NEUROSCIENCE

Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody

Philippe Albouy[1,2,3]*, Lucas Benjamin[1], Benjamin Morillon[4]†, Robert J. Zatorre[1,2]†

Does brain asymmetry for speech and music emerge from acoustical cues or from domain-specific neural networks? We selectively filtered temporal or spectral modulations in sung speech stimuli for which verbal and melodic content was crossed and balanced. Perception of speech decreased only with degradation of temporal information, whereas perception of melodies decreased only with spectral degradation. Functional magnetic resonance imaging data showed that the neural decoding of speech and melodies depends on activity patterns in left and right auditory regions, respectively. This asymmetry is supported by specific sensitivity to spectrotemporal modulation rates within each region. Finally, the effects of degradation on perception were paralleled by their effects on neural classification. Our results suggest a match between acoustical properties of communicative signals and neural specializations adapted to that purpose.

Speech and music represent the most cognitively complex, and arguably uniquely human, use of sound. To what extent do these two domains depend on separable neural mechanisms? What is the basis for such specialization? Several studies have proposed that left hemisphere neural specialization of speech (*1*) and right hemisphere specialization of pitch-based aspects of music (*2*) emerge from differential analysis of acoustical cues in the left and right auditory cortices (ACs). However, domain-specific accounts suggest that speech and music are processed by dedicated neural networks, the lateralization of which cannot be explained by low-level acoustical cues (*3*–*6*).

Despite consistent empirical evidence in its favor, the acoustical cue account has been computationally underspecified: Concepts such as spectrotemporal resolution (*7*–*9*), time integration windows (*10*), and oscillations (*11*) have all been proposed to explain hemispheric specializations. However, it is difficult to test these concepts directly within a neurally viable framework, especially using naturalistic speech or musical stimuli. The concept of spectrotemporal receptive fields (*12*) provides a computationally rigorous and neurophysiologically plausible approach to the neural decomposition of acoustical cues. This model proposes that auditory neurons act as spectrotemporal modulation (STM) rate filters, based on both single-cell recordings in animals (*13*, *14*) and neuroimaging in humans (*15*, *16*). STM may provide a mechanistic basis to account for lateralization in AC (*17*), but a direct relationship among acoustical STM features, hemispheric asymmetry, and behavioral performance during processing of complex signals such as speech and music has not been investigated.

We created a stimulus set in which 10 original sentences were crossed with 10 original melodies, resulting in 100 naturalistic a cappella songs (Fig. 1) (stimuli are available at www.zlab.mcgill.ca/downloads/albouy_20190815/). This orthogonalization of speech and melodic domains across stimuli allows the dissociation of speech-specific (or melody-specific) from nonspecific acoustic features, thereby controlling for any potential acoustic bias (*3*). We created two separate stimulus sets, one with French and one with English sentences, to allow for reproducibility and to test generality across languages. We then parametrically degraded each stimulus selectively in either the temporal or spectral dimension using a manipulation that decomposes the acoustical signal using the STM framework (*18*).

We first investigated the importance of STM rates on sentence or melody recognition scores in a behavioral experiment (Fig. 2A). Native French (n = 27) and English (n = 22) speakers were presented with pairs of stimuli and asked to discriminate either the speech or the melodic content. Thus, the stimulus set across the two tasks was identical; only the instructions differed. The degradation of information in the temporal dimension impaired sentence recognition ($t_{(48)}$ = 13.61 < 0.001, one-sample t test against zero of the slope of the linear fit relating behavior to the degree of acoustic degradation) but not melody recognition ($t_{(48)}$ =

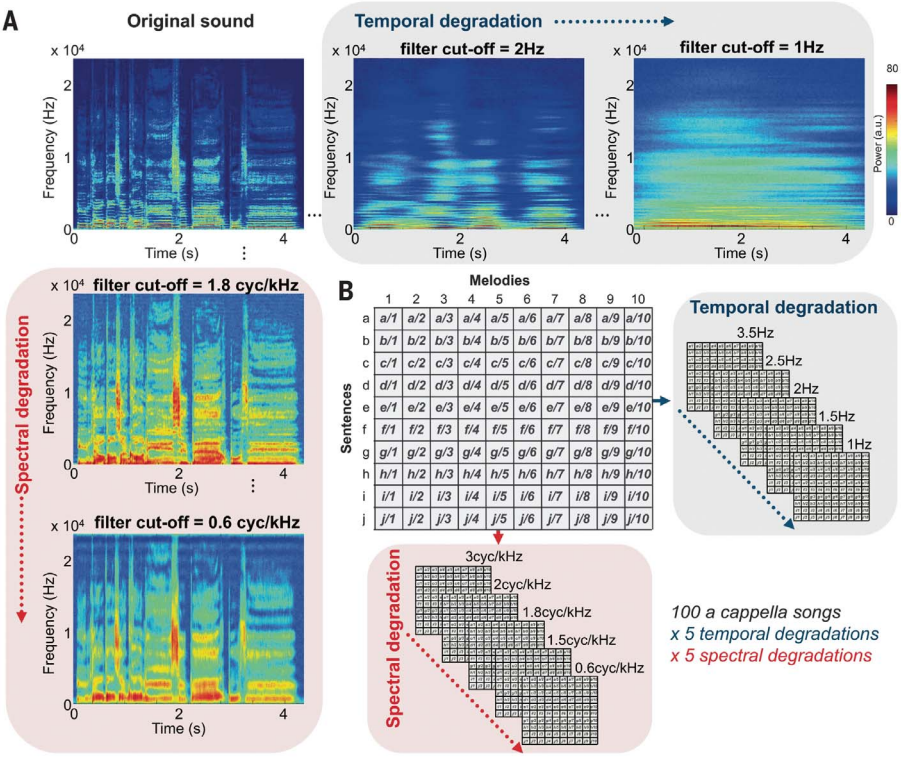

Fig. 1. Spectrotemporal filtering and stimulus set. (**A**) Spectral and temporal degradations applied on an original a cappella song. (**B**) One hundred a cappella songs in each language were recorded following a 10 × 10 matrix with 10 melodies (number code) and 10 sentences (letter code). Stimuli were then filtered either in the spectral or in the temporal dimension with five filter cutoffs, resulting in 1000 degraded stimuli for each language.

[1]Cognitive Neuroscience Unit, Montreal Neurological Institute, McGill University, Montreal, QC, Canada. [2]International Laboratory for Brain, Music and Sound Research (BRAMS); Centre for Research in Brain, Language and Music; Centre for Interdisciplinary Research in Music, Media, and Technology, Montreal, QC, Canada. [3]CERVO Brain Research Centre, School of Psychology, Laval University, Quebec, QC, Canada. [4]Aix Marseille University, Inserm, INS, Institut de Neurosciences des Systèmes, Marseille, France.
*Corresponding author. Email: philippe.albouy@psy.ulaval.ca
†These authors contributed equally to this work.

0.62, $p = 0.53$), whereas degradation of information in the spectral dimension impaired melody recognition ($t_{(48)} = 8.24 < 0.001$) but not sentence recognition ($t_{(48)} = -1.28$, $p = 0.20$; Fig. 2, B and C). This double dissociation was confirmed by a domain-by-degradation interaction (2 × 2 repeated-measures ANOVA: $F_{(1,47)} = 207.04$, $p < 0.001$). Identical results were observed for the two language groups (see fig. S2 and the supplementary results for complementary analyses).

We then investigated the impact of STM rates on the neural responses to speech and melodies using functional magnetic resonance imaging (fMRI). Blood oxygenation level–dependent (BOLD) activity was recorded while 15 French speakers who had participated in the behavioral experiment listened to blocks of five songs degraded either in the temporal or spectral dimension. Participants attended to either the speech or the melodic content (Fig. 3A). BOLD signal in bilateral ACs scaled with both temporal and spectral degradation cutoffs [i.e., parametric modulation with quantity of temporal or spectral information; $p < 0.05$ familywise error (FWE) corrected; Fig. 3B and table S1]. These regions were located lateral to primary ACs and correspond to the ventral auditory stream of information processing, covering both parabelt areas and the lateral anterior superior temporal gyrus [parabelt and auditory area 4 (A4) regions; see (*19*)], but there was no significant difference in the hemispheric response to either dimension (whole-brain two-sample t tests; all $p > 0.05$).

To investigate more fine-grained encoding of speech and melodic contents, we performed a multivariate pattern analysis on the fMRI data. Ten-category classifications (separately for melodies and sentences) using whole-brain searchlight analyses (support vector machine, leave-one-out cross-validation procedure, cluster corrected) revealed that the neural encoding of sentences significantly depends on neural activity patterns in left A4 [TE.3; subregion of AC; see (*19*)], whereas the neural decoding of melodies significantly depends on neural activity patterns in right A4 ($p < 0.05$ cluster corrected; Fig. 3, C and D, and table S1; other, subthreshold clusters are reported in fig. S3). To ensure that this effect was generalizable to the population, we performed a complementary information prevalence analysis within temporal lobe masks (see the materials and methods). For the decoding of sentences, a prevalence value of up to 70% was observed in left A4 ($p = 0.02$, corrected), whereas a prevalence value of up to 69% was observed for the decoding of melodies in right A4 ($p = 0.03$, corrected; see table S1). Finally, we tested whether the classification accuracy was better for sentence or melody in the right or the left hemisphere. We computed a lateralization index on accuracy scores [(R – L)/(R + L)] and observed a significant asymmetry in opposite directions for the two domains in region A4 (Fig. 3F, table S1, and fig. S4; $p < 0.05$, cluster corrected at the whole-brain level).

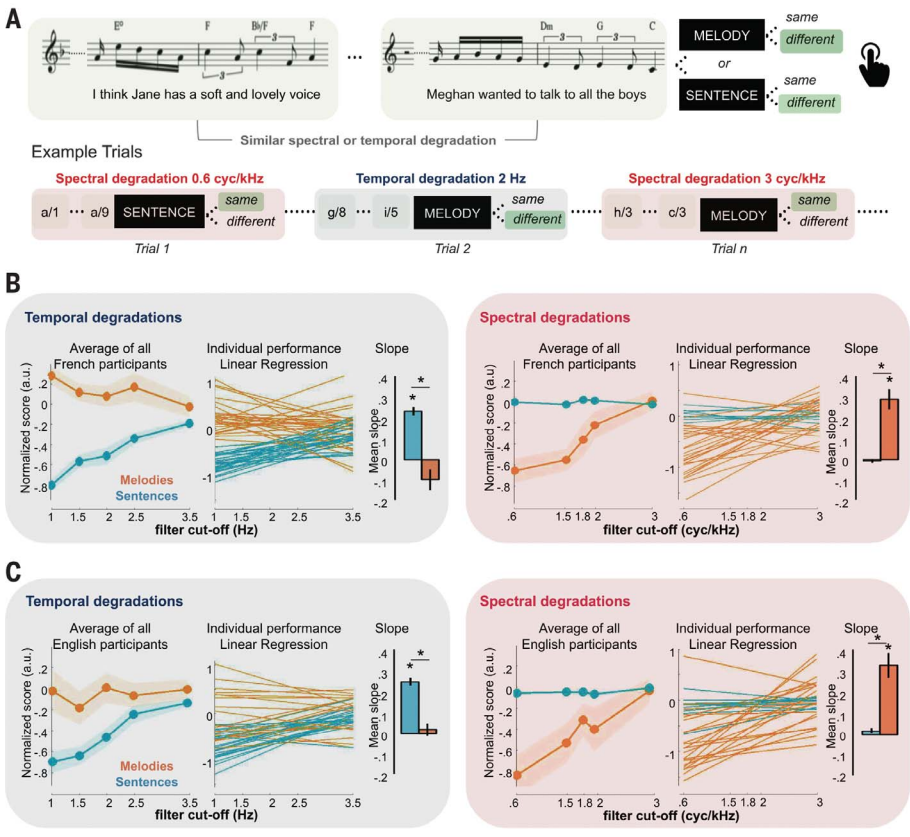

Fig. 2. Behavioral experiment. (**A**) Participants listened to degraded (either in the spectral or temporal dimension) a cappella songs presented in pairs. After the second song, a visual instruction indicated the domain of interest (sentences or melodies). Lower panel shows example trials. (**B**) Behavioral performance of French-speaking listeners. Aqua shading indicates temporal degradations and orange shading indicates spectral degradations. Average performance across participants (95% confidence interval) and individual performance modeled with linear regression are shown for both types of degradations. (**C**) Same as (B) but for English-speaking listeners. Error bars indicate SEM. Asterisks indicate significant differences.

We then tested the relationship between neural specialization of left and right hemispheres for speech and melodic contents and behavioral processing of these two domains. We estimated linear and nonlinear statistical dependencies by computing the normalized mutual information [NMI (*20*)] between the confusion matrices extracted from classification of neural data (whole brain, for each searchlight) and those from behavioral data recorded offline (for each participant and each domain). To investigate the correspondence between neural and behavioral patterns (pattern of errors) instead of mere accuracy (diagonal), these analyses were done after removing the diagonal information (Fig. 4A). NMI was significantly higher in left than right A4 for sentences, whereas the reverse pattern was observed for melodies, as measured by the lateralization index ($p < 0.05$, cluster corrected; see the materials and methods, table S1, and fig. S5).

We next tested whether the origin of the observed lateralization was related to attentional processes by investigating the decoding accuracy and NMI lateralization index as a function of attention to sentences or melodies. Whole-brain analyses did not reveal any significant cluster, suggesting that the previously observed hemispheric specialization is robust to attention and thus is more likely to be linked to automatic than to top-down processes (see fig. S6 and the supplementary results for details).

Finally, we investigated whether the hemispheric specialization for speech and melodic contents was directly related to a differential acoustic sensitivity of left and right ACs to STMs, as initially hypothesized. We estimated the impact of temporal or spectral degradations on decoding accuracy by computing the accuracy change (with negative indicating accuracy loss and positive indicating accuracy gain) between decoding accuracy computed on all trials (all degradation types) and on a

specific degradation type (temporal or spectral). We observed a domain-by-degradation interaction in bilateral ACs (left and right area A4; $p < 0.05$, cluster corrected; Fig. 4C and fig. S7). For sentences, accuracy loss was observed only in the left A4 for temporal as compared with spectral degradations ($p < 0.001$, Tukey corrected; all others, $p > 0.16$), whereas the reverse pattern was observed for melodies only in right A4 ($p = 0.003$, Tukey corrected; all others, $p > 0.29$).

This differential sensitivity to acoustical cues in left and right ACs was also observed in the brain–behavior relationship. We investigated the effect of degradations on the NMI lateralization index. We first show a significant domain-by-degradation interaction observed in area A4 ($p < 0.05$, cluster corrected; Fig. 4D, left; table S1; and fig. S8). The main effect of degradation (temporal > spectral) was then analyzed with two-sample t tests for each domain to reveal that the NMI lateralization index was affected in opposite directions by temporal and spectral degradations (A4 and

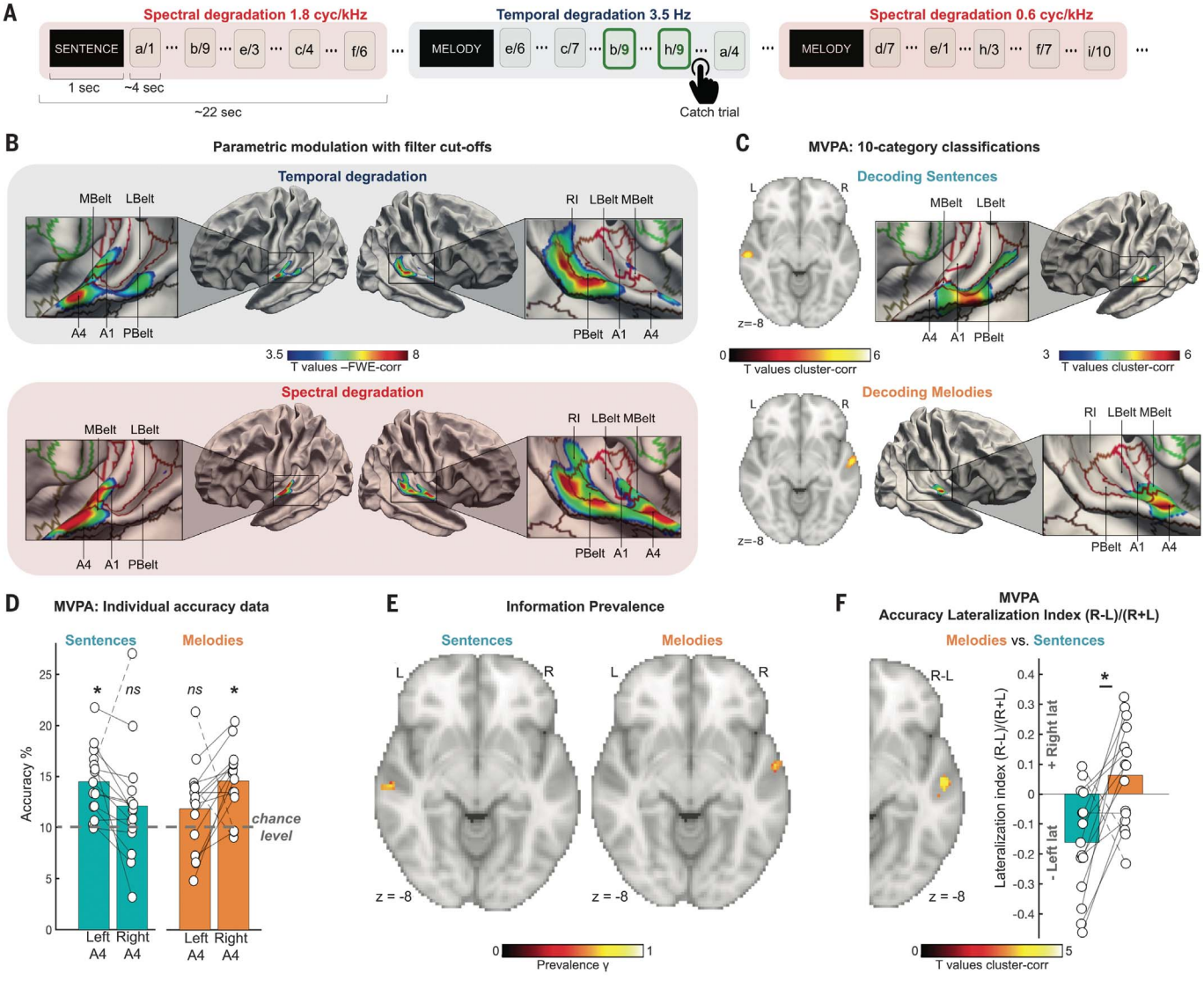

Fig. 3. fMRI experiment. (**A**) fMRI design. BOLD activity was collected while participants listened to blocks of five songs (degraded either in the temporal or spectral dimension). To control for attention, participants were asked to detect two catch trials (with high filter cutoffs containing melody (or sentence) repetition (1-back task). (**B**) Univariate fMRI results ($p < 0.05$, voxelwise FWE corrected). Shown is parametric modulation of BOLD signal with temporal (top) or spectral (bottom) filter cutoffs. (**C**) Multivariate analysis: accuracy (minus chance) maps for 10-category classifications of sentences and melodies ($p < .05$, cluster corrected). t-values are plotted on FsAverage surface (Freesurfer https://surfer.nmr.mgh.harvard.edu/). Regions of interest are extracted from the atlas by Glasser *et al*. (*19*): A1, primary auditory cortex; A4, auditory 4 complex (TE.3); RI, retroinsular cortex; PBelt, parabelt complex; LBelt, lateral belt complex; MBelt, medial belt complex. (**D**) Decoding accuracy in significant clusters of (C) presented as a function of domain and regions (chance level at 10%). Solid/dashed lines: participants showing the expected/reversed effect (sentences: $n = 14/1$; melodies: $n = 13/2$). (**E**) Prevalence analysis for 10-category classifications of sentences and melodies computed in an anatomically defined mask covering the entire temporal lobe (see the materials and methods). Highlighted areas are those where the majority null hypothesis (prevalence $\gamma < 50\%$ of the population) can be rejected at a level of $\alpha = 0.05$ ($p < 0.05$, corrected). (**F**) Left panel: Multivariate analysis for 10-category classification of melodies versus sentences for accuracy values analyzed in terms of lateralization index [$(R - L)/(R + L)$; $p < 0.05$, cluster corrected]. Right panel: Accuracy lateralization values in the significant cluster. Negative values indicate left-lateralized accuracy, whereas positive values indicate right-lateralized accuracy. Solid/dashed lines: participants showing the expected/reversed effect ($n = 13/2$). Bar plots show mean accuracy. White circles indicate individual data. Asterisks indicate significant effects; ns, nonsignificant effects.

superior temporal sulcus dorsal anterior regions, see table S1; $p < 0.05$, cluster corrected; Fig. 4D, right, and fig. S9). Post hoc tests (one-sample t tests) revealed that for sentences, NMI was left lateralized for spectral degradations ($t_{(14)} = -2.32$, $p = 0.03$), but the lateralization vanished for temporal degradations ($t_{(14)} = 0.44$, $p = 0.66$). By contrast, for melodies, NMI was right lateralized for temporal degradations ($t_{(14)} = 3.46$, $p = 0.004$) and the lateralization vanished for spectral degradations ($t_{(14)} = -0.24$, $p = 0.80$).

Years of debate have centered on the theoretically important question of the representation of speech and music in the brain (*2*, *6*, *21*). Here, we take advantage of the STM framework to establish a rigorous demonstration that: (i) perception of speech content is most affected by degradation of information in the temporal dimension, whereas perception of melodic content is most affected by degradation in the spectral dimension (Fig. 2, B and C); (ii) neural decoding of speech and melodic contents primarily depends on neural activity patterns in the left and right AC regions, respectively (Fig. 3, C to F, and fig. S4); (iii) in turn, this neural specialization for each stimulus domain is dependent on the specific sensitivity to STM rates of each auditory region (Fig. 4C and fig S7); and (iv) the perceptual effect of temporal or spectral degradation on speech or melodic content is mirrored specifically within each hemispheric auditory region (as revealed by mutual information), thereby demonstrating the brain–behavior relationship necessary to conclude that STM features are processed differentially for each stimulus domain within each hemisphere (Fig. 4D and figs. S8 and S9).

These results extend seminal studies on the robustness of speech comprehension to spectral degradation (*17*, *22*) and are also consistent with observations that the temporal modulation rate of speech samples from many languages is substantially higher than that of music samples across genres (*23*). It remains to be seen whether such a result also applies to other languages, such as tone languages, for which spectral information is arguably more important, and to musical pieces with complex rhythmic and harmonic variations or belonging to musical systems different from the Western tonal melodies used here.

The idea that auditory cognition depends on processing of spectrotemporal energy patterns

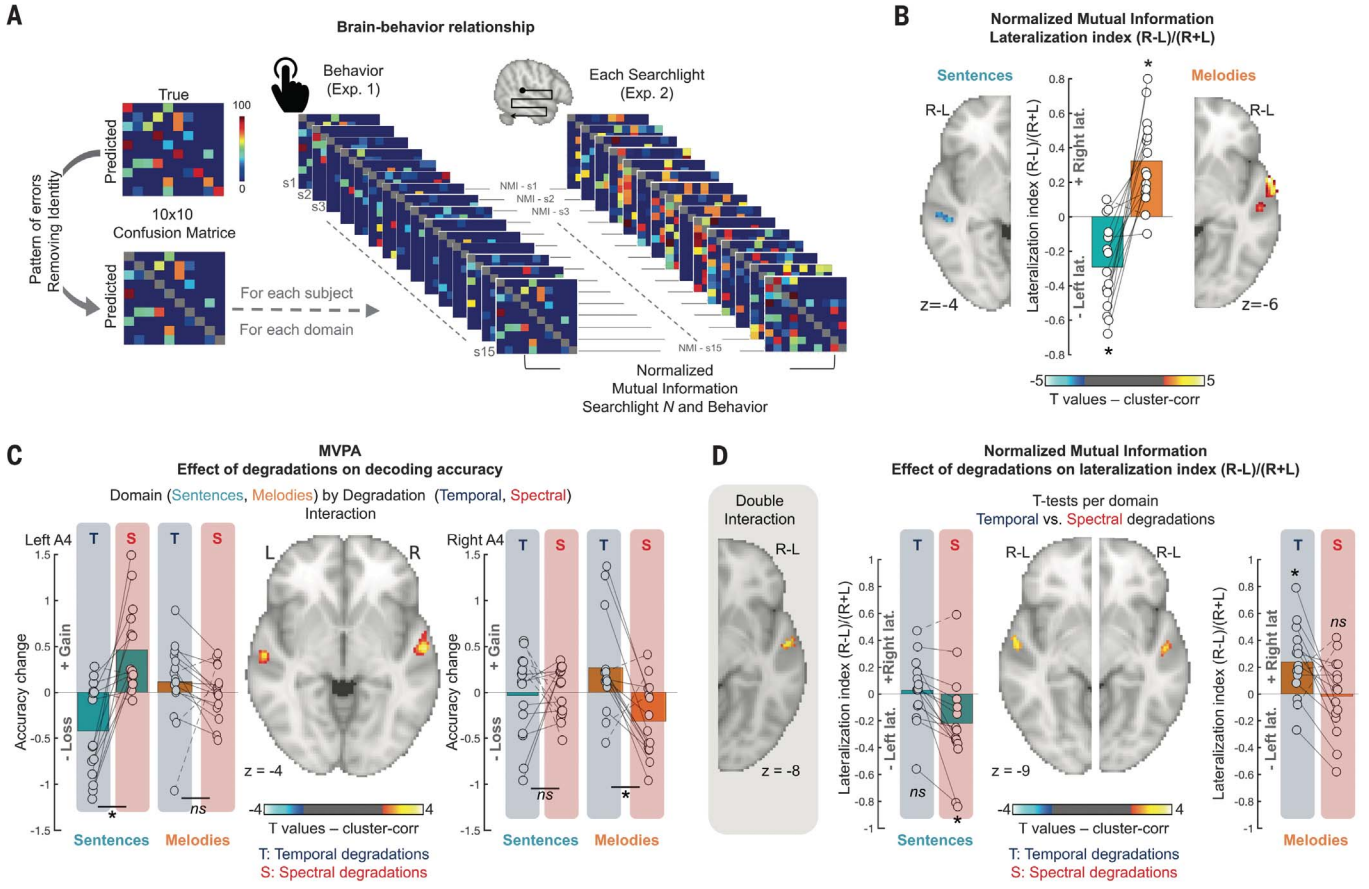

Fig. 4. Effect of degradations on decoding accuracy and NMI. (**A**) NMI computed between confusion matrices extracted from behavioral and fMRI data at the whole-brain level (searchlight procedure) for each participant and for each domain (sentences or melodies). (**B**) NMI results presented in terms of lateralization index [(R − L)/(R + L); $p < 0.05$, cluster corrected]. Light blue clusters indicate a left-lateralized NMI for sentences; red and yellow clusters indicate a right-lateralized NMI for melodies. White circles indicate individual data. Solid/dashed lines: participants showing the expected/reversed effect ($n = 14/1$). (**C**) Accuracy change for 10-category classification of sentences (aqua) and melodies (orange) presented as a function of degradation type (blue-shaded bars: temporal degradation; red-shaded bars: spectral degradation) revealing a domain-by-degradation interaction ($p < 0.05$, cluster corrected). Same conventions as in (B) (solid/dashed lines: $n = 13/2$ for sentences, $n = 12/3$ for melodies). (**D**) Effect of acoustic degradations on NMI lateralization index [(R − L)/(R + L); $p < 0.05$, cluster corrected]. Left panel: Domain-by-degradation interaction. Right panel: Effect of degradation (temporal versus spectral) performed per domain (sentences and melodies). Bar plots indicate the mean lateralization index per domain and degradation type (temporal and spectral). Same conventions as in (B) and (C) (solid/dashed lines: $n = 14/1$ for sentences, $n = 12/3$ for melodies). Asterisks indicate significant lateralization index; ns, nonsignificant lateralization index.

and that these features often trade off against one another is supported by human psychophysics (*17*, *18*), recordings from cat inferior colliculus (*13*), and human neuroimaging (*6*, *7*, *15*–*17*). During passive listening of short, isolated stimuli lacking semantic content, preferences for high spectral versus temporal modulation are distributed in an anterior–posterior dimension of the AC, with relatively weaker hemispheric differences (*6*, *7*, *15*, *16*). Our results suggest that this purely acoustic lateralization may be enhanced during the iterative analysis of temporally structured natural stimuli (*24*) in the most anterior and inferior auditory (A4) patches, which are known to analyze complex acoustic features and their relationships, or sound categories, thus fitting well with their encoding of relevant speech or musical features (*6*, *25*, *26*). We hypothesize that hemispheric lateralization of STM cues scales with the strength of the dynamical interactions between acoustic and higher-level (motor, syntactic, working memory, etc.) processes, which are typically maximized with complex, cognitively engaging stimuli that require decoding of feature relationships to extract meaning (speech or melodic content), as used here.

More generally, studies across numerous species have indicated a match between ethologically relevant stimulus features and the spectrotemporal response functions of their auditory nervous systems, suggesting efficient adaptation to the statistical properties of relevant sounds, especially communicative ones (*27*). This is consistent with the theory of efficient neural coding (*28*). Our study shows that in addition to speech, this theory can be applied to melodic information, a form-bearing dimension of music. Humans have developed two means of auditory communication: speech and music. Our study suggests that these two domains exploit opposite extremes of the spectrotemporal continuum, with a complementary specialization of two parallel neural systems, one in each hemisphere, that maximizes the efficiency of encoding of their respective acoustical features.

REFERENCES AND NOTES

1. D. Poeppel, *Speech Commun.* **41**, 245–255 (2003).
2. R. J. Zatorre, P. Belin, V. B. Penhune, *Trends Cogn. Sci.* **6**, 37–46 (2002).
3. C. McGettigan, S. K. Scott, *Trends Cogn. Sci.* **16**, 269–276 (2012).
4. I. Peretz, M. Coltheart, *Nat. Neurosci.* **6**, 688–691 (2003).
5. A. D. Friederici, *Philos. Trans. R. Soc. London B Biol. Sci.* **375**, 20180391 (2020).
6. S. Norman-Haignere, N. G. Kanwisher, J. H. McDermott, *Neuron* **88**, 1281–1296 (2015).
7. R. J. Zatorre, P. Belin, *Cereb. Cortex* **11**, 946–953 (2001).
8. J. Obleser, F. Eisner, S. A. Kotz, *J. Neurosci.* **28**, 8116–8123 (2008).
9. M. Schönwiesner, R. Rübsamen, D. Y. von Cramon, *Eur. J. Neurosci.* **22**, 1521–1528 (2005).
10. A. Boemio, S. Fromm, A. Braun, D. Poeppel, *Nat. Neurosci.* **8**, 389–395 (2005).
11. B. Morillon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18688–18693 (2010).
12. T. Chi, P. Ru, S. A. Shamma, *J. Acoust. Soc. Am.* **118**, 887–906 (2005).
13. F. A. Rodríguez, H. L. Read, M. A. Escabí, *J. Neurophysiol.* **103**, 887–903 (2010).
14. J. Fritz, S. Shamma, M. Elhilali, D. Klein, *Nat. Neurosci.* **6**, 1216–1223 (2003).
15. M. Schönwiesner, R. J. Zatorre, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14611–14616 (2009).
16. R. Santoro *et al.*, *PLOS Comput. Biol.* **10**, e1003412 (2014).
17. A. Flinker, W. K. Doyle, A. D. Mehta, O. Devinsky, D. Poeppel, *Nat. Hum. Behav.* **3**, 393–405 (2019).
18. T. M. Elliott, F. E. Theunissen, *PLOS Comput. Biol.* **5**, e1000302 (2009).
19. M. F. Glasser *et al.*, *Nature* **536**, 171–178 (2016).
20. R. A. Ince *et al.*, *Hum. Brain Mapp.* **38**, 1541–1573 (2017).
21. D. Schön *et al.*, *Neuroimage* **51**, 450–461 (2010).
22. R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, M. Ekelid, *Science* **270**, 303–304 (1995).
23. N. Ding *et al.*, *Neurosci. Biobehav. Rev.* **81**, (pt. B), 181–187 (2017).
24. A. M. Leaver, J. P. Rauschecker, *J. Neurosci.* **30**, 7604–7612 (2010).
25. T. Overath, J. H. McDermott, J. M. Zarate, D. Poeppel, *Nat. Neurosci.* **18**, 903–911 (2015).
26. M. Chevillet, M. Riesenhuber, J. P. Rauschecker, *J. Neurosci.* **31**, 9345–9352 (2011).
27. L. H. Arnal, A. Flinker, A. Kleinschmidt, A. L. Giraud, D. Poeppel, *Curr. Biol.* **25**, 2051–2056 (2015).
28. J. Gervain, M. N. Geffen, *Trends Neurosci.* **42**, 56–65 (2019).
29. P. Albouy, L. Benjamin, B. Morillon, R. J. Zatorre, Data for: Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody, Open Science Framework (2020); https://doi.org/10.17605/OSF.IO/9UB78.

ACKNOWLEDGMENTS

We thank S. Norman-Haignere, A.-L. Giraud, and E. Coffey for comments on a previous version of the manuscript; C. Soden for creating the melodies; A.-K. Barbeau for singing the stimuli; and M. Generale and M. de Francisco for expertise with recording. **Funding:** This work was supported by a foundation grant from the Canadian Institute for Health Research to R.J.Z. P.A. is funded by a Banting Fellowship. R.J.Z. is a senior fellow of the Canadian Institute for Advanced Research. B.M.'s research is supported by grants ANR-16-CONV-0002 (ILCB) and ANR-11-LABX-0036 (BLRI) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). **Author contributions:** Conceptualization: B.M., P.A., R.J.Z.; Methodology: P.A., L.B., B.M., R.J.Z.; Analysis: P.A., L.B.; Investigation: L.B., P.A.; Resources: R.J.Z.; Writing original draft: P.A., B.M., R.J.Z.; Writing – review & editing: P.A., L.B., B.M., R.J.Z.; Visualization: P.A.; Supervision: B.M., R.J.Z. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Sound files can be found at www.zlab.mcgill.ca/downloads/albouy_20190815/. A demo of the behavioral task can be found at: https://www.zlab.mcgill.ca/spectro_temporal_modulations/. Data and code used to generate the findings of this study are accessible online (*29*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/367/6481/1043/suppl/DC1
Materials and Methods
Figs. S1 to S9
Table S1
Supplementary Results
References (*30*–*32*)

View/request a protocol for this paper from *Bio-protocol*.

2 September 2019; accepted 2 January 2020
10.1126/science.aaz3468

Science

Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody

Philippe Albouy, Lucas Benjamin, Benjamin Morillon and Robert J. Zatorre

Science **367** (6481), 1043-1047.
DOI: 10.1126/science.aaz3468

Speech versus music in the brain

To what extent does the perception of speech and music depend on different mechanisms in the human brain? What is the anatomical basis underlying this specialization? Albouy *et al.* created a corpus of a cappella songs that contain both speech (semantic) and music (melodic) information and degraded each stimulus selectively in either the temporal or spectral domain. Degradation of temporal information impaired speech recognition but not melody recognition, whereas degradation of spectral information impaired melody recognition but not speech recognition. Brain scanning revealed a right-left asymmetry for speech and music. Classification of speech content occurred exclusively in the left auditory cortex, whereas classification of melodic content occurred only in the right auditory cortex.

Science, this issue p. 1043

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.