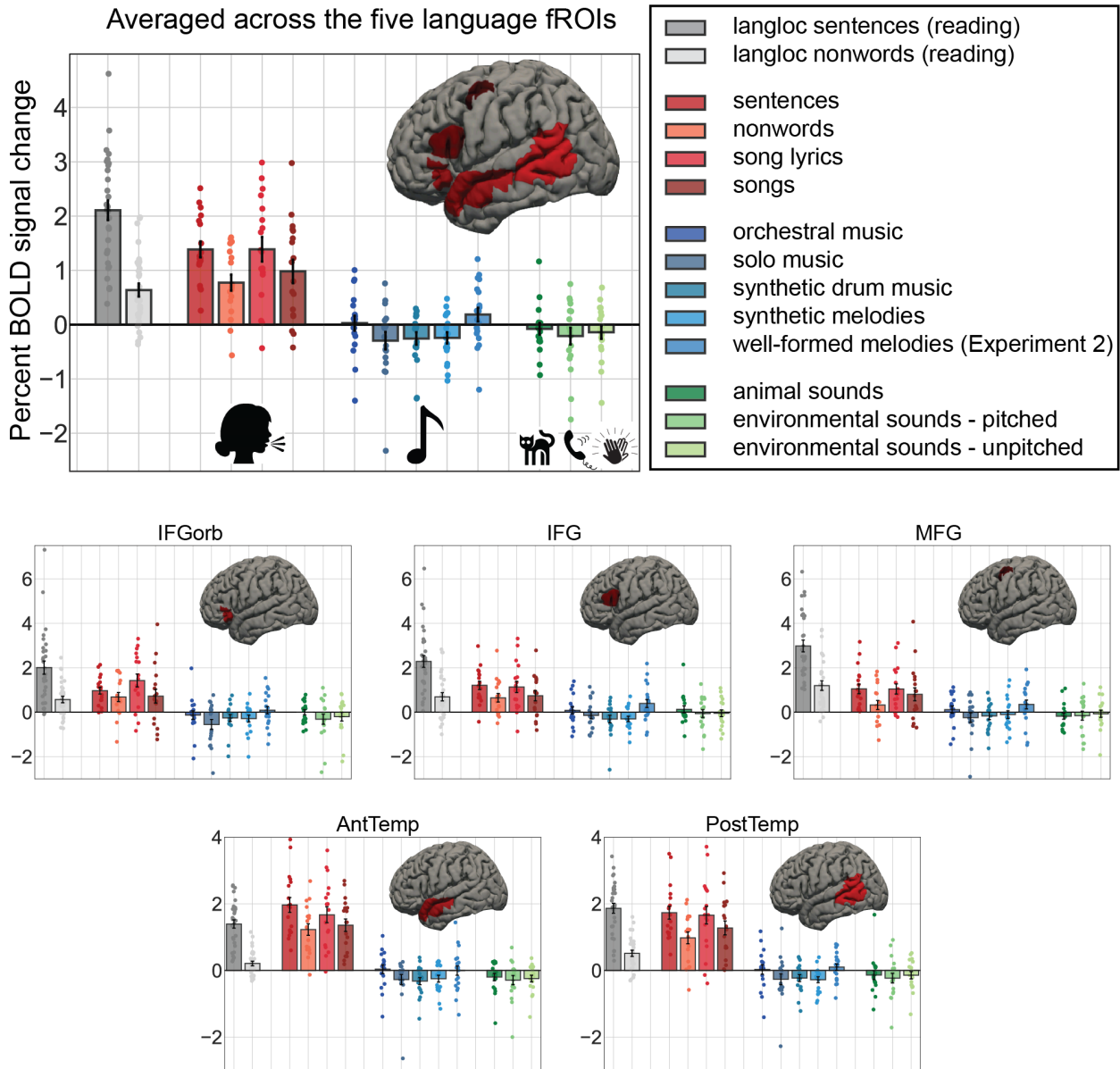
Figure 2. Responses of the language fROIs (pooling across the network – top, and for each fROI individually – bottom) to the language localizer conditions (in grey), to the four auditory conditions containing speech in Experiment 1 (red shades), to the five music conditions in Experiments 1 and 2 (blue shades), and to the three non-linguistic/non-music auditory conditions (green shades) in Experiment 1. Here and elsewhere, the error bars represent standard errors of the mean by participants. For the language localizer results, we include here all participants in Experiments 1 and 2. The responses to the music conditions cluster around the fixation baseline, are much lower than the responses to sentences, and are not higher than the responses to non-music sounds.

Contrast	Language network	LIFGorb	LIFG	LMFG	LAnt Temp	LPost Temp
music > fixation						
orchestral music >fixation	b=0.028 se=0.059 df=162.000 d=0.060 t=0.477 p=0.634	b=-0.129 se=0.188 df=35.995 d=-0.229 t=-0.686 p=1.000	b=0.082 se=0.157 df=36.000 d=0.174 t=0.521 p=1.000	b=0.117 se=0.160 df=35.994 d=0.243 t=0.731 p=1.000	b=0.040 se=0.126 df=35.995 d=0.106 t=0.319 p=1.000	b=0.030 se=0.139 df=36.000 d=0.072 t=0.216 p=1.000
single-instrument music >fixation	b=-0.294 se=0.069 df=162.000 d=-0.537 t=-4.280 p<0.001***	b=-0.552 se=0.217 df=36.000 d=-0.848 t=-2.542 p=0.075	b=-0.141 se=0.151 df=35.997 d=-0.311 t=-0.932 p=1.000	b=-0.243 se=0.211 df=36.000 d=-0.385 t=-1.154 p=1.000	b=-0.272 se=0.148 df=35.995 d=-0.614 t=-1.846 p=0.365	b=-0.264 se=0.159 df=36.000 d=-0.553 t=-1.658 p=0.530
synthetic drum music >fixation	b=-0.255 se=0.054 df=162.000 d=-0.611 t=-4.742 p<0.001***	b=-0.258 se=0.150 df=36.000 d=-0.571 t=-1.715 p=0.475	b=-0.306 se=0.167 df=35.998 d=-0.611 t=-1.832 p=0.375	b=-0.168 se=0.157 df=36.000 d=-0.357 t=-1.070 p=1.000	b=-0.319 se=0.103 df=21.604 d=-1.035 t=-3.108 p=0.025*	b=-0.226 se=0.101 df=36.000 d=-0.749 t=-2.253 p=0.150
synthetic melodies >fixation	b=-0.243 se=0.051 df=162.000 d=-0.623 t=-4.735 p<0.001***	b=-0.286 se=0.150 df=36.000 d=-0.636 t=-1.910 p=0.320	b=-0.299 se=0.117 df=36.000 d=-0.853 t=-2.557 p=0.075	b=-0.108 se=0.172 df=36.000 d=-0.209 t=-0.629 p=1.000	b=-0.247 se=0.100 df=36.000 d=-0.823 t=-2.464 p=0.095	b=-0.276 se=0.087 df=21.603 d=-1.058 t=-3.183 p=0.020*
well-formed melodies (Expt 2) >fixation	b=0.186 se=0.063 df=175.468 d=0.367 t=2.949 p=0.004**	b=0.090 se=0.161 df=40.000 d=0.177 t=0.557 p=1.000	b=0.393 se=0.176 df=39.987 d=0.705 t=2.231 p=0.155	b=0.348 se=0.194 df=40.000 d=0.567 t=1.791 p=0.405	b=-0.003 se=0.136 df=40.000 d=-0.007 t=-0.020 p=1.000	b=0.101 se=0.094 df=40.000 d=0.339 t=1.068 p=1.000
music > nonwords						
orchestral music >nonwords	b=-0.746 se=0.092 df=157.708 d=-0.978 t=-8.097 p<0.001***	b=-0.811 se=0.276 df=36.000 d=-0.981 t=-2.945 p=0.030*	b=-0.569 se=0.142 df=18.000 d=-0.779 t=-4.015 p=0.005**	b=-0.210 se=0.221 df=18.000 d=-0.276 t=-0.954 p=1.0	b=-1.187 se=0.147 df=18.000 d=-1.884 t=-8.101 p<0.001***	b=-0.950 se=0.205 df=18.000 d=-1.427 t=-4.646 p<0.001***
single-instrument music >nonwords	b=-1.068 se=0.100 df=157.689 d=-1.314 t=-10.714 p<0.001***	b=-1.234 se=0.296 df=36.000 d=-1.388 t=-4.167 p<0.001***	b=-0.791 se=0.222 df=18.000 d=-1.101 t=-3.567 p=0.010*	b=-0.571 se=0.235 df=18.000 d=-0.661 t=-2.431 p=0.130	b=-1.500 se=0.196 df=18.000 d=-2.236 t=-7.648 p<0.001***	b=-1.244 se=0.234 df=17.998 d=-1.765 t=-5.315 p<0.001***
synthetic drum music >nonwords	b=-1.029 se=0.087 df=157.720 d=-1.408 t=-11.839 p<0.001***	b=-0.94 se=0.212 df=18.000 d=-1.246 t=-4.430 p<0.001***	b=-0.956 se=0.182 df=18.000 d=-1.275 t=-5.252 p<0.001***	b=-0.496 se=0.245 df=18.000 d=-0.658 t=-2.026 p=0.290	b=-1.546 se=0.187 df=18.000 d=-2.621 t=-8.262 p<0.001***	b=-1.207 se=0.177 df=18.000 d=-2.012 t=-6.817 p<0.001***

synthetic melodies -nonwords	b=-1.017 se=0.087 df=157.683 d=-1.421 t=-11.623 p<0.001***	b=-0.969 se=0.209 df=18.000 d=-1.286 t=-4.642 p<0.001***	b=-0.949 se=0.153 df=18.000 d=-1.441 t=-6.223 p<0.001***	b=-0.435 se=0.252 df=18.000 d=-0.556 t=-1.727 p=0.505	b=-1.474 se=0.195 df=36.000 d=-2.513 t=-7.541 p<0.001***	b=-1.256 se=0.176 df=18.000 d=-2.164 t=-7.136 p<0.001***
well-formed melodies (Expt 2) >nonwords (visual)	b=-0.447 se=0.088 df=175.805 d=-0.557 t=-5.090 p<0.001***	b=-0.533 se=0.210 df=20.000 d=-0.666 t=-2.542 p=0.095	b=-0.402 se=0.197 df=20.000 d=-0.443 t=-2.044 p=0.270	b=-0.691 se=0.243 df=20.000 d=-0.707 t=-2.838 p=0.050	b=-0.229 se=0.137 df=20.000 d=-0.436 t=-1.668 p=0.555	b=-0.379 se=0.119 df=20.000 d=-0.797 t=-3.192 p=0.025*
music > non-linguistic, non-music auditory conditions						
music (combined) >animal sounds	b=-0.114 se=0.060 df=427.875 d=-0.177 t=-1.915 p=0.056	b=-0.306 se=0.148 df=72.000 d=-0.422 t=-2.069 p=0.210	b=-0.295 se=0.146 df=72.000 d=-0.451 t=-2.021 p=0.235	b=0.080 se=0.151 df=72.000 d=0.111 t=0.528 p=1.000	b=-0.002 se=0.090 df=72.000 d=-0.004 t=-0.023 p=1.000	b=-0.048 se=0.094 df=72.000 d=-0.088 t=-0.513 p=1.000
music (combined) >environmental (pitched)	b=0.019 se=0.06 df=427.902 d=0.028 t=0.307 p=0.759	b=0.005 se=0.144 df=72.000 d=0.006 t=0.033 p=1.000	b=-0.104 se=0.133 df=72.000 d=-0.156 t=-0.781 p=1.000	b=0.055 se=0.159 df=72.000 d=0.071 t=0.347 p=1.000	b=0.092 se=0.094 df=72.000 d=0.171 t=0.975 p=1.000	b=0.045 se=0.094 df=72.000 d=0.081 t=0.475 p=1.000
music (combined) >environmental (unpitched)	b=-0.052 se=0.063 df=427.861 d=-0.079 t=-0.823 p=0.411	b=-0.109 se=0.163 df=72.000 d=-0.140 t=-0.666 p=1.000	b=-0.118 se=0.152 df=72.000 d=-0.182 t=-0.778 p=1.000	b=-0.030 se=0.151 df=72.000 d=-0.040 t=-0.198 p=1.000	b=0.042 se=0.097 df=72.000 d=0.083 t=0.429 p=1.000	b=-0.043 se=0.100 df=72.000 d=-0.082 t=-0.426 p=1.000
(melodic contour + linguistic content) > linguistic content						
songs >lyrics	b=-0.408 se=0.102 df=157.895 d=-0.377 t=-4.014 p<0.001***	b=-0.705 se=0.287 df=18.000 d=-0.569 t=-2.454 p=0.125	b=-0.394 se=0.195 df=18.000 d=-0.400 t=-2.025 p=0.290	b=-0.243 se=0.219 df=18.000 d=-0.226 t=-1.107 p=1.000	b=-0.313 se=0.163 df=18.000 d=-0.356 t=-1.925 p=0.350	b=-0.384 se=0.171 df=18.000 d=-0.392 t=-2.246 p=0.185

Table 1. Statistical results (two-sided) for the contrasts between the music conditions and fixation, nonwords, animal sounds, and environmental sounds in Experiments 1 and 2, and for the contrast between songs and lyrics in Experiment 1. Abbreviations: b=the beta estimate for the effect; se=standard error of the mean by participants; df=degrees of freedom; d=Cohen's d (Westfall et al., 2014; Brysbaert & Stevens, 2018); t=the t statistic; p=the significance value (for the individual fROIs, these values have been FDR-corrected for the number of fROIs (n=5)). In light grey, we highlight the results that are not consistent with the role of the language regions in music perception.

Is the language network sensitive to structure in music?

Experiments 1 and 2 (fMRI): Because most prior claims about the overlap between language and music concern the processing of *structure*, given the parallels that can be drawn between the syntactic structure of language and the tonal and rhythmic structure in music (e.g., Lerdahl & Jackendoff, 1977, 1983; cf. Jackendoff, 2009), we used three contrasts to test whether language

304 regions are sensitive to music structure. First and second, in Experiment 1, we compared the
305 responses to synthetic melodies vs. their scrambled counterparts, and to synthetic drum music vs.
306 the scrambled drum music condition. The former targets both tonal and rhythmic structure, and
307 the latter selectively targets rhythmic structure. The reason to examine rhythmic structure is that
308 some patient studies have argued that pitch contour processing relies on the right hemisphere,
309 and rhythm processing draws on the left hemisphere (e.g., Zatorre, 1984; Peretz, 1990; Alcock et
310 al., 2000; cf. Boebinger, 2021 for fMRI evidence of bilateral responses to both tonal and
311 rhythmic structure processing and for lack of spatial segregation between the two), so although
312 most prior work examining the language-music relationship has focused on tonal structure,
313 rhythmic structure may *a priori* be more likely to overlap with linguistic syntactic structure given
314 their alleged co-lateralization based on the patient literature. And third, in Experiment 2, we
315 compared the responses to well-formed melodies vs. melodies with a sour note. A brain region
316 that responds to structure in music should respond more strongly to intact than scrambled music
317 (similar to how language regions respond more strongly to sentences than lists of words; e.g.,
318 Fedorenko et al., 2010; Diachek, Blank, Siegelman et al., 2020), and exhibit sensitivity to
319 structure violations (similar to how language regions respond more strongly to sentences that
320 contain grammatical errors: e.g., Embick et al., 2000; Newman et al., 2001; Kuperberg et al.,
321 2003; Cooke et al., 2006; Friederici et al., 2010; Herrmann et al., 2012; Fedorenko et al., 2020).
322
323 The language regions did not show strong sensitivity to structural manipulations in music
324 (**Figure 3**; **Table 2**). In Experiment 1, the responses to synthetic melodies did not significantly
325 differ from (or were weaker than) the responses to the scrambled counterparts, and the responses
326 to synthetic drum music did not significantly differ from the responses to scrambled drum music.
327 In Experiment 2, at the network level, we observed a small but reliable ($p<0.05$) effect of *sour-
328 note > well-formed melodies*. This effect was not significant in any of the five individual fROIs
329 (even prior to the FDR correction). Moreover, as discussed above, the responses elicited by the
330 well-formed melodies were very low: around the level of the fixation baseline. The responses to
331 both the well-formed melodies and sour-note melodies are below the response elicited by the
332 unstructured (and meaningless) language localizer control condition (nonword sequences).
333
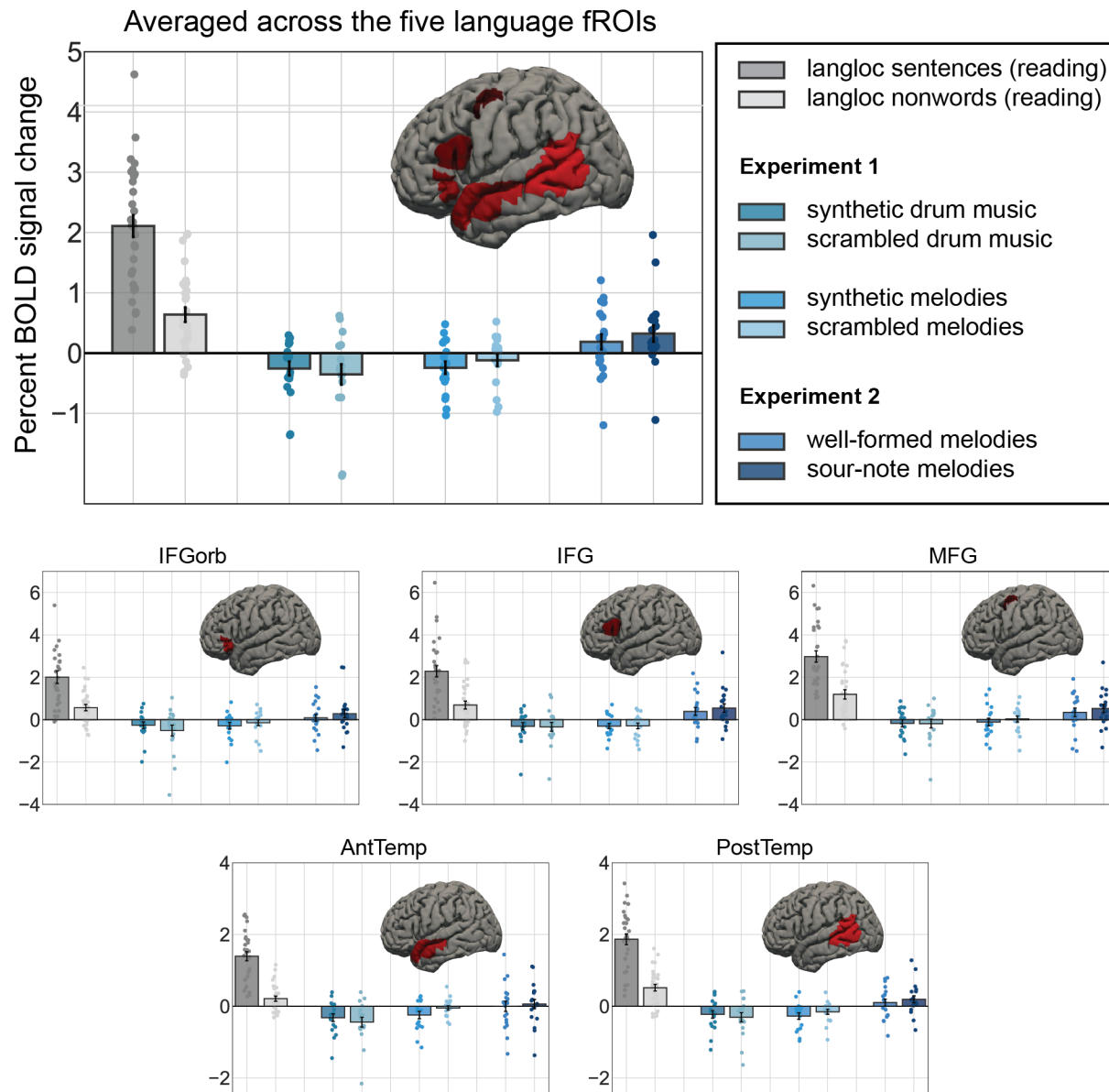
Figure 3. Responses of the language fROIs (pooling across the network – top, and for each fROI individually – bottom) to the language localizer conditions (in grey), and to the three sets of conditions targeting structure in music (in blue). The error bars represent standard error of the mean by participants. For the language localizer results, we include here participants in Experiments 1 and 2. The responses to the music conditions cluster around the fixation baseline, and are much lower than the response to sentences. One of the three critical contrasts (*sour-note > well-formed* melodies) elicits a small but reliable effect at the network level, but it is not individually significant in any of the five fROIs.

Contrast	Language network	LIFGorb	LIFG	LMFG	LAnt Temp	LPost Temp
synthetic drum music >scrambled drum music	b=0.099 se=0.073 df=157.824 d=0.140	b=0.252 se=0.191 df=18.000 d=0.288	b=0.027 se=0.176 df=18.000 d=0.034	b=0.014 se=0.186 df=18.000 d=0.018	b=0.124 se=0.103 df=18.000 d=0.247 t=1.210	b=0.079 se=0.110 df=18.000 d=0.165

	t=1.358 p=0.176	t=1.322 p=1.000	t=0.156 p=1.000	t=0.073 p=1.000	p=1.000	t=0.718 p=1.000
synthetic melodies >scrambled synthetic melodies	b=-0.124 se=0.061 df=157.717 d=-0.238 t=-2.015 p=0.046*	b=-0.147 se=0.130 df=18.000 d=-0.245 t=-1.133 p=1.000	b=-0.009 se=0.153 df=18.000 d=-0.017 t=-0.057 p=1.000	b=-0.143 se=0.202 df=18.000 d=-0.216 t=-0.708 p=1.000	b=-0.199 se=0.101 df=18.000 d=-0.572 t=-1.971 p=0.320	b=-0.121 se=0.106 df=18.000 d=-0.365 t=-1.142 p=1.000
sour-note melodies >well-formed melodies	b=0.138 se=0.069 df=175.886 d=0.186 t=2.007 p=0.046*	b=0.199 se=0.097 df=20.000 d=0.250 t=2.042 p=0.275	b=0.156 se=0.104 df=20.000 d=0.188 t=1.495 p=0.750	b=0.182 se=0.084 df=20.000 d=0.214 t=2.174 p=0.210	b=0.062 se=0.051 df=20.000 d=0.106 t=1.218 p=1.000	b=0.091 se=0.054 df=20.000 d=0.215 t=1.687 p=0.535

Table 2. Statistical results (two-sided) for the contrasts between the synthetic drum music and scrambled drum music, synthetic melodies and scrambled melodies, and sour-note and well-formed melodies contrasts in Experiments 1 and 2. Abbreviations: b=the beta estimate for the effect; se=standard error of the mean by participants; df=degrees of freedom; d=Cohen's d (Westfall et al., 2014; Brysbaert & Stevens, 2018); t=the t statistic; p=the significance value (for the individual fROIs, these values have been FDR-corrected for the number of fROIs (n=5)). In light grey, we highlight the results that are not consistent with the role of the language regions in music perception.

Experiment 3 (behavioral): In Experiment 3, we further asked whether individuals with severe deficits in processing linguistic syntax also exhibit difficulties in processing music structure. To do so, we assessed participants' ability to discriminate well-formed ("good") melodies from melodies with a sour note ("bad"), while controlling for their response bias (how likely they are overall to say that something is well-formed) by computing *d'* for each participant (Green & Swets, 1966), in addition to proportion correct. We then compared the *d'* values of each individual with aphasia to the distribution of *d'* values of healthy control participants using a Bayesian test for single case assessment (Crawford & Garthwaite, 2007) as implemented in the *psycho* package in R (Makowski, 2018). (Note that for the linguistic syntax tasks, it was not necessary to conduct statistical tests comparing the performance of each individual with aphasia to the control distribution because the performance of each individual with aphasia was lower than 100% of the control participants' performances.) We similarly compared the proportion correct on the MBEA scale task of each individual with aphasia to the distribution of accuracies of healthy controls. If linguistic and music syntax draw on the same resources, then individuals with linguistic syntactic impairments should also exhibit deficits on tasks requiring the processing of music syntax.

In the critical music task, where participants were asked to judge the well-formedness of musical structure, neurotypical control participants responded correctly, on average, on 87.1% of trials, suggesting that the task was sufficiently difficult to preclude ceiling effects. Patients with severe aphasia showed intact sensitivity to music structure. The three patients had accuracies of 89.4% (PR), 94.4% (SA), and 97.8% (PP), falling on the higher end of the controls' performance range (**Figure 4**; **Table 3**). Crucially, none of the three aphasic participants' *d'* scores were lower than the average control participants' *d'* scores (M = 2.75, SD = 0.75). In fact, the patients' *d'* scores

376 were high: SA's *d'* was 3.51, higher than 83.91% (95% Credible Interval (CI) [75.20, 92.03]) of
377 the control population, PR's *d'* was 3.09, higher than 67.26% (95% CI [56.60, 78.03]) of the
378 control population, and PP's *d'* was 3.99, higher than 94.55% (95% CI [89.40, 98.57]) of the
379 control population. In the Scale task from the Montreal Battery for the Evaluation of Aphasia,
380 the control participants' performance showed a similar distribution to that reported in Peretz et
381 al. (2003). All participants with aphasia performed within the normal range, with two
382 participants making no errors. PR and PP's score was higher than 85.24% (95% CI [76.94,
383 93.06]) of the control population, providing a conceptual replication of the results from the well-
384 formed/sour-note melody discrimination task. SA's score was higher than 30.57% (95% CI
385 [20.00, 41.50]) of the control population.
386

Participant	SA	PR	PP	Controls
Critical Music Task	170/180	161/180	176/180	M = 156.5/180 SD = 15.8 Min = 109/180 Max = 177/180 N=45
Montreal Battery for the Evaluation of Amusia				
(Critical for this study) Task 1 (Scale)	27/30	30/30	30/30	M = 28/30 SD = 1.89 Min = 23/30 Max = 30/30 N = 45
Task 2 (Interval; "Same Contour" on MBEA CD)	26/30	22/30	18/30	
Task 3 (Contour; "Different Contour" on MBEA CD)	22/30	23/30	18/30	
Task 4 (Rhythm; "Rhythmic Contour" on MBEA CD)	25/30	25/30	22/30	
Task 5 (Meter; "Metric" on MBEA CD)	28/30	22/30	24/30	
Task 6 (Incidental Memory)	28/30	28/30	22/30	

387 **Table 3.** Results for participants with aphasia and control participants on the critical music task and the
388 Scale task of the MBEA (Peretz et al., 2003). For participants with aphasia, we report the results from all
389 six MBEA tasks, for completeness.
390

Critical music task

Scale subset of MBEA

0.4   0.6   0.8   1.0
Proportion correct out of 180

0.4   0.6   0.8   1.0
Proportion correct out of 30
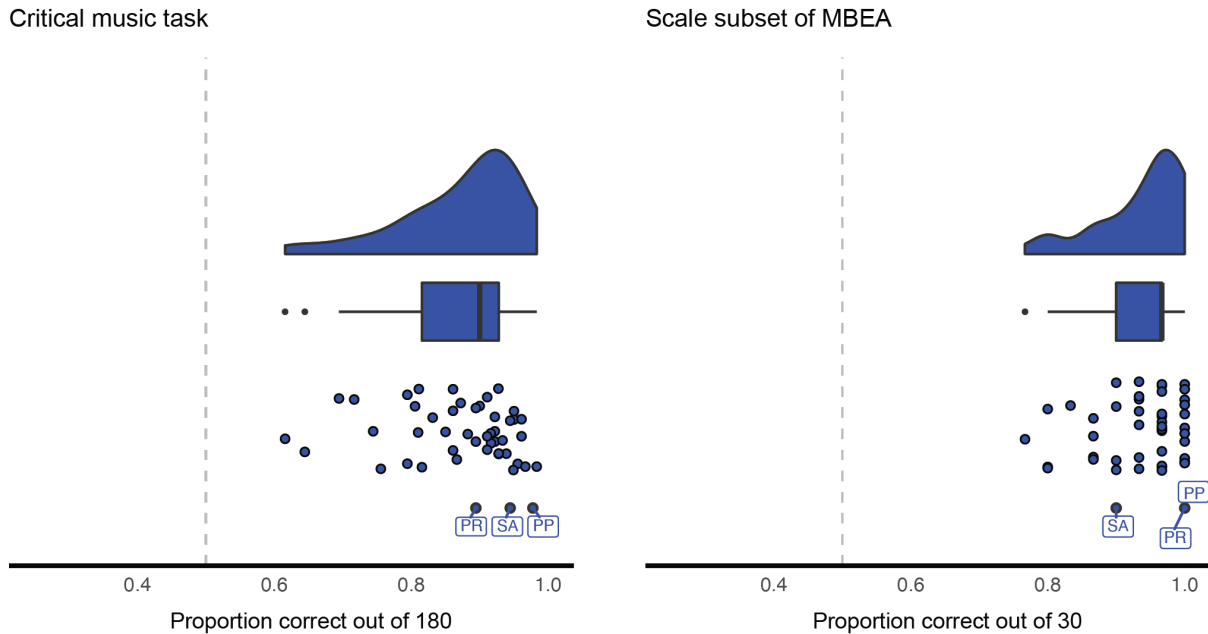
Figure 4. Performance of the control and aphasic participants on two measures of music syntax processing: the critical music task (left), the Scale task of the MBEA (right). The densities show the distribution of proportion correct scores in the control participants and the boxplot shows the quartiles of the control population (the whiskers show 1.5x interquartile range and points represent outliers). The dots show individual participants (for the aphasic individuals, the initials indicate the specific participant). Dashed grey lines indicate chance performance.

Does music elicit a response in the language network of native speakers of a tonal language?

The above analyses focus on the language network's responses to diverse music stimuli and its sensitivity to music structure in English native speakers. However, some have argued that responses to music may differ in speakers of languages that use pitch to make lexical or grammatical distinctions (e.g., Deutsch et al., 2006, 2009; Bidelman et al., 2011; Creel et al., 2018; Ngo et al., 2016). In Experiment 4, we therefore tested whether language regions of Mandarin native speakers respond to music. Similar to Experiment 1, we compared the response to the music condition against a) the fixation baseline, b) the foreign language condition, and c) a non-linguistic, non-music condition (environmental sounds). A brain region that supports music processing should respond more strongly to music than the fixation baseline and the foreign condition; if the response is further selective, it should be stronger than the response elicited by environmental sounds.

Results from Mandarin native speakers replicated the results from Experiment 1: the music condition did not elicit a strong response in the language network (**Figure 5**; **Table 4**). Although the response to music was above the fixation baseline at the network level and in some fROIs, the response did not differ from (or was lower than) the responses elicited by an unfamiliar foreign language (Russian) and environmental sounds.

16

418

Averaged across the five language fROIs
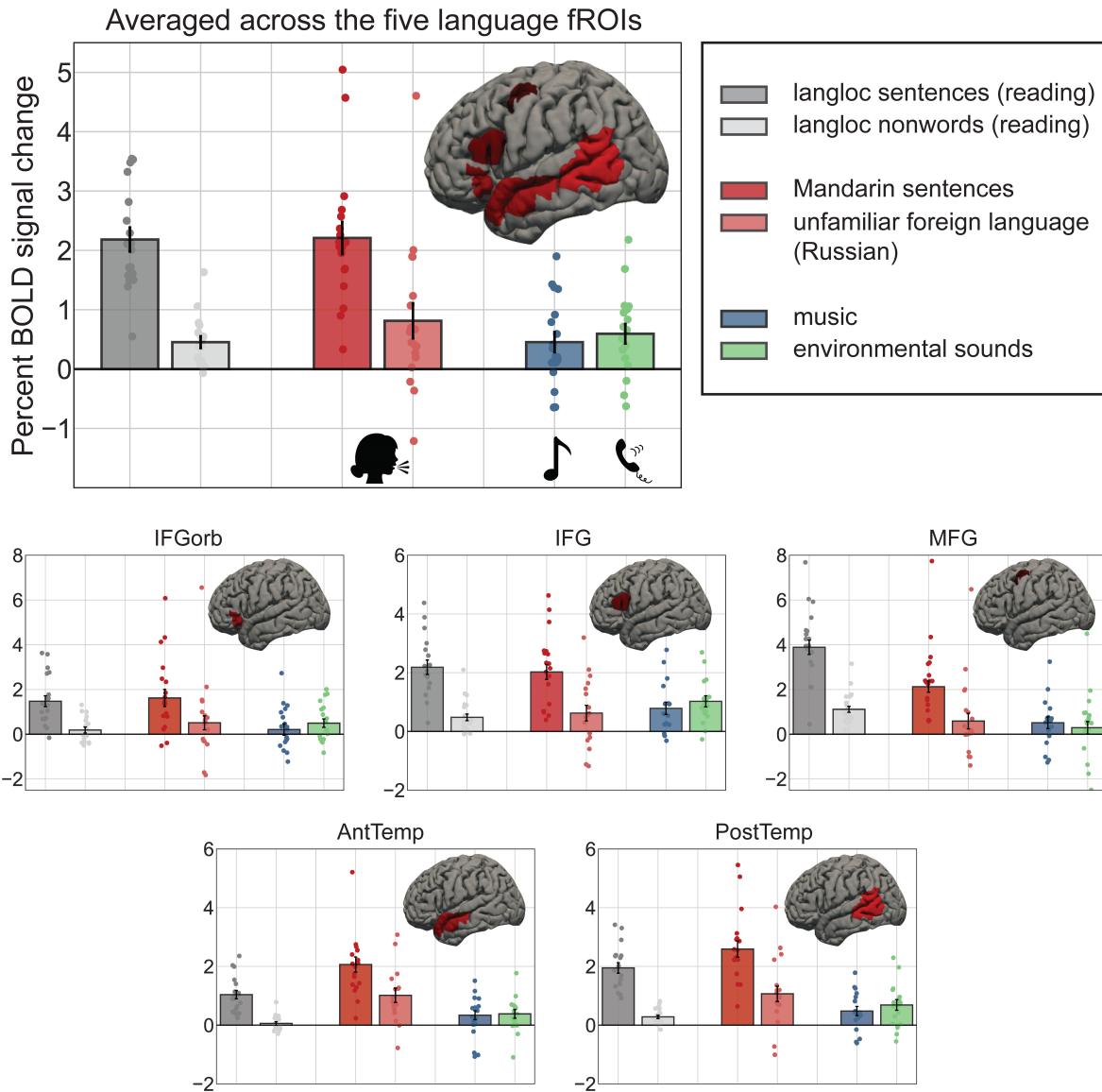
419
420 **Figure 5.** Responses of the language fROIs (pooling across the network – top, and for each fROI
421 individually – bottom) to the language localizer conditions (in grey), to the two auditory conditions
422 containing speech (red shades), to the music condition (blue), and to the non-linguistic/non-music
423 auditory condition (green) in Experiment 4. The error bars represent standard error of the mean by
424 participants. The response to the music condition is much lower than the response to sentences, and is not
425 higher than the response to foreign language and environmental sounds.
426

Contrast	Language network	LIFGorb	LIFG	LMFG	LAnt Temp	LPost Temp
music >fixation	b=0.454 se=0.080 df=157.503 d=0.734	b=0.299 se=0.222 df=36.000 d=0.449	b=0.761 se=0.201 df=21.604 d=1.263	b=0.480 se=0.252 df=36.000 d=0.634	b=0.268 se=0.166 df=36.000 d=0.538	b=0.462 se=0.151 df=21.603 d=1.018

17

	t=5.687 p<0.001***	t=1.346 p=0.935	t=3.790 p=0.005**	t=1.901 p=0.325	t=1.614 p=0.575	t=3.049 p=0.030*
music >foreign	b=-0.359 se=0.141 df=162.000 d=-0.308 t=-2.547 p=0.012*	b=-0.360 se=0.416 df=18.000 d=-0.258 t=-0.865 p=1.000	b=0.123 se=0.309 df=18.000 d=0.124 t=0.398 p=1.000	b=-0.219 se=0.473 df=18.000 d=-0.149 t=-0.463 p=1.000	b=-0.703 se=0.240 df=18.000 d=-0.870 t=-2.926 p=0.045*	b=-0.638 se=0.254 df=18.000 d=-0.686 t=-2.511 p=0.110
music >environmental sounds	b=-0.141 se=0.108 df=157.749 d=-0.154 t=-1.299 p=0.196	b=-0.249 se=0.187 df=18.000 d=-0.280 t=-1.328 p=1.000	b=-0.240 se=0.193 df=18.000 d=-0.302 t=-1.248 p=1.000	b=0.038 se=0.304 df=18.000 d=0.030 t=0.125 p=1.000	b=-0.042 se=0.147 df=18.000 d=-0.065 t=-0.285 p=1.000	b=-0.210 se=0.179 df=18.000 d=-0.310 t=-1.171 p=1.000

Table 4. Statistical results (two-sided) for the contrasts between the music condition and fixation, foreign language, and environmental sounds in Experiment 4. Abbreviations: b=the beta estimate for the effect; se=standard error of the mean by participants; df=degrees of freedom; d=Cohen's d (Westfall et al., 2014; Brysbaert & Stevens, 2018); t=the t statistic; p=the significance value (for the individual fROIs, these values have been FDR-corrected for the number of fROIs (n=5)). In light grey, we highlight the results that are not consistent with the role of the language regions in music perception.

Discussion

We here tackled a much investigated but still debated question: do the brain regions of the language network support the processing of music, especially music structure? Across three fMRI experiments and an investigation of patients with severe aphasia, we obtained a clear answer: the brain regions of the language network, which support the processing of linguistic syntax (e.g., Fedorenko et al., 2010, 2020; Pallier et al., 2011; Bautista & Wilson, 2016; Blank et al., 2016), do not support—and are not needed for—music processing (see Table XX for a summary of the results). We found overall low responses to diverse kinds of music in the language brain regions (**Figure 2**; see Sueoka et al., 2022, for complementary evidence from the inter-subject correlation approach), including in speakers of a tonal language (**Figure 5**), and little or no sensitivity to the manipulations of music structure (**Figure 3**). We further found that the ability to make well-formedness judgments about the tonal structure of music was preserved in severely aphasic patients who cannot make grammaticality judgments for sentences (**Figure 4**). These results align with prior neuropsychological patient evidence of language/music dissociations (e.g., Luria et al., 1965; Brust, 1980; Marin, 1982; Basso & Capitani, 1985; Polk & Kertesz, 1993; Peretz et al., 1994, 1997; Piccirilli et al., 2000; Peretz & Coltheart, 2003; Slevc et al., 2016; Chiapetta et al., 2022), but stand in sharp contrast to numerous reports arguing for shared structure processing mechanisms in the two domains (e.g., Patel et al., 1998; Koelsch et al., 2000; Maess et al., 2001; Koelsch et al., 2002; Levitin & Menon, 2003; see Kunert & Slevc, 2015; LaCroix et al., 2016, for reviews).

Below, we discuss several issues that are relevant for interpreting the current results and/or that these results inform, and outline some limitations of scope of our study.

460   *1. Theoretical considerations about the language-music relationship.*
461
462   Why might we *a priori* think that the language network, or some of its components, may be
463   important for processing music in general, or for processing music structure specifically?
464   Similarities between language and music have long been noted and discussed. For example, as
465   summarized in Jackendoff (2009; see also Patel, 2008), both capacities are human-specific,
466   involve the production of sound (though this is not always the cases for language: cf. sign
467   languages, or written language in literate societies), and have multiple culture-specific variants.
468   However, Jackendoff (2009) notes that i) most cognitive capacities / mechanisms that have been
469   argued to be common to language and music are not *uniquely* shared by language and music, and
470   ii) language and music differ in several critical ways, and these differences are important to
471   consider alongside potential similarities when theorizing about possible shared representations
472   and computations.
473
474   To elaborate on the first point: the cognitive capacity that has perhaps received the most attention
475   in discussions of cognitive and neural mechanisms that may be shared by language and music is
476   the combinatorial capacity of the two domains (e.g., Riemann, 1877, as cited in Swain, 1995;
477   Lindblom & Sundberg, 19769; Fay, 1971; Sundberg & Lindblom, 1976; Lerdahl & Jackendoff,
478   1977, 1983; Roads, 1979; Krumhansl & Keil, 1982). In particular, in language, words can be
479   combined into complex hierarchical structures to form novel phrases and sentences, and in
480   music, notes and chords can similarly be combined to form novel melodies. Further, in both
481   domains, the combinatorial process is constrained by a set of rules. However, this capacity can
482   be observed, in some form, in many other domains, from visual processing, to math, to social
483   cognition, to motor planning, to general reasoning. Similarly, other cognitive capacities
484   necessary to process language and music—including a large long-term memory store for
485   previously encountered elements and patterns, a working memory capacity needed to integrate
486   information as it comes in, an ability to form expectations about upcoming elements, and an
487   ability to engage in joint action—are important for information processing in other domains. An
488   observation that some mental capacity is necessary for multiple domains is compatible with at
489   least two architectures: one where the relevant capacity is implemented (perhaps in a similar
490   way) in each relevant set of domain-specific circuits, and another where the relevant capacity is
491   implemented in a centralized mechanism that all domains draw on (e.g., Fedorenko & Shain,
492   2021). Those arguing for overlap between language and music processing advocate a version of
493   the latter. Critically, any shared mechanism that language and music would draw on should also
494   support information processing in other domains that require the relevant computation. A
495   possible exception, according to Jackendoff (2009), may be the fine-scale vocal motor control
496   that is needed for speech and vocal music production (cf. sign language or instrumental music),
497   but not any other behaviors.
498
499   More importantly, aside from the similarities that have been noted between language and music,
500   numerous differences characterize the two domains. Most notable are their different functions.
501   Language enables humans to express propositional meanings, and thus to share thoughts with
502   one another. The function of music has long been debated (e.g., Darwin, 1871; Pinker, 1994; see

503 e.g., McDermott, 2008 and Mehr et al., 2020, for a summary of key ideas), but most proposed
504 functions have to do with emotional or affective processing, often with a social component[1]
505 (Jackendoff, 2009; Savage et al., 2020). If function drives the organization of the brain (and
506 biological systems more generally; e.g., Rueffler et al., 2012) by imposing particular
507 computational demands on each domain (e.g., Mehr et al., 2020), these fundamentally different
508 functions of language and music provide a theoretical reason to expect cognitive and neural
509 separation between them. Besides, even the components of language and music that appear
510 similar on the surface (e.g., combinatorial processing) differ in deep and important ways (e.g.,
511 Patel, 2008; Jackendoff, 2009; Slevc, 2009).
512
513 *2. Functional selectivity of the language network.*
514
515 The current results add to the growing body of evidence that the left-lateralized fronto-temporal
516 brain network that supports language processing is highly selective for linguistic input (e.g.,
517 Fedorenko et al., 2011; Monti et al., 2009, 2012; Deen et al., 2015; Pritchett et al., 2018;
518 Jouravlev et al., 2019; Ivanova et al., 2020, 2021; Benn, Ivanova et al., 2021; Liu et al., 2020;
519 Deen & Freiwald, 2021; Paunov et al., 2021; see Fedorenko & Blank, 2020 for a review) and not
520 critically needed for many forms of complex cognition (e.g., Lecours & Joanette, 1980; Varley &
521 Siegal, 2000; Varley et al., 2005; Apperly et al., 2006; Woolgar et al., 2018; Ivanova et al., 2021;
522 see Fedorenko & Varley, 2016 for a review). Importantly, this selectivity holds across all
523 components of the language network, including the parts that fall within 'Broca's area' in the left
524 inferior frontal gyrus. As discussed in the introduction, many claims about shared structure
525 processing in language and music have focused specifically on Broca's area (e.g., Patel, 2003;
526 Fadiga et al., 2009; Fitch & Martins, 2014). The evidence presented here shows that the
527 language-responsive parts of Broca's area, which are robustly sensitive to linguistic syntactic
528 manipulations (e.g., Just et al., 1996; Stromswold et al., 1996; Ben-Shachar et al., 2003; Caplan
529 et al., 2008; Peelle et al., 2010; Blank et al., 2016; see Friederici, 2011, for a meta-analysis), do
530 not respond when we listen to music and are not sensitive to structure in music. These results
531 rule out the hypothesis that language and music processing rely on the same mechanism housed
532 in Broca's area.
533
534 It is also worth noting that the underlying premise of the latter hypothesis—of a special
535 relationship between Broca's area and the processing of linguistic syntax (e.g., Caramazza &
536 Zurif, 1976; Friederici, 2018)—has been questioned and overturned. *First*, syntactic processing
537 appears to not be carried out focally, but instead to be distributed across the entire language
538 network, with all of its regions showing sensitivity to syntactic manipulations (e.g., Fedorenko et
539 al., 2010, 2020; Pallier et al., 2011; Blank et al., 2016; Shain, Blank et al., 2020), and with
540 damage to different components leading to similar syntactic comprehension deficits (e.g., Caplan
541 et al., 1996; Dick et al., 2001; Wilson & Saygin, 2004; Mesulam et al., 2014; Mesulam et al.,

---

[1] Although some have discussed the notions of 'meaning' in music (e.g., Meyer, 1961; Raffman, 1993; Cross & Tolbert, 2009; Koelsch, 2001), it is uncontroversial that music cannot be used to express propositional thought (for discussion, see Patel, 2008; Jackendoff, 2009; Slevc, 2009).

542 2015). And *second*, the language-responsive part of Broca's area, like other parts of the language
543 network, is sensitive to both syntactic processing and word meanings, and even sub-lexical
544 structure (Fedorenko et al., 2010, 2012b, 2020; Regev et al., 2021; Shain et al., 2021b). The lack
545 of segregation between syntactic and lexico-semantic processing is in line with the idea of
546 'lexicalized syntax' where the rules for how words can combine with one another are highly
547 dependent on the particular lexical items (e.g., Goldberg, 2002; Jackendoff, 2002, 2007; Sag et
548 al., 2003; Levin & Rappaport-Hovav, 2005; Bybee, 2010; Jackendoff and Audring, 2020), and is
549 contra the idea of 'abstract syntax' where the combinatorial rules are blind to the
550 content/meaning of the to-be-combined elements (e.g., Chomsky, 1965, 1995; Fodor, 1983;
551 Pinker & Prince, 1988; Pinker, 1991, 1999; Pallier et al., 2011).
552
553 *3. Overlap in structure processing in language and music outside of the core language*
554 *network?*
555
556 We have here focused on the core fronto-temporal language network. Could structure processing
557 in language and music draw on shared resources elsewhere in the brain? The prime candidate is
558 the domain-general executive control, or Multiple Demand (MD), network (e.g., Duncan &
559 Owen, 2000; Duncan, 2001, 2010; Assem et al., 2020), which supports functions like working
560 memory and inhibitory control. Indeed, according to Patel's Shared Structural Integration
561 Resource Hypothesis (SSIRH; 2003, 2008, 2012), language and music draw on separate
562 representations, stored in distinct cortical areas, but rely on the same working memory store to
563 integrate incoming elements into evolving structures. Relatedly, Slevc et al. (2013; see Asano et
564 al., 2021 for a related proposal) have recently argued that another executive resource—inhibitory
565 control—may be required for structure processing in both language and music. Although it is
566 certainly possible that some aspects of linguistic and/or musical processing would require
567 domain-general executive resources, based on the available evidence from the domain of
568 language, we would argue that any such engagement does not reflect the engagement of
569 computations like syntactic structure building. In particular, Blank & Fedorenko (2017) found
570 that activity in the brain regions of the domain-general MD network does not closely 'track'
571 linguistic stimuli, as evidenced by low inter-subject correlations during the processing of
572 linguistic input (see Paunov et al., 2021 for a replication). Further, Diachek, Blank, Siegelman et
573 al. (2020) showed in a large-scale fMRI investigation that the MD network is not engaged during
574 language processing in the absence of secondary task demands (cf. the core language network,
575 which is not sensitive to task demands and responds robustly even during passive
576 listening/reading). And Shain, Blank et al. (2020; also, Shain et al., 2021a) have shown that the
577 language network, but not the MD network, is sensitive to linguistic surprisal and working-
578 memory integration costs (see also Wehbe et al., 2021). In tandem, this evidence argues against
579 the role of executive resources in core linguistic computations like those related to lexical access
580 and combinatorial processing, including syntactic parsing and semantic composition (see also
581 Hasson et al., 2015 and Dasgupta & Gershman, 2021 for general arguments against the
582 separation between memory and computation in the brain). Thus, although the contribution of
583 executive resources to music processing deserves further investigation, any overlap within the
584 executive system between linguistic and music processing cannot reflect core linguistic

585 computations, as those seem to be carried out by the language network (see Fedorenko & Shain,
586 2021, for a review).
587
588 Because we had included a localizer for the MD network in our fMRI experiments (based on a
589 spatial working memory task; Fedorenko et al., 2013; Blank et al., 2014; Shashidhara et al.,
590 2019), we examined the responses of these executive brain regions to the music conditions and
591 other conditions in the current study. We found that music conditions elicit a response at or
592 below the fixation baseline, with the exception of the conditions in Experiment 2, which included
593 an explicit task (well-formedness judgments) (the results are available at: https://osf.io/68y7c/).
594 The above-baseline responses to the music conditions accompanied by a task align with the
595 general sensitivity of the executive network to task demands and its role in goal-directed
596 behaviors (e.g., Duncan, 2010; Assem et al., 2020; Diachek, Blank, Siegelman et al., 2020). The
597 fact that the condition with music violations elicits a stronger response than the well-formed
598 condition is in line with the sensitivity of this system to unexpected events across domains, at
599 least in task-based paradigms (e.g., Corbetta & Shulman, 2002; Fouragnan et al., 2018; Corlett et
600 al., 2021). And importantly, the fact that passively listening to rich structured musical stimuli
601 does not elicit an above-baseline response in the MD network argues against the possible role of
602 this network in core computations related to music structure processing. In interpreting past
603 studies, and in any future studies, it is / will be important to rule out extraneous task demands as
604 the source of overlap between music and language processing.
605
606 *4. What brain system processes music, including its structure?*
607
608 We have shown here that the language system shows little or no response when we listen to
609 music. It is worth briefly talking about the brain areas that *are* sensitive to structure in music.
610 Norman-Haignere et al. (2015; see also Boebinger et al., 2020) reported robust selectivity of
611 parts of the auditory cortex for music over diverse kinds of other sounds, including speech (see
612 Peretz et al., 2015, for review and discussion). They further showed that these music-selective
613 components are sensitive to the scrambling of music structure in stimuli similar to those used
614 here in Experiment 1 (see also Fedorenko et al., 2012c, Boebinger, 2021; responses of music-
615 sensitive areas to the conditions of Experiments 1 and 2 are available at: https://osf.io/68y7c/).
616
617 *5. Overlap between music processing and other aspects of speech / language.*
618
619 The current study investigated the role of the language network—which supports 'high-level'
620 comprehension and production—in music processing. As a result, the claims we make are
621 restricted to those aspects of language that are supported by this network. These include the
622 processing of word meanings and combinatorial (syntactic and semantic) processing, but exclude
623 speech perception, prosodic processing, higher-level discourse structure building, and at least
624 some aspects of pragmatic reasoning. Some of these components of language (e.g., pragmatic
625 reasoning) seem *a priori* unlikely to share resources with music. Others (e.g., speech perception)
626 have been shown to robustly dissociate from music (Norman-Haignere et al., 2015; Kell et al.,
627 2018; Boebinger et al., 2020). However, some components of speech and language may, and

628   some do, draw on the same resources as aspects of music. For example, aspects of pitch
629   perception have been argued to overlap between speech and music based on behavioral and
630   neuropsychological evidence (e.g., Wong & Perrachione, 2007; Perrachione et al., 2013; Patel et
631   al., 2008). Indeed, brain regions selectively responsive to different kinds of pitched sounds have
632   been previously reported (Patterson et al., 2002; Penagos et al., 2004; Norman-Haignere et al.,
633   2013, 2015). Some studies have also suggested that music training may improve general rapid
634   auditory processing and pitch encoding that are important for speech perception and language
635   comprehension (e.g., Overy, 2003; Tallal & Gaab, 2006; Wong et al., 2007), although at least
636   some of these effects likely originate in the brainstem and subcortical auditory regions (e.g.,
637   Wong et al., 2007). Other aspects of high-level auditory perception, including aspects of rhythm,
638   may turn out to overlap as well, and deserve further investigation (see Patel, 2008, for an
639   extensive review).
640
641   In conclusion, we have here provided extensive evidence against the role of the language
642   network in music perception, including the processing of music structure. Although the
643   relationship between music and aspects of speech and language will likely continue to generate
644   interest in the research community, and aspects of speech and language other than those
645   implemented in the core fronto-temporal network (Fedorenko & Thompson-Schill, 2014;
646   Fedorenko, 2020) may indeed share some processing resources with (aspects of) music, we hope
647   that the current study helps bring clarity to the debate about structure processing in language and
648   music.
649
650
651   **Materials and methods**
652
653   *Participants*
654
655   <u>Experiments 1, 2, and 4 (fMRI)</u>:
656
657   48 individuals (age 18-51, mean 24.3; 28 (~58%) females) from the Cambridge/Boston, MA
658   community participated for payment across three fMRI experiments (n=18 in Experiment 1;
659   n=20 in Experiment 2; n=18 in Experiment 4; 8 participants overlapped between Experiments 1
660   and 2). 33 participants were right-handed and four left-handed, as determined by the Edinburgh
661   handedness inventory (Oldfield, 1971), or self-report (see Willems et al., 2014, for arguments for
662   including left-handers in cognitive neuroscience experiments); the handedness data for the
663   remaining 11 participants (one in Experiment 2 and 10 in Experiment 4) were not collected. All
664   but one participant (with no handedness information) in Experiment 4 showed typical left-
665   lateralized language activations in the language localizer task described below (as assessed by
666   numbers of voxels falling within the language parcels in the left vs. right hemisphere (LH vs.
667   RH), using the following formula: (LH-RH)/(LH+RH); e.g., Jouravlev et al., 2020; individuals
668   with values of 0.25 or greater were considered to have a left-lateralized language system). For
669   the participant with right-lateralized language activations (with a lateralization value of -0.25 or
670   lower), we used right-hemisphere language regions for the analyses (see SI-3 for an analysis

671 where the LH language regions were used for this participant; the critical results were not
672 affected). Participants in Experiments 1 and 2 were native English speakers; participants in
673 Experiment 4 were native Mandarin speakers and proficient speakers of English (none had any
674 knowledge of Russian, which was used as an unfamiliar foreign-language condition in
675 Experiment 4). All participants gave informed written consent in accordance with the
676 requirements of the Institutional Review Board.
677
678 Experiment 3 (behavioral):
679
680 *Individuals with aphasia*. Three participants with severe and chronic aphasia were recruited to
681 the study (SA, PR, and PP). All participants gave informed consent in accordance with the
682 requirements of the Institutional Review Board. Background information on each participant is
683 presented in **Table 5**. Anatomical scans are shown in **Figure 6A** and extensive perisylvian
684 damage in the left hemisphere, encompassing areas where language activity is observed in
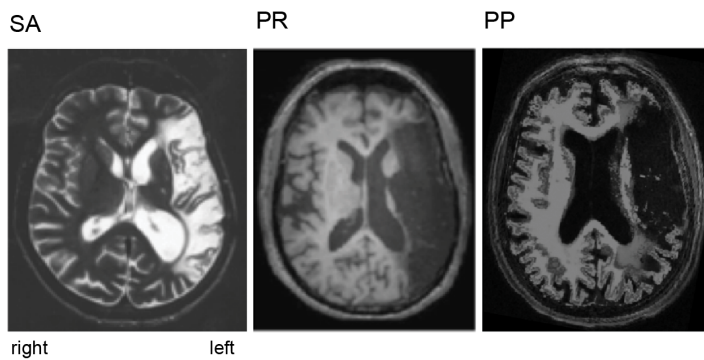685 neurotypical individuals is illustrated in **Figure 6B**.
686

Patient	Sex	Age (years) at testing	Time post-onset (years) at testing	Handedness	Etiology	Premorbid musical experience	Premorbid employment
SA	M	67	21	R	Subdural empyema	Sang in choir; basic sight-reading ability	Police sergeant
PR	M	68	14	L	Left hemisphere stroke	Drummer in band; basic sight-reading ability	Retail manager
PP	M	77	10	R	Left hemisphere stroke	Childhood musical training. No adult experience.	Minerals trader

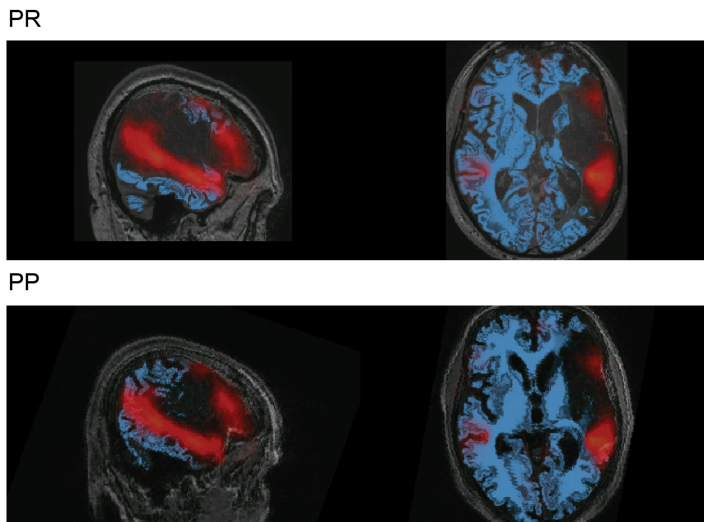687 **Table 5.** Background information on the aphasic participants.
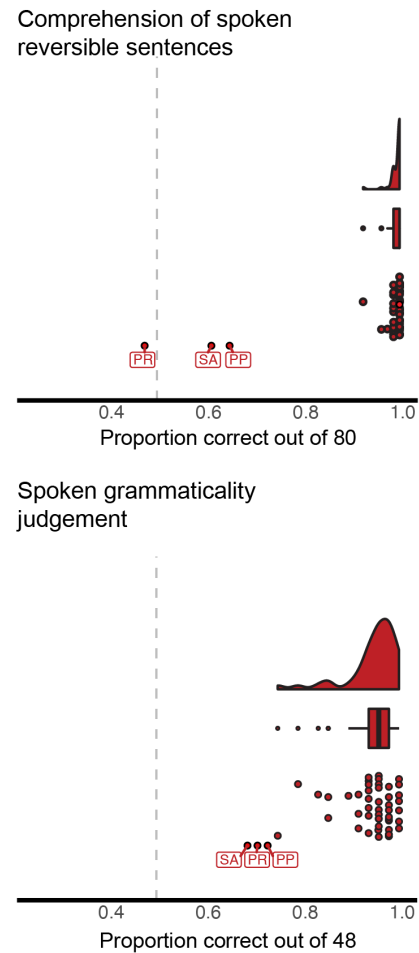688

A. Anatomical scans

SA   PR   PP

right   left

B. Language network overlay

PR

PP

C. Language tasks

Comprehension of spoken reversible sentences

Proportion correct out of 80

Spoken grammaticality judgement

Proportion correct out of 48

Figure 6: **A.** Anatomical scans (T2-weighted for SA, T1-weighted for PR and PP) of the aphasic participants (all scans were performed during the chronic phase, as can be seen from the ventricular enlargement). Note that the right side of the image represents the left side of the brain. **B.** P.R.'s (top) and P.P.'s (bottom) anatomical scans (blue-tinted) shown with the probabilistic activation overlap map for the fronto-temporal language network overlaid (SA's raw anatomical data were not available). The map was created by overlaying thresholded individual activation maps (red-tinted) for the *sentences > nonwords* contrast (Fedorenko et al., 2010) in 220 neurotypical participants (none of whom were participants in any experiments in the current study). As the images show, the language network falls largely within the lesioned tissue in the left hemisphere. **C.** Performance of the control and aphasic participants on two measures of linguistic syntax processing (see Design, materials, and procedure – Experiment 3): the comprehension of spoken reversible sentences (top), and the spoken grammaticality judgments (bottom). The densities show the distribution of proportion correct scores in the control participants and the boxplot shows the quartiles of the control population (the whiskers show 1.5x interquartile range and points represent outliers). The dots show individual participants (for the aphasic individuals, the initials indicate the specific participant). Dashed grey lines indicate chance performance.

Control participants. We used Amazon's Mechanical Turk platform to recruit normative samples for the music tasks and a subset of the language tasks that are most critical to linguistic syntactic comprehension. Ample evidence now shows that online experiments yield data that closely mirror the data patterns in experiments conducted in a lab setting (e.g., Crump et al., 2013). Data from participants with IP addresses in the US who self-reported being native English speakers were included in the analyses. 50 participants performed the critical music task, and the Scale task from the MBEA (Peretz et al., 2003), as detailed below. Data from participants who responded incorrectly to the catch trial in the MBEA Scale task (n=5) were excluded from the analyses, for a final sample of 45 control participants for the music tasks. A separate sample of 50 participants performed the *Comprehension of spoken reversible sentences* task. Data from one participant who completed fewer than 75% of the questions and another participant who did not report being a native English speaker were excluded for a final sample of 48 control participants. Finally, a third sample of 50 participants performed the *Spoken grammaticality judgment* task. Data from one participant who did not report being a native English speaker were excluded for a final sample of 49 control participants.

Design, materials, and procedure

Experiments 1, 2, and 4 (fMRI):

Each participant completed a language localizer task (Fedorenko et al., 2010) and one or more of the critical music perception experiments, along with one or more tasks for unrelated studies. The scanning sessions lasted approximately two hours.

Language localizer. This task is described in detail in Fedorenko et al. (2010) and subsequent studies from the Fedorenko lab (e.g., Fedorenko et al., 2011; Blank et al., 2014; Blank et al., 2016; Pritchett et al., 2018; Paunov et al., 2019; Fedorenko et al., 2020; Shain et al., 2020, among others) and is available for download from https://evlab.mit.edu/funcloc/). Briefly, participants read sentences and lists of unconnected, pronounceable nonwords in a blocked design. Stimuli were presented one word/nonword at a time at the rate of 450ms per word/nonword. Participants read the materials passively and performed a simple button-press task at the end of each trial (included in order to help participants remain alert). Each participant completed two ~6 minute runs. This localizer task has been extensively validated and shown to be robust to changes in the materials, modality of presentation (visual vs. auditory; see the results of Experiments 1 and 4 for additional replications of modality robustness), and task (Fedorenko et al., 2010; Fedorenko, 2014; Scott et al., 2017; Diachek, Blank, Siegelman et al., 2020). Further, a network that corresponds closely to the localizer contrast (*sentences > nonwords*) emerges robustly from whole-brain task-free data—voxel fluctuations during rest (e.g., Braga et al., 2020), providing further support for the idea that this network constitutes a 'natural kind' in the brain and a meaningful unit of analysis.

Experiment 1. Participants passively listened to diverse stimuli across 18 conditions in a long-event-related design (five conditions were not relevant to the current study and therefore not

750 included in the analyses). All stimuli were 9s in length. The conditions were selected to probe
751 responses to diverse kinds of music, to examine sensitivity to structure scrambling in music, to
752 compare responses to songs vs. spoken lyrics, and to compare responses to music stimuli vs.
753 other auditory stimuli.
754
755 The four non-vocal music conditions (all Western tonal music) included orchestral music, single-
756 instrument music, synthetic drum music, and synthetic melodies. The orchestral music condition
757 consisted of 12 stimuli (**SI-Table 4a**) selected from classical orchestras or jazz bands. The
758 single-instrument music condition consisted of 12 stimuli (**SI-Table 4b**) that were played on one
759 of the following instruments: cello (n=1), flute (n=1), guitar (n=4), piano (n=4), sax (n=1), or
760 violin (n=1). The synthetic drum music condition consisted of 12 stimuli synthesized using
761 percussion patches from MIDI files taken from freely available online collections. The stimuli
762 were synthesized using the MIDI toolbox for MATLAB (writemidi).
763
764 The synthetic melodies condition consisted of 12 stimuli transcribed from folk tunes obtained
765 from freely available online collections. Each melody was defined by a sequence of notes with
766 corresponding pitches and durations. Each note was composed of harmonics 1 through 10 of the
767 fundamental presented in equal amplitude, with no gap in-between notes. Phase discontinuities
768 between notes were avoided by ensuring that the starting phase of the next note was equal to the
769 ending phase of the previous note.
770
771 The synthetic drum music and the synthetic melodies conditions had scrambled counterparts to
772 probe sensitivity to music structure. The scrambled drum music condition was created by
773 jittering the inter-note-interval (INI). The amount of jitter was sampled from a uniform
774 distribution (from -0.5 to 0.5 beats). The scrambled INIs were truncated to be no smaller than 5%
775 of the distribution of INIs from the intact drum track. The total distribution of INIs was then
776 scaled up or down to ensure that the total duration remained unchanged. The scrambled melodies
777 condition was created by scrambling both pitch and rhythm information. Pitch information was
778 scrambled by randomly re-ordering the sequence of pitches and then adding jitter to disrupt the
779 key. The amount of jitter for each note was sampled from a uniform distribution centered on the
780 note's pitch after shuffling (from -3 to +3 semitones). The duration of each note was also jittered
781 (from -0.2 to 0.2 beats). To ensure the total duration was unaffected by jitter, N/2 positive jitter
782 values were sampled, where N is the number of notes, and then a negative jitter was added with
783 the same magnitude for each of the positive samples, such that the sum of all jitters equaled 0. To
784 ensure the duration of each note remained positive, the smallest jitters were added to the notes
785 with the smallest durations. Specifically, the note durations and sampled jitters were sorted by
786 their magnitude, summed, and then the jittered durations were randomly re-ordered.
787
788 To allow for a direct comparison between music and linguistic conditions within the same
789 experiment, we included auditory sentences and auditory nonword sequences. The sentence
790 condition consisted of 24 lab-constructed stimuli (half recorded by a male, and half by a female).
791 Each stimulus consisted of a short story (each three sentences long) describing common,

792 everyday events. Any given participant heard 12 of the stimuli (6 male, 6 female). The nonword
793 sequence condition consisted of 12 stimuli (recorded by a male).
794 We also included two other linguistic conditions: songs and spoken lyrics. These conditions were
795 included to test whether the addition of a melodic contour to speech (in songs) would increase
796 the responses of the language regions. Such a pattern might be expected of a brain region that
797 responds to both linguistic content and music structure. The songs and the lyrics conditions each
798 consisted of 24 stimuli. We selected songs with a tune that was easy to sing without
799 accompaniment. These materials were recorded by four male singers: each recorded between 2
800 and 11 song-lyrics pairs. The singers were actively performing musicians (e.g., in a capella
801 groups) but were not professionals. Any given participant heard either the song or the lyrics
802 version of an item for 12 stimuli in each condition.
803
804 Finally, to assess the specificity of the potential responses to music, we included three non-music
805 conditions: animal sounds and two kinds of environmental sounds (pitched and unpitched). The
806 animal sounds condition and the environmental sounds conditions each consisted of 12 stimuli
807 taken from in-lab collections. If individual recordings were shorter than 9s, then several
808 recordings of the same type of sound were concatenated together (100ms gap in between). We
809 included the pitch manipulation in order to test for general responsiveness to pitch—a key
810 component of music—in the language regions. The materials for all conditions are available at
811 OSF: https://osf.io/68y7c/.
812
813 The remaining five conditions (consisting of three acoustically manipulated versions of the
814 sentence condition, and two acoustically manipulated versions of the synthetic melodies
815 condition) were of no relevance to the current study and are therefore not discussed.
816
817 For each participant, stimuli were randomly divided into six sets (corresponding to runs) with
818 each set containing two stimuli from each condition. The order of the conditions for each run
819 was selected from four predefined palindromic orders, which were constructed so that conditions
820 targeting similar mental processes (e.g., orchestral music and single-instrument music) were
821 separated by other conditions (e.g., speech or animal sounds). Each run contained three 10s
822 fixation periods: at the beginning, in the middle, and at the end. Otherwise, the stimuli were
823 separated by 3s fixation periods, for a total run duration of 456s (7min 36s). All but two
824 participants completed all six runs (and thus got a total of 12 experimental events per condition);
825 the remaining two completed four runs (and thus got 8 events per condition).
826
827 Because, as noted above, we have previously established that the language localizer is robust to
828 presentation modality, we used the visual localizer to define the language regions. However, in SI-
829 2 we show that the critical results are similar when auditory contrasts (*sentences > nonwords* in
830 Experiment 1, or *Mandarin sentences > foreign* in Experiment 4) are instead used to define the
831 language regions.
832
833 *Experiment 2.* Participants listened to well-formed melodies (adapted and expanded from
834 Fedorenko et al., 2009) and melodies with a structural violation in a long-event-related design, and

judged the well-formedness of the melodies. As discussed in the Introduction, this type of manipulation is commonly used to probe sensitivity to music structure, including in studies examining language-music overlap (e.g., Patel et al., 1998; Koelsch et al., 2000, 2002; Maess et al., 2001; Tillmann et al, 2003; Fedorenko et al., 2009; Slevc et al., 2009; Kunert et al., 2015; Musso et al., 2015). The melodies were between 11 and 14 notes. The well-formed condition consisted of 90 melodies, which were tonal and ended in a tonic note with an authentic cadence in the implied harmony. All melodies were isochronous, consisting of quarter notes except for the final half note. The first five notes established a strong sense of key. Each melody was then altered to create a version with a "sour" note: the pitch of one note (from among the last four notes in a melody) was altered up or down by one or two semitones, so as to result in a non-diatonic note while keeping the melodic contour (the up-down pattern) the same. The structural position of the note that underwent this change varied among the tonic, the fifth, and the major third. The full set of 180 melodies was distributed across two lists following a Latin Square design. Any given participant heard stimuli from one list. The materials are available at OSF: https://osf.io/68y7c/.

For each participant, stimuli were randomly divided into two sets (corresponding to runs) with each set containing 45 melodies (22 or 23 per condition). The order of the conditions, and the distribution of inter-trial fixation periods, was determined by the optseq2 algorithm (Dale et al., 1999). The order was selected from among four predefined orders, with no more than four trials of the same condition in a row. In each trial, participants were presented with a melody for three seconds followed by a question, presented visually on the screen, about the well-formedness of the melody ("Is the melody well-formed?"). To respond, participants had to press one of two buttons on a button box within two seconds. When participants answered, the question was replaced by a blank screen for the remainder of the two-second window; if no response was made within the two-second window, the experiment advanced to the next trial. Responses received within one second after the end of the previous trial were still recorded to account for the possible slow responses. The screen was blank during the presentation of the melodies. Each run contained 151s of fixation interleaved among the trials, for a total run duration of 376s (6min 16s). All but four participants completed both runs (due to experimenter error, two participants completed two runs from different lists which means they heard both versions of some melodies; because their neural data looked similar to the rest of the participants, we chose to include their data); the remaining four completed one run. Due to a script error, participants only heard the first 12 notes of each melody during the three seconds stimulus presentation. Therefore, we only analyzed the 80 pairs (160 of the 180 total melodies) where the contrastive note appeared within the first 12 notes.

Experiment 4. Participants passively listened to single-instrument music, environmental sounds, sentences in an unfamiliar foreign language (Russian), and Mandarin sentences in a blocked design. All stimuli were 5-5.95s in length. The conditions were selected to probe responses to music, and to compare responses to music stimuli vs. other auditory stimuli. The critical music condition consisted of 60 stimuli selected from classical pieces by J.S. Bach played on cello, flute, or violin (n=15 each) and jazz music played on saxophone (n=15). The environmental sounds condition consisted of 60 stimuli selected from in-lab collections and included both pitched and unpitched stimuli. The foreign language condition consisted of 60 stimuli selected from Russian audiobooks

(short stories by Paustovsky, and "Fathers and Sons" by Turgenev). The foreign language condition was included because creating a 'nonwords' condition (the baseline condition we typically use for defining the language regions; Fedorenko et al., 2010) is challenging in Mandarin given that most words are monosyllabic, thus most syllables carry some meaning. As a result, sequences of syllables are more akin to lists of words. Therefore, we included the unfamiliar foreign language condition, which we know also works well as a baseline (Ayyash, Malik-Moraleda et al., 2020). The Mandarin sentence condition consisted of 240 stimuli (120 lab-constructed sentences, each recorded by a male and a female native speaker). The Mandarin sentence stimuli were divided into four lists, each consisting of 60 unique sentences (half recorded by a male, and half by a female) and 60 unique nonword sequences (half recorded by a male, and half by a female). The materials are available at OSF: https://osf.io/68y7c/. The experiment also included five (speech) conditions of no relevance to the current study which are therefore not discussed.

Stimuli were grouped into blocks with each block consisting of three stimuli and lasting 18s (stimuli were padded with silence to make each trial exactly six seconds long). For each participant, blocks were divided into 10 sets (corresponding to runs), with each set containing two blocks from each condition. The order of the conditions for each run was selected from eight predefined palindromic orders. Each run contained three 14s fixation periods: at the beginning, in the middle, and at the end, for a total run duration of 366s (6min 6s). Five participants completed eight of the 10 runs (and thus got 16 blocks per condition; the remaining thirteen completed six runs (and thus got 12 blocks per condition). (We had created enough materials for 10 runs, but based on observing robust effects for several key contrasts in the first few participants who completed six to eight runs, we administered 6-8 runs to the remaining participants.)

Because we have previously found that an English localizer works well in native speakers of diverse languages, including Mandarin, as long as they are proficient in English (Malik-Moraleda, Ayyash et al., 2021), we used the same localizer in Experiment 4 as the one used in Experiments 1 and 2, for consistency. However, in SI-2 (**SI-Figure 2c**, **SI-Table 2c**) we show that the critical results are similar when the *Mandarin sentences > foreign* contrast is instead used to define the language regions.

Experiment 3 (behavioral):

Language assessments. Participants with aphasia were assessed for the integrity of lexical processing using word-to-picture matching tasks in both spoken and written modalities (ADA Spoken and Written Word-Picture Matching; Franklin et al., 1992). Productive vocabulary was assessed through picture naming. In the spoken modality, the Boston Naming Test was employed (Kaplan et al., 2001), and in writing, the PALPA Written Picture Naming subtest (Kay et al., 1992). Sentence processing was evaluated in both spoken and written modalities through comprehension (sentence-to-picture matching) of reversible sentences in active and passive voice. In a reversible sentence, the heads of both noun phrases are plausible agents, and therefore, word order (in a word-order-based language like English) is the only cue to who is

921 doing what to whom. Participants also completed spoken and written grammaticality judgment
922 tasks, where they made a yes/no decision as to the grammaticality of a word string. The task
923 employed a subset of sentences from Linebarger et al. (1983).
924
925 All three participants exhibited severe language impairments that disrupted both comprehension
926 and production (**Table 6**). For lexical-semantic tasks, all three participants displayed residual
927 comprehension ability for high imageability/picturable vocabulary, although more difficulty was
928 evident on the synonym matching test, which included abstract words. They were all severely
929 anomic in speech and writing. Sentence production was severely impaired with output limited to
930 single words, social speech (expressions, like "How are you?"), and other formulaic expressions
931 (e.g., "and so forth"). Critically, all three performed at or close to chance level on spoken and
932 written comprehension of reversible sentences and grammaticality judgments; each patient's
933 scores were lower than all of the healthy controls (**Table 6** and **Figure 6C**).
934

Participant	SA	PR	PP	Controls
Lexical-semantic assessments				
ADA Spoken Word-Picture Matching (chance = 16.5)	60/66	61/66	64/66	N/A
ADA Written Word-Picture Matching (chance = 16.5)	62/66	66/66	58/66	N/A
ADA spoken synonym matching (chance = 80)	123/160	121/160	135/160	N/A
ADA written synonym matching (chance = 80)	121/160	145/160	143/160	N/A
Boston Naming Test (NB: accepting both spoken and written responses)	4/60	4/60	11/60	N/A
PALPA 54 Written Picture Naming	24/60	2/60	1/60	N/A
Syntactic assessments				
Comprehension of spoken reversible sentences (chance = 40)	49/80	38/80	52/80	Mean = 79.5/80 SD = 1.03 Min = 74/80 Max = 80/80 N=48
Comprehension of written reversible sentences (chance = 40)	42/80	49/80	51/80	N/A
Spoken grammaticality judgments (chance = 24)	33/48	34/48	35/48	Mean = 45.5/48 SD = 2.52 Min = 36/48

				Max = 48/48 N=49
Written grammaticality judgments (chance = 24)	29/48	24/48	29/48	N/A

Table 6. Results of language assessments for participants with aphasia and healthy controls. For each test, we show number of correctly answered questions out of the total number of questions.

Critical music task. Participants judged the well-formedness of the melodies from Experiment 2. Judgments were intended to reflect the detection of the key violation in the sour versions of the melodies. The full set of 180 melodies was distributed across two lists following a Latin Square design. All participants heard all 180 melodies. The control participants heard the melodies from one list, followed by the melodies from the other list, with the order of lists counter-balanced across participants. For the participants with aphasia, each list was further divided in half, and each participant was tested across four sessions, with 45 melodies per session.

Montreal Battery for the Evaluation of Amusia. To obtain another measure of music competence/sensitivity to music structure, we administered the Montreal Battery for the Evaluation of Amusia (MBEA) (Peretz et al., 2003). The battery consists of six tasks that assess musical processing components described by Peretz & Coltheart (2003): three target melodic processing, two target rhythmic processing, and one assesses memory for melodies. Each task consists of 30 experimental trials (and uses the same set of 30 base melodies) and is preceded by practice examples. Some of the tasks additionally include a catch trial, as described below. For the purposes of the current investigation, the critical task is the "Scale" task. Participants are presented with pairs of melodies that they have to judge as identical or not. On half of the trials, one of the melodies is altered by modifying the pitch of one of the tones to be out of scale. Like our critical music task, this task aims to test participants' ability to represent and use tonal structure in Western music, except that instead of making judgments on each individual melody, participants compare two melodies on each trial. This task thus serves as a conceptual replication (Schmidt, 2009). One trial contains stimuli designed to be easy, intended as a catch trial to ensure that participants are paying attention. In this trial, the comparison melody has all its pitches set at random. This trial is excluded when computing the scores.

Control participants performed just the Scale task. Participants with aphasia performed all six tasks, distributed across three testing sessions to minimize fatigue.

fMRI data acquisition, preprocessing, and first-level modeling (for Experiments 1, 2, and 4)

Data acquisition. Whole-brain structural and functional data were collected on a whole-body 3 Tesla Siemens Trio scanner with a 32-channel head coil at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. T1-weighted structural images were collected in 176 axial slices with 1 mm isotropic voxels (repetition time (TR) = 2,530 ms; echo time (TE) = 3.48 ms). Functional, blood oxygenation level-dependent (BOLD) data were acquired using an EPI sequence with a 90º flip angle and using GRAPPA with an acceleration

factor of 2; the following parameters were used: thirty-one 4.4 mm thick near-axial slices acquired in an interleaved order (with 10% distance factor), with an in-plane resolution of 2.1 mm × 2.1 mm, FoV in the phase encoding (A >> P) direction 200 mm and matrix size 96 × 96 voxels, TR = 2000 ms and TE = 30 ms. The first 10 s of each run were excluded to allow for steady state magnetization (see OSF https://osf.io/68y7c/ for the pdf of the scanning protocols).

Preprocessing. Data preprocessing was carried out with SPM12 (using default parameters, unless specified otherwise) and supporting, custom MATLAB scripts. Preprocessing of functional data included motion correction (realignment to the mean image of the first run using 2nd-degree b-spline interpolation), normalization into a common space (Montreal Neurological Institute (MNI) template) (estimated for the mean image using trilinear interpolation), resampling into 2 mm isotropic voxels, smoothing with a 4 mm FWHM Gaussian filter, and high-pass filtering at 128s.

First-level modeling. For both the language localizer task and the critical experiments, a standard mass univariate analysis was performed in SPM12 whereby a general linear model (GLM) estimated, for each voxel, the effect size of each condition in each experimental run. These effects were each modeled with a boxcar function (representing entire blocks/events) convolved with the canonical Hemodynamic Response Function (HRF). The model also included first-order temporal derivatives of these effects, as well as nuisance regressors representing entire experimental runs, offline-estimated motion parameters, and timepoints classified as outliers (scan-to-scan differences in global BOLD signal above 5 standard deviations, or scan-to-scan motion above 0.9 mm).

Definition of the language functional regions of interest (for Experiments 1, 2, and 4)

For each critical experiment, we defined a set of language functional regions of interest (fROIs) using group-constrained, subject-specific localization (Fedorenko et al., 2010). In particular, each individual map for the *sentences > nonwords* contrast from the language localizer was intersected with a set of five binary masks. These masks (**Figure 2**; available at OSF: https://osf.io/68y7c/) were derived from a probabilistic activation overlap map for the same contrast in a large set of participants (n=220) using watershed parcellation, as described in Fedorenko et al. (2010) for a smaller set of participants. These masks covered the fronto-temporal language network in the left hemisphere. Within each mask, a participant-specific language fROI was defined as the top 10% of voxels with the highest *t*-values for the localizer contrast.

Statistical Analyses and the results for sanity-check analyses

All analyses were performed with linear mixed-effects models using the "lme4" package in R with *p*-value approximation performed by the "lmerTest" package (Bates et al., 2015; Kuznetsova et al., 2017). Effect size (Cohen's d) was calculated using the method from Westfall et al. (2014) and Brysbaert & Stevens (2018).

1. Validation of the language fROIs and sanity-check analyses(for Experiments 1, 2, and 4)

To ensure that the language fROIs behave as expected (i.e., show a reliably greater response to the sentences condition compared to the nonwords condition), we used an across-runs cross-validation procedure (e.g., Nieto-Castañón & Fedorenko, 2012). In this analysis, the first run of the localizer was used to define the fROIs, and the second run to estimate the responses (in percent BOLD signal change, PSC) to the localizer conditions, ensuring independence (e.g., Kriegeskorte et al., 2009); then the second run was used to define the fROIs, and the first run to estimate the responses; finally, the extracted magnitudes were averaged across the two runs to derive a single response magnitude for each of the localizer conditions. Statistical analyses were performed on these extracted PSC values. Consistent with much previous work (e.g., Fedorenko et al., 2010; Mahowald & Fedorenko 2016; Diachek, Blank, Siegelman et al., 2020), each of the language fROIs showed a robust *sentences > nonwords* effect (all ps < 0.001).

2. Sanity check and critical analyses (for Experiments 1, 2, and 4)

To estimate the responses in the language fROIs to the conditions of the critical experiments here and in the critical analyses, the data from all the runs of the language localizer were used to define the fROIs, and the responses to each condition were then estimated in these regions. Statistical analyses were then performed on these extracted PSC values. For Experiments 1 and 4, we repeated the analyses using alternative language localizer contrasts to define the language fROIs (auditory *sentences > nonwords* in Experiment 1, and *Mandarin sentences > foreign* in Experiment 4), which yielded quantitatively and qualitatively similar responses (see SI-2).

We conducted two sets of sanity check analyses. First, to ensure that auditory conditions that contain meaningful linguistic content elicit strong responses in the language regions relative to perceptually similar conditions with no discernible linguistic content, we compared the auditory sentences condition with the auditory nonwords condition (Experiment 1) or with the foreign language condition (Experiment 4). Indeed, as expected, the auditory sentence condition elicited a stronger response than the auditory nonwords condition (Experiment 1) or the foreign language condition (Experiment 4). These effects were robust at the network level (ps < 0.001; **SI-Table 1a**). Further, the *sentences > nonwords* effect was significant in all but one language fROI in Experiment 1, and the *sentences > foreign* effect was significant in all language fROIs in Experiment 4 (ps < 0.05; **SI-Table 1a**).

And second, to ensure that the music conditions elicit strong responses in auditory cortex, we extracted the responses from a bilateral anatomically defined auditory cortical region (area Te1.2 from the Morosan et al., 2001 cytoarchitectonic probabilistic atlas) to the six critical music conditions: orchestral music, single instrument music, synthetic drum music, and synthetic melodies in Experiment 1; well-formed melodies in Experiment 2; and the music condition in Experiment 4. Statistical analyses, comparing each condition to the fixation baseline, were

1059 performed on these extracted PSC values. As expected, all music conditions elicited strong
1060 responses in a primary auditory area bilaterally (all ps ≅ 0.001; **SI-Table 1b**; **SI-Figure 1**).
1061
1062 ***Critical analyses***
1063
1064 To characterize the responses in the language network to music perception, we asked three
1065 questions. First, we asked whether music conditions elicit strong responses in the language
1066 regions. Second, we investigated whether the language network is sensitive to structure in music,
1067 as would be evidenced by stronger responses to intact than scrambled music, and stronger
1068 responses to structural violations compared to no-violation control. And third, we asked whether
1069 music conditions elicit strong responses in the language regions of individuals with high
1070 sensitivity to linguistic pitch—native speakers of a tonal language (Mandarin).
1071
1072 For each contrast (the contrasts relevant to the three research questions are detailed below), we
1073 used two types of linear mixed-effect regression models:
1074 i) the language network model, which examined the language network as a whole; and
1075 ii) the individual language fROI models, which examined each language fROI separately.
1076
1077 Treating the language network as an integrated system is reasonable given that the regions of this
1078 network a) show similar functional profiles, both with respect to selectivity for language over
1079 non-linguistic processes (e.g., Fedorenko et al., 2011; Pritchett et al., 2018; Jouravlev et al.,
1080 2019; Ivanova et al., 2020, 2021) and with respect to their role in lexico-semantic and syntactic
1081 processing (e.g., Fedorenko et al., 2012b; Blank et al., 2016; Fedorenko et al., 2020); and b)
1082 exhibit strong inter-region correlations in both their activity during naturalistic cognition
1083 paradigms (e.g., Blank et al., 2014; Braga et al., 2020; Paunov et al., 2019) and key functional
1084 markers, like the strength or extent of activation in response to language stimuli (e.g., Mahowald
1085 & Fedorenko, 2016; Mineroff, Blank et al., 2018). However, because we want to allow for the
1086 possibility that language regions differ in their response to music, we supplement the network-
1087 wise analyses with the analyses of the five language fROIs separately.
1088
1089 For each network-wise analysis, we fit a linear mixed-effect regression model predicting the
1090 level of BOLD response in the language fROIs in the contrasted conditions. The model included
1091 a fixed effect for condition and random intercepts for fROI and participant. Here and elsewhere,
1092 the p-value was estimated by applying the Satterthwaite's method-of-moment approximation to
1093 obtain the degrees of freedom (Giesbrecht & Burns, 1985; Fai & Cornelius, 1996; as described in
1094 Kuznetsova et al., 2017).
1095
1096 $$Effect\ size \sim condition + (1 \mid fROI) + (1 \mid SubjectID)$$
1097
1098 For each fROI-wise analysis, we fit a linear mixed-effect regression model predicting the level of
1099 BOLD response in each of the five language fROIs in the contrasted conditions. The model
1100 included a fixed effect for condition and a random intercept for participant. For each analysis, the
1101 result was FDR-corrected for the five fROIs.

1102

1103 $$Effect\ size \sim\ condition + (1\ |\ SubjectID)$$

1104

1105

1106 **Data availability**

1107

1108 The datasets generated during and/or analyzed during the current study are available in the OSF
1109 repository, https://osf.io/68y7c/.

1110

1111

1112 **Code availability**

1113

1114 Scripts for statistical analysis are available at: https://osf.io/68y7c/.

1115

1116

References

Alcock, K. J., Wade, D., Anslow, P., & Passingham, R. E. (2000). Pitch and timing abilities in adult left-hemisphere-dysphasic and right-hemisphere-damaged subjects. *Brain. Lang.*, 75(1), 47-65.

Amalric, M., & Dehaene, S. (2018). Cortical circuits for mathematical knowledge: evidence for a major subdivision within the brain's semantic networks. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 373(1740), 20160515.

Apperly, I. A., Samson, D., Carroll, N., Hussain, S., & Humphreys, G. (2006). Intact first-and second-order false belief reasoning in a patient with severely impaired grammar. *Soc. Neurosci.*, 1(3-4), 334-348.

Assem, M., Glasser, M. F., Van Essen, D. C., & Duncan, J. (2020). A domain-general cognitive core defined in multimodally parcellated human cortex. *Cereb. Cortex.*, 30(8), 4361-4380.

Baillet, S. (2014) Forward and Inverse Problems of MEG/EEG. In: Jaeger, D., Jung, R. (Eds.) Encyclopedia of Computational Neuroscience. New York, NY: Springer.

Baroni, M., Maguire, S., & Drabkin, W. (1983). The concept of musical grammar. *Music. Anal.*, 2(2), 175-208.

Basso, A., & Capitani, E. (1985). Spared musical abilities in a conductor with global aphasia and ideomotor apraxia. *J.Neurol. Neurosurg. Psychiatry.*, 48(5), 407-412.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.*, 67(1), 1–48.

Bautista, A., & Wilson, S. M. (2016). Neural responses to grammatically and lexically degraded speech. *Lang. Cogn. and Neurosci.*, 31(4), 567-574.

Ben-Shachar, M., Hendler, T., Kahn, I., Ben-Bashat, D., & Grodzinsky, Y. (2003). The neural reality of syntactic transformations: Evidence from functional magnetic resonance imaging. *Psychol. Sci.*, 14(5), 433-440.

Benn, Y.*, Ivanova, A.*, Clark, O., Mineroff, Z., Seikus, C., Santos Silva, J., Varley, R.^, & Fedorenko, E.^ (2021). No evidence for a special role of language in feature-based categorization. *bioRxiv*.

Bernstein, L. (1976). The unanswered question: Six talks at Harvard. Cambridge, MA: Harvard University Press.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain. Cogn.*, 77(1), 1-10.

Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D. A., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition.*, 81(1), B11-B20.

Bishop, D.V.M., & Norbury, C.F. (2002). Exploring the borderlands of autistic disorder and specific language impairment: a study using standardised diagnostic instruments. *J. Child. Psychol. Psychiatry*, 43(7), 917-29.

Blank, I., Kanwisher, N. & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *J. Neurophysiol.*, 112(5), 1105-1118.

Blank, I., Balewski, Z., Mahowald, K. & Fedorenko, E. (2016). Syntactic processing is distributed across the language system. *Neuroimage*, 127, 307-323.

Blank, I. & Fedorenko, E. (2017). Domain-general brain regions do not track linguistic input as closely as language-selective regions. *J. Neurosci.*, 37(41), 9999-10011.

Boebinger, D., Norman-Haignere, S., McDermott, J., & Kanwisher, N. (2020). Cortical music selectivity does not require musical training. https://www.biorxiv.org/content/10.1101/2020.01.10.902189v1

Boilès, C. L. (1973). Reconstruction of proto-melody. *Anuario Interamericano de Investigacion Musical*, 9, 45-63.

Bortolini, U., Leonard, L.B., & Caselli, M.C. (1998). Specific Language Impairment in Italian and English: evaluating alternative accounts of grammatical deficits. *Lang. Cogn. Process.*, 13(1), 1-20.

Braga, R. M., DiNicola, L. M., Becker, H. C., & Buckner, R. L. (2020). Situating the left-lateralized language network in the broader organization of multiple specialized large-scale distributed networks. *J. Neurophysiol.*, 124(5), 1415-1448.

Brust, J. C. (1980). Music and language: musical alexia and agraphia. *Brain.*, 103(2), 367-392.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. Journal of cognition, 1(1).

Bybee, J. (2010). Language, usage and cognition. Cambridge: Cambridge University Press.

Caplan, D., Hildebrandt, N., & Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain.*, 119(3), 933-949.

Caplan, D., Stanczak, L., & Waters, G. (2008). Syntactic and thematic constraint effects on blood oxygenation level dependent signal correlates of comprehension of relative clauses. *J. Cogn. Neurosci.*, 20(4), 643-656.

1209

1210 Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in
1211   language comprehension: Evidence from aphasia. *Brain. Lang.*, 3(4), 572-582.

1212

1213 Chen, G., Taylor, P. A., & Cox, R. W. (2017). Is the statistic value all we should care about in
1214   neuroimaging?. *Neuroimage*, 147, 952-959.

1215

1216 Chiappetta, B., Patel, A. D., & Thompson, C. K. (2022). Musical and linguistic syntactic
1217   processing in agrammatic aphasia: An ERP study. *Journal of Neurolinguistics*, 62, 101043.

1218

1219 Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge, MA: MIT press.

1220

1221 Chomsky, N. (1995). The minimalist program. Cambridge, MA: MIT Press.

1222

1223 Cooke, A., Grossman, M., DeVita, C., Gonzalez-Atavales, J., Moore, P., Chen, W., Gee, J., &
1224   Detre, J. (2006). Large-scale neural network for sentence processing. *Brain. Lang.*, 96(1),
1225   14-36.

1226

1227 Cooper, R. (1973) Propositions pour un modele transformationnel de description musicale.
1228   *Musique en Jeu*, 10, 70-88.

1229

1230 Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in
1231   the brain. *Nat. Rev. Neurosci.*, 3(3), 201-215.

1232

1233 Corlett, P. R., Mollick, J. A., & Kober, H. (2021). Substrates of Human Prediction Error for
1234   Incentives, Perception, Cognition, and Action. https://doi.org/10.31234/osf.io/pf89k

1235

1236 Crawford, J. R., & Garthwaite, P. H. (2007). Comparison of a single case to a control or
1237   normative sample in neuropsychology: Development of a Bayesian approach. *Cogn.*
1238   *Neuropsychol.*, 24(4), 343–372.

1239

1240 Creel, S. C., Weng, M., Fu, G., Heyman, G. D., & Lee, K. (2018). Speaking a tone language
1241   enhances musical pitch perception in 3–5‐year‐olds. *Dev. Sci.*, 21(1), e12503.

1242

1243 Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical
1244   Turk as a tool for experimental behavioral research. *PLoS One*, 8(3), e57410.

1245

1246 Dale, A. M. (1999). Optimal experimental design for event‐related fMRI. *Hum. Brain Mapp.*,
1247   8(2-3), 109-114.

1248

1249 Darwin C. *The Descent of Man, and Selection in Relation to Sex*. London: John Murray; 1871.

1250

1251 Dasgupta, I., & Gershman, S. J. (2021). Memory as a Computational Resource. *Trends. Cogn.*
1252   *Sci*., 25(3), 240-251.

1253

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cereb. Cortex.*, 25(11), 4596-4609.

Deen, B., & Freiwald, W. A. (2021). Parallel systems for social and spatial reasoning within the cortical apex. *bioRxiv*.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *J. Acoust. Soc. Am.*, 119(2), 719-722.

Deutsch, D., Dooley, K., Henthorn, T., & Head, B. (2009). Absolute pitch among students in an American music conservatory: Association with tone language fluency. *J. Acoust. Soc. Am.*, 125(4), 2398-2403.

Diachek, E.*, Blank, I.*, Siegelman, M.*, Affourtit, J. & Fedorenko, E. (2020). The domain-general multiple demand (MD) network does not support core aspects of language comprehension: a large-scale fMRI investigation. *J. Neurosci.*, 40(23), 4536–4550.

Dick, F., Bates, E., Wulfeck, B., Utman, J. A., Dronkers, N., & Gernsbacher, M. A. (2001). Language deficits, localization, and grammar: evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychol. Rev.*, 108(4), 759-788.

Ding, J., Martin, R. C., Hamilton, A. C., & Schnur, T. T. (2020). Dissociation between frontal and temporal-parietal contributions to connected speech in acute stroke. *Brain.*, 143(3), 862-876.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends. Neurosci.*, 23(10), 475-483.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nat. Rev. Neurosci.*, 2(11), 820-829.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends. Cogn. Sci.*, 14(4), 172-179.

Duncan, J. (2013). The structure of cognition: attentional episodes in mind and brain. *Neuron.*, 80(1), 35-50.

Embick, D., Marantz, A., Miyashita, Y., O'Neil, W., & Sakai, K. L. (2000). A syntactic specialization for Broca's area. *Proc. Natl. Acad. Sci. U.S.A.*, 97(11), 6150-6154.

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Ann. N. Y. Acad. Sci.*, 1169(1), 448-58.

1299 Fancourt, A. (2013). Exploring musical cognition in children with Specific Language
1300   Impairment. Doctoral thesis, Goldsmiths, University of London.
1301

1302 Fay, T. (1971). Perceived hierarchic structure in language and music. *J. Music. Theory.*, 15(1/2),
1303   112-137.
1304

1305 Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration
1306   in language and music: Evidence for a shared system. *Mem. Cognit.*, 37(1), 1-9.
1307

1308 Fedorenko, E., Hsieh, P.-J., Nieto-Castañon, A., Whitfield-Gabrieli, S. & Kanwisher, N. (2010).
1309   A new method for fMRI investigations of language: Defining ROIs functionally in
1310   individual subjects. *J. Neurophysiol.*, 104(2), 1177-94.
1311

1312 Fedorenko, E., Behr, M. & Kanwisher, N. (2011). Functional specificity for high-level linguistic
1313   processing in the human brain. *Proc. Natl. Acad. Sci. U.S.A.*, 108(39), 16428-16433.
1314

1315 Fedorenko, E., Duncan, J. & Kanwisher, N. (2012a). Language-selective and domain-general
1316   regions lie side by side within Broca's area. *Curr. Biol.*, 22(21), 2059-2062.
1317

1318 Fedorenko, E., Nieto-Castañon, A. & Kanwisher, N. (2012b). Lexical and syntactic
1319   representations in the brain: An fMRI investigation with multi-voxel pattern analyses.
1320   *Neuropsychologia.*, 50(4), 499-513.
1321

1322 Fedorenko, E., McDermott, J., Norman-Haignere, S. & Kanwisher, N. (2012c). Sensitivity to
1323   musical structure in the human brain. *J. Neurophysiol.*, 108(12), 3289-3300.
1324

1325 Fedorenko, E., Duncan, J. & Kanwisher, N. (2013). Broad domain-generality in focal regions of
1326   frontal and parietal cortex. *Proc. Natl. Acad. Sci. U.S.A.*, 110(41), 16616-16621.
1327

1328 Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension.
1329   *Front. Psychol.*, 5, 335.
1330

1331 Fedorenko, E., & Thompson-Schill, S. L. (2014). Reworking the language network. *Trends.*
1332   *Cogn. Sci*., 18(3), 120-126.
1333

1334 Fedorenko, E. & Varley, R. (2016). Language and thought are not the same thing: Evidence from
1335   neuroimaging and neurological patients. *Ann. N. Y. Acad. Sci.*, 1369(1), 132-153.
1336

1337 Fedorenko, E. & Blank, I. (2020). Broca's Area Is Not a Natural Kind. *Trends. Cogn. Sci*., 24(4),
1338   270-284.
1339

1340 Fedorenko, E., Blank, I., Siegelman, M. & Mineroff, Z. (2020). Lack of selectivity for syntax
1341   relative to word meanings throughout the language network. *Cognition.*, 203, 104348.
1342

Fedorenko, E. (2020). The brain network that supports high-level language processing. In
       Gazzaniga, Ivry, Mangun (Ed.), Cognitive Neuroscience: The Biology of the Mind (5th
       edition). Cambridge, MA: MIT Press.

Fedorenko, E. (2021). The early origins and the growing popularity of the individual-subject
       analytic approach in human neuroscience. *Curr. Opin. Behav. Sci.*, in press.

Fedorenko, E., & Shain, C. (2021). Similarity of computations across domains does not imply
       shared implementation: The case of language comprehension. *Current Directions in
       Psychological Science*, 30(6), 526-534.

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B. T., Mohlberg, H.,
       Amunts, K., & Zilles, K. (2008). Cortical folding patterns and predicting cytoarchitecture.
       Cereb. Cortex., 18(8), 1973-1980.

Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action:
       Lashley revisited. *Ann. N. Y. Acad. Sci.*, 1316(1), 87-104.

Fodor, J. D. (1983). Phrase structure parsing and the island constraints. *Linguist. Philos.*, 6(2),
       163-223.

Fouragnan, E., Retzler, C., & Philiastides, M. G. (2018). Separate neural representations of
       prediction error valence and surprise: Evidence from an fMRI meta-analysis. *Hum. Brain.
       Mapp.*, 39(7), 2887-2906.

Franklin, S., Turner, J.E., Ellis, A.W. (1992). ADA Comprehension Battery. Action for
       Dysphasic Adults, Canterbury House, Royal Street, London SE1 7LL.

Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. D., & Von Cramon, D. Y.
       (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cereb.
       Cortex.*, 16(12), 1709-1717.

Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and
       intelligibility in auditory language comprehension. *Hum. Brain. Mapp.*, 31(3), 448-457.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function.
       Physiol. Rev., 91(4), 1357-1392.

Friederici, A. D. (2018). The neural basis for human syntax: Broca's area and beyond. *Curr.
       Opin. Behav. Sci.*, 21, 88-92.

Frost, M. A., & Goebel, R. (2012). Measuring structural–functional correspondence: spatial
       variability of specialised brain regions after macro-anatomical alignment. *Neuroimage*,
       59(2), 1369-1381.

1388 Giesbrecht, F., & Burns, J. (1985). Two-Stage Analysis Based on a Mixed Model: Large-Sample
1389   Asymptotic Theory and Small-Sample Simulation Results. *Biometrics., 41*(2), 477-486.

1390

1391 Goldberg, A. E. (2002). "Construction Grammar." *Encyclopedia of Cognitive Science*.
1392   Macmillan Reference Limited Nature Publishing Group.

1393

1394 Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York:
1395   Wiley.

1396

1397 Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: memory as an
1398   integral component of information processing. *Trends. Cogn. Sci.*, *19*(6), 304-313.

1399

1400 Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity:
1401   behavior, function, and structure. *Neuron.*, 76(3), 486-502.

1402

1403 Herrmann, B., Obleser, J., Kalberlah, C., Haynes, J. D., & Friederici, A. D. (2012). Dissociable
1404   neural imprints of perception and grammar in auditory functional imaging. *Hum. Brain.*
1405   *Mapp.*, 33(3), 584-595.

1406

1407 Hoch, L., Poulin-Charronnat, B., & Tillmann, B. (2011). The influence of task-irrelevant music
1408   on language processing: syntactic and semantic structures. *Front. Psychol.*, 2, 112.

1409

1410 Hrong-Tai Fai, A., & Cornelius, P. L. (1996). Approximate F-tests of multiple degree of freedom
1411   hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *J.*
1412   *Stat. Comput. Simul.*, *54*(4), 363-378.

1413

1414 Ivanova, A., Srikant, S., Sueoka, Y., Kean, H., Dhamala, R., O'Reilly, U-M., Bers, M. U., &
1415   Fedorenko, E. (2020). Comprehension of computer code relies primarily on domain-general
1416   executive resources. *eLife*, 9:e58906.

1417

1418 Ivanova, A., Mineroff, Z., Zimmerer, V., Kanwisher, N., Varley, R. & Fedorenko, E. (2021) The
1419   language network is recruited but not required for non-verbal semantic processing.
1420   https://www.biorxiv.org/content/10.1101/696484v1.

1421

1422 Jackendoff, R. (2002). English particle constructions, the lexicon, and the autonomy of syntax. In
1423   Dehé, N., Jackendoff, R., McIntyre, A., & Urban, S. (Eds.) *Verb-particle explorations*, (pp.
1424   67-94). Berlin: De Gruyter.

1425

1426 Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Res.*,
1427   1146, 2-22.

1428

1429 Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Percept.*,
1430   26(3), 195-204.

1431

1432 Jackendoff, R., & Audring, J. (2020). The texture of the lexicon: relational morphology and the
1433   parallel architecture. Oxford: Oxford University Press.

1434

1435 Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts
1436   in music. *J. Cogn. Neurosci.*, 7(2), 153-164.

1437

1438 Jouravlev, O., Zheng, D., Balewski, Z., Pongos, A., Levan, Z., Goldin-Meadow, S., &
1439   Fedorenko, E. (2019). Speech-accompanying gestures are not processed by the language-
1440   processing mechanisms. *Neuropsychologia*, 132, 107132.

1441

1442 Jouravlev, O., Kell, A., Mineroff, Z., Haskins, A.J., Ayyash, D., Kanwisher, N. & Fedorenko, E.
1443   (2020). Reduced language lateralization in autism and the broader autism phenotype as
1444   assessed with robust individual-subjects analyses. *Autism Res.*, 13(10), 1746-1761.

1445

1446 Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain
1447   activation modulated by sentence comprehension. *Science*, 274(5284), 114-116.

1448

1449 Kaplan, E., Goodglass, H., & Weintraub, S. (2001). Boston Naming Test. 2nd Ed. Philadelphia,
1450   PA: Lippincott Williams & Wilkins.

1451

1452 Kay, J., Lesser, R., & Coltheart, M. (1992). Psycholinguistic Assessments of Language
1453   Processing in Aphasia (PALPA). Hove: Erlbaum.

1454

1455 Kell, A. J., Yamins, D. L., Shook, E. N., Norman-Haignere, S. V., & McDermott, J. H. (2018). A
1456   task-optimized neural network replicates human auditory behavior, predicts brain responses,
1457   and reveals a cortical processing hierarchy. *Neuron*, 98(3), 630-644.

1458

1459 Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). The neural bases of sentence
1460   comprehension: a fMRI examination of syntactic and lexical processing. *Cereb. Cortex*,
1461   11(3), 223-237.

1462

1463 Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music
1464   processing: "nonmusicians" are musical. *J. Cogn. Neurosci.*, 12(3), 520-541.

1465

1466 Koelsch, S., Gunter, T. C., Schröger, E., Tervaniemi, M., Sammler, D., & Friederici, A. D.
1467   (2001). Differentiating ERAN and MMN: an ERP study. *NeuroReport*, 12(7), 1385-1389.

1468

1469 Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D.
1470   (2002). Bach speaks: A cortical" language-network" serves the processing of music.
1471   *Neuroimage*, 17(2), 956-966.

1472

1473 Koelsch, S. (2006). Significance of Broca's area and ventral premotor cortex for music-syntactic
1474   processing. *Cortex*, 42(4), 518-520.

1475

1476 Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical
1477   syntactic structure in music. *Proc. Natl. Acad. Sci. U.S.A.*, 110(38), 15443-15448.

1478

1479 Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in
1480   systems neuroscience: the dangers of double dipping. *Nat. Neurosci.*, 12(5), 535.
1481
1482 Krumhansl, C. L., & Keil, F. C. (1982). Acquisition of the hierarchy of tonal functions in music.
1483   *Mem. Cognit.*, 10(3), 243-251.
1484
1485 Kunert, R., & Slevc, L. R. (2015). A Commentary on:"Neural overlap in processing music and
1486   speech". *Front. Hum. Neurosci.*, 9, 330.
1487
1488 Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and
1489   language syntax interact in Broca's area: An fMRI study. *PLoS One*, 10(11), e0141069.
1490
1491 Kunert, R., Willems, R. M., & Hagoort, P. (2016). Language influences music harmony
1492   perception: effects of shared syntactic integration resources beyond attention. *R. Soc. Open*
1493   *Sci.*, 3(2), 150685.
1494
1495 Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003).
1496   Distinct patterns of neural modulation during the processing of conceptual and syntactic
1497   anomalies. *J. Cogn. Neurosci.*, 15(2), 272-293.
1498
1499 Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear
1500   mixed effects models. *J. Stat. Softw.*, 82(13), 1-26.
1501
1502 LaCroix, A., Diaz, A. F., & Rogalsky, C. (2015). The relationship between the neural
1503   computations for speech and music perception is context-dependent: an activation likelihood
1504   estimate study. *Front. Psychol.*, 6, 1138.
1505
1506 Lerdahl, F., & Jackendoff, R. (1977). Toward a formal theory of tonal music. *J. Music Theory*,
1507   21(1), 111-171.
1508
1509 Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music*
1510   *Percept.*, 1(2), 229-252.
1511
1512 Levin, B., & Rappaport-Hovav, M. (2005). Argument realization. Cambridge: Cambridge
1513   University Press.
1514
1515 Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the
1516   brain: a possible role for Brodmann Area 47 in temporal coherence. *Neuroimage*, 20(4),
1517   2142-2152.
1518
1519 Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical
1520   structure in so-called agrammatic aphasics. *Cognition*, 13(3), 361-392.
1521
1522 Lindblom, B., & Sundberg, J. (1969). Towards a generative theory of melody. Speech
1523   Transmission Laboratory. *Quarterly Progress and Status Reports*, 10, 53-86.
1524

Luria, A. R., Tsvetkova, L. S., & Futer, D. S. (1965). Aphasia in a composer. *J. Neurol. Sci.*, 2(3), 288-292.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci.*, 4(5), 540-545.

Mahowald, K. & Fedorenko, E. (2016). Reliable individual-level neural markers of high-level language processing: A necessary precursor for relating neural variability to behavioral and genetic variability. *Neuroimage*, 139, 74-93.

Makowski, (2018). The psycho Package: an Efficient and Publishing-Oriented Workflow for Psychological Science. *J. Open Source Softw.*, 3(22), 470, https://doi.org/10.21105/joss.00470

Malik-Moraleda, S.*, Ayyash, D.*, Gallée, J., Affourtit, J., Hoffman, M., Mineroff, Z., Jouravlev, O., & Fedorenko, E. (2021). The universal language network: A cross-linguistic investigation spanning 45 languages and 11 language families. *bioRxiv*.

Marin, O.S.M. (1982). Neurological Aspects of Music Perception and Performance. New York: Academic Press.

Matchin, W., & Hickok, G. (2020). The cortical organization of syntax. *Cereb. Cortex*, 30(3), 1481-1498.

Mehr, S., Krasnow, M., Bryant, G., & Hagen, E. (2020). Origins of music in credible signaling. *Behav. Brain Sci.*, 1-41.

Mesulam, M. M., Rogalski, E. J., Wieneke, C., Hurley, R. S., Geula, C., Bigio, E. H., Thompson, C. K., & Weintraub, S. (2014). Primary progressive aphasia and the evolving neurology of the language network. *Nat. Rev. Neurol.*, 10(10), 554.

Mesulam, M. M., Thompson, C. K., Weintraub, S., & Rogalski, E. J. (2015). The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia. *Brain*, 138(8), 2423-2437.

McDermott, J. (2008). The evolution of music. *Nature*, 453(7193), 287-288.

Mineroff, Z.*, Blank, I.*, Mahowald, K. & Fedorenko, E. (2018). A robust dissociation among the language, multiple demand, and default mode networks: evidence from inter-region correlations in effect size. *Neuropsychologia*, 119, 501-511.

Mollica, F., Shain, C., Affourtit, J., Kean, H., Siegelman, M., & Fedorenko, E. (2020), October 21-24). Another look at the constituent structure of sentences in the human brain [Poster presentation]. SNL 2020, virtual.

Monti, M. M., Parsons, L. M., & Osherson, D. N. (2009). The boundaries of language and thought in deductive inference. *Proc. Natl. Acad. Sci. U.S.A.*, 106(30), 12554-12559.

Monti, M. M., Parsons, L. M., & Osherson, D. N. (2012). Thought beyond language: Neural dissociation of algebra and natural language. *Psychol. Sci.*, 23(8), 914-922.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage*, 13(4), 684-701.

Musso, M., Weiller, C., Horn, A., Glauche, V., Umarova, R., Hennig, J., Schneider, A., & Rijntjes, M. (2015). A single dual-stream framework for syntactic computations in music and language. *Neuroimage*, 117, 267-283.

Newman, A. J., Pancheva, R., Ozawa, K., Neville, H. J., & Ullman, M. T. (2001). An event-related fMRI study of syntactic and semantic violations. *J. Psycholinguist. Res.*, 30(3), 339-364.

Ngo, M. K., Vu, K. P. L., & Strybel, T. Z. (2016). Effects of music and tonal language experience on relative pitch performance. *Am. J. Psychol.*, 129(2), 125-134.

Nieto-Castañon, A. & Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage*, 63(3), 1646-1669.

Norman-Haignere, S., Kanwisher, N., & McDermott, J. H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *J. Neurosci.*, 33(50), 19451-19469.

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, 88(6), 1281-1296.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1), 97-113.

Pallier, C., Devauchelle, A. D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proc. Natl. Acad. Sci. U.S.A.*, 108(6), 2522-2527.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *J. Cogn. Neurosci.*, 10(6), 717-733.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.*, 6(7), 674-681.

Patel, A. D. (2008). Music, Language, and the Brain. Oxford: Oxford University Press.

Patel, A. D., Iversen, J. R., Wassenaar, M., & Hagoort, P. (2008). Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology*, 22(7-8), 776-789.

Patel, A.D. (2012). Language, music, and the brain: a resource-sharing framework. In: P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and Music as Cognitive Systems* (pp. 204-223). Oxford: Oxford University Press.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4), 767-776.

Paunov, A., Blank, I. A., & Fedorenko, E. (2019). Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *J. Neurophysiol.*, 121, 1244-1265.

Paunov, A., Blank, I., Jouravlev, O., Mineroff, Z., Gallée, J., & Fedorenko, E. (2021). Differential tracking of linguistic vs. mental state content in naturalistic stimuli by language and Theory of Mind (ToM) brain networks. *bioRxiv*.

Peelle, J. E., Troiani, V., Wingfield, A., & Grossman, M. (2010). Neural processing during older adults' comprehension of spoken sentences: age differences in resource allocation and connectivity. *Cereb. Cortex*, 20(4), 773-782.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J. Neurosci.*, 24(30), 6810-6815.

Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, 113(4), 1185-1205.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, 117(6), 1283-1301.

Peretz, I., Belleville, S., & Fontaine, S. (1997). Dissociations between music and language functions after cerebral resection: a new case of amusia without aphasia. *Canadian journal of experimental psychology*, 51(4), 354-368.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: the Montreal Battery of Evaluation of Amusia. *Ann. N. Y. Acad. Sci.*, 999(1), 58-75.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nat. Neurosci.*, 6(7), 688-691.

Peretz, I., Vuvan, D., Lagrois, M. É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 370(1664), 20140090.

Perrachione, T. K., Fedorenko, E. G., Vinke, L., Gibson, E., & Dilley, L. C. (2013). Evidence for shared cognitive processing of pitch in music and language. *PLoS One*, 8(8), e73372.

Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychon. Bull. Rev.*, 20(2), 310-317.

Piccirilli, M., Sciarma, T., & Luzzi, S. (2000). Modularity of music: evidence from a case of pure amusia. *Journal of Neurology, Neurosurgery & Psychiatry*, 69(4), 541-545.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28(1-2), 73-193.

Pinker, S. (1991). Rules of language. *Science*, 253(5019), 530-535.

Pinker, S. (1994), The Language Instinct: How the Mind Creates Language, New York: Harper Collins Publishers, Inc

Pinker, S. (1999). Out of the minds of babes. *Science*, 283(5398), 40-41.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data?. *Trends. Cogn. Sci.*, 10(2), 59-63.

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron*, 72(5), 692-697.

Polk, M., & Kertesz, A. (1993). Music and language in degenerative disease of the brain. *Brain Cogn.*, 22(1), 98-117.

Poulin-Charronnat, B., Bigand, E., Madurell, F., & Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. *Cognition*, 94, B67-B78.

Pritchett, B., Hoeflin, C., Koldewyn, K., Dechter, E. & Fedorenko, E. (2018). High-level language processing regions are not engaged in action observation or imitation. *J. Neurophysiol.*, 120(5), 2555-2570.

Riemann, H. (1877). Musikalische Syntaxis: Grundriss einer harmonischen Satzbildungslehre. Leipzig: Breitkopf und Härtel.

Roads, C., & Wieneke, P. (1979). Grammars as representations for music. *Comput. Music J.*, 3(1), 48-55.

Roberts, I. (2012). Comments and a conjecture inspired by Fabb and Halle. In Rebuschat, P., Rohrmeier, M., Hawkins, J. A., & Cross, I. (Eds.) *Language and Music as Cognitive Systems* (pp. 51-66.). Oxford: Oxford University Press.

Röder, B., Stock, O., Neville, H., Bien, S., & Rösler, F. (2002). Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: a functional magnetic resonance imaging study. *Neuroimage*, 15(4), 1003-1014.

Rogalsky, C., & Hickok, G. (2011). The role of Broca's area in sentence comprehension. *J. Cogn. Neurosci.*, 23(7), 1664-1680.

Rueffler, C., Hermisson, J., & Wagner, G. P. (2012). Evolution of functional specialization and division of labor. *Proc. Natl. Acad. Sci. U.S.A.*, 109(6), E326-E335.

Sag, I., Wasow, T., & Bender, E. (2003). Formal syntax, an introduction. CSLI publication.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Elger, C. E., Friederici, A. D., Grigutsch, M., Huppertz, H.-J., Knosche, T. R., Wellmer, J., Widman, G., & Schulze-Bonhaged, A. (2009). Overlap of musical and linguistic syntax processing: intracranial ERP evidence. *Ann. N. Y. Acad. Sci.*, 1169(1), 494-498.

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

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H. J., Wellmer, J., Widman, G., Elger, C. E., Friederici, A. D., & Schulze-Bonhaged, A. (2013). Co-localizing linguistic and musical syntax with intracranial EEG. *Neuroimage*, 64, 134-146.

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2020). Music as a coevolved system for social bonding. *Behav. Brain Sci.*, 1-36.

Schmidt, S. (2009). Shall We Really Do It Again? The Powerful Concept of Replication Is Neglected in the Social Sciences. *Rev. Gen. Psychol.*, 13(2), 90-100.

Scott, T.L., Gallée, J., & Fedorenko, E. (2017). A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cogn. Neurosci.*, 8(3), 167-176.

Shain, C.*, Blank, I.*, Van Shijndel, M., Schuler, W. & Fedorenko, E. (2020). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, 138, 107307.

Shain, C., Blank, I., Fedorenko, E., Gibson, E., & Schuler, W. (2021a). Robust effects of working memory demand during naturalistic language comprehension in language-selective cortex. *bioRxiv*.

Shain, C., Kean, H., Lipkin, B., Affourtit, J., Siegelman, M., Mollica, F., & Fedorenko, E. (2021b). 'Constituent length' effects in fMRI do not provide evidence for abstract syntactic processing. *bioRxiv*.

1753 Shashidhara, S., Mitchell, D. J., Erez, Y., & Duncan, J. (2019). Progressive recruitment of the
1754   frontoparietal multiple-demand system with increased task complexity, time pressure, and
1755   reward. *J. Cogn. Neurosci.*, 31(11), 1617-1630.
1756
1757 Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-
1758   paced reading time evidence for shared processing of linguistic and musical syntax.
1759   *Psychon. Bull. Rev.*, 16(2), 374-381.
1760
1761 Slevc, L. R., Reitman, J., & Okada, B. (2013). Syntax in music and language: the role of
1762   cognitive control. In Proceedings of the Annual Meeting of the Cognitive Science Society
1763   (Vol. 35, No. 35).
1764
1765 Slevc, L. R., & Okada, B. M. (2015). Processing structure in language and music: a case for
1766   shared reliance on cognitive control. *Psychon. Bull. Rev.*, 22(3), 637-652.
1767
1768 Slevc, L. R., Faroqi-Shah, Y., Saxena, S., & Okada, B. M. (2016). Preserved processing of
1769   musical structure in a person with agrammatic aphasia. *Neurocase*, 22(6), 505-511.
1770
1771 Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic
1772   comprehension by positron emission tomography. *Brain. Lang.*, 52(3), 452-473.
1773
1774 Sueoka, Y., Paunov, A., Ivanova, A., Blank, I.A., & Fedorenko, E. (2022). The language
1775   network reliably 'tracks' naturalistic meaningful non-verbal stimuli.
1776
1777 Sundberg, J., & Lindblom, B. (1976). Generative theories in language and music descriptions.
1778   *Cognition*, 4(1), 99-122.
1779
1780 Swain, J. P. (1995). The concept of musical syntax. *Music. Q.*, 79(2), 281-308.
1781
1782 Tahmasebi, A. M., Davis, M. H., Wild, C. J., Rodd, J. M., Hakyemez, H., Abolmaesumi, P., &
1783   Johnsrude, I. S. (2012). Is the link between anatomical structure and function equally strong
1784   at all cognitive levels of processing?. *Cereb. Cortex.*, 22(7), 1593-1603.
1785
1786 Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*.
1787   Society for Industrial and Applied Mathematics.
1788
1789 Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in
1790   musical priming. *Cogn. Brain Res.*, 16(2), 145-161.
1791
1792 Tillmann, B., Koelsch, S., Escoffier, N., Bigand, E., Lalitte, P., Friederici, A. D., & von Cramon,
1793   D. Y. (2006). Cognitive priming in sung and instrumental music: activation of inferior
1794   frontal cortex. *Neuroimage*, 31(4), 1771-1782.
1795
1796 Tillmann, B. (2012). Music and Language Perception: Expectations, Structural Integration, and
1797   Cognitive Sequencing. *Top. Cogn. Sci.*, 4(4), 568-584.
1798

Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B. J., Zhuang, J., Papoutsi, M., & Stamatakis, E. A. (2011). Left inferior frontal cortex and syntax: function, structure and behaviour in patients with left hemisphere damage. *Brain*, 134(2), 415-431.

Van de Cavey, J., & Hartsuiker, R. J. (2016). Is there a domain-general cognitive structuring system? Evidence from structural priming across music, math, action descriptions, and language. *Cognition*, 146, 172-184.

Varley, R., & Siegal, M. (2000). Evidence for cognition without grammar from causal reasoning and 'theory of mind' in an agrammatic aphasic patient. *Curr. Biol.*, 10(12), 723-726.

Varley, R. A., Klessinger, N. J., Romanowski, C. A., & Siegal, M. (2005). Agrammatic but numerate. *Proc. Natl. Acad. Sci. U.S.A.*, 102(9), 3519-3524.

Vázquez-Rodríguez, B., Suárez, L. E., Markello, R. D., Shafiei, G., Paquola, C., Hagmann, P., van den Heuvel, M. P., Bernhardt, B. C., Spreng, R. N. & Misic, B. (2019). Gradients of structure–function tethering across neocortex. *Proc. Natl. Acad. Sci. U.S.A.*, 116(42), 21219-21227.

Wehbe, L., Blank, I., Shain, C., Futrell, R., Levy, R., Malsburg, T. Smith, N., Gibson, E., Fedorenko, E. (2021). Incremental language comprehension difficulty predicts activity in the language network but not the multiple demand network https://doi.org/10.1101/2020.04.15.043844.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. Journal of Experimental Psychology: General, 143(5), 2020-2045. http://dx.doi.org/10.1037/xge0000014

Willems, R. M., Van der Haegen, L., Fisher, S. E., & Francks, C. (2014). On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nat. Rev. Neurosci.*, 15(3), 193-201.

Wilson, S. M., & Saygın, A. P. (2004). Grammaticality judgment in aphasia: Deficits are not specific to syntactic structures, aphasic syndromes, or lesion sites. *J. Cogn. Neurosci.*, 16(2), 238-252.

Wilson, S. M., Galantucci, S., Tartaglia, M. C., & Gorno-Tempini, M. L. (2012). The neural basis of syntactic deficits in primary progressive aphasia. *Brain. Lang.*, 122(3), 190-198.

Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Appl. Psycholinguist.*, 28(4), 565.

Woolgar, A., Duncan, J., Manes, F., & Fedorenko, E. (2018). Fluid intelligence is supported by the multiple-demand system not the language system. *Nat. Hum. Behav.*, 2(3), 200-204.

1845   Zatorre, R. J. (1984). Musical perception and cerebral function: A critical review. *Music*
1846   *Percept.*, 2(2), 196-221.
1847