

of anti-PirB.1. As we have reported previously (8), NgR^{-/-} CGN neurite outgrowth is inhibited by AP-Nogo66 or myelin to the same extent as that in WT neurons (50% and 49%) (Fig. 3). Anti-PirB.1 antibody treatment of NgR^{+/-} neurons partially reversed inhibition by either AP-Nogo66 or myelin, as discussed above with WT neurons. Anti-PirB.1 treatment of NgR^{-/-} neurons resulted in a similar partial disinhibition on AP-Nogo66 but did not provide any further rescue, suggesting that NgR is not involved in AP-Nogo66 inhibition. In contrast, anti-PirB.1 treatment of NgR^{+/-} neurons restored neurite outgrowth on myelin to nearly control levels. Thus, it appears that PirB, but not NgR, is required for substrate inhibition by AP-Nogo66 in CGNs, but only accounts for it partly. In contrast, PirB and NgR both contribute to the substrate inhibition imparted by myelin.

Because NgR is required for growth-cone collapse in response to various myelin inhibitors (6, 7), it is possible that PirB is also involved in this more acute response. For this experiment, we used sensory neurons from the DRG of 3-week-old mice, confirmed to express PirB (fig. S3). Similar to what has been shown by others (6, 7), we found that growth cones in this culture system have a high baseline level of collapse (~30%), which is further increased (to ~75%) by incubation with AP-Nogo66 or myelin (Fig. 4). As reported previously (6, 7), this collapse was largely abolished in NgR^{-/-} neurons. In addition, blocking PirB function with anti-PirB.1 was also sufficient to reverse growth-cone collapse by these inhibitors.

Together, these results support a previously unknown role for PirB as a necessary receptor for neurite inhibition by myelin extracts and, more specifically, by the myelin-associated inhibitors Nogo66, MAG, and OMgp. PirB appears to be a more substantial mediator of substrate inhibition than NgR, as removal of PirB function alone (either genetically or with antibodies) partially disinhibits growth on both myelin extracts and inhibitors, whereas genetic removal of NgR alone does not disinhibit on any of these substrates. However, NgR appears to play an adjunct role in mediating inhibition by myelin extracts (but not Nogo66), because genetic removal of NgR can augment the disinhibition caused by anti-PirB antibodies on myelin (but not on Nogo66). Other co-receptors or modulators may also contribute in parallel, such as p75, TROY, LINGO, and gangliosides (17, 18). Conversely, other PirB ligands, including major histocompatibility complex class I proteins, may contribute to the inhibitory action of myelin (13, 19, 20). Our finding of collaboration between PirB and NgR may help to explain the surprising lack of enhanced CST regeneration after dorsal spinal cord hemisection in NgR knockout mice (6, 8), despite the reported regeneration or sprouting seen in rodents infused with the NgR ectodomain (21). Thus, it might be necessary to remove both PirB and NgR to achieve extensive regeneration in vivo. In addition, because on Nogo66 substrate the genetic removal of NgR does not further augment the partial dis-

inhibitory effect of PirB removal, it is likely that there are additional binding receptors for Nogo66.

Although PirB appears to be a more important receptor for substrate inhibition than NgR, inactivation of either PirB or NgR alone is sufficient to block the acute growth-cone collapse caused by the addition of myelin inhibitors. This observation suggests that collapse is a more demanding process, requiring both PirB and NgR activities, acting either in parallel or together. In this context, it is of interest that PirB and NgR receptors have recently been shown to play similar roles in limiting plasticity of synaptic connections in the visual cortex. In mice lacking either receptor, eye closure during a critical developmental period results in excessive strengthening of connections via the open eye (13, 22). The mechanisms responsible for the effect of both receptors in mediating growth-cone collapse could also underlie the commonality of their role in ocular dominance plasticity.

The mechanism by which PirB signals to inhibit axon growth in response to myelin inhibitors is not clear. However, PirB has been shown to antagonize the function of integrin receptors (23) and to recruit both src homology 2-containing protein tyrosine phosphatase (SHP)-1 and SHP-2 phosphatases (13, 24); either or both of these events could attenuate normal neurite outgrowth. In humans, one or more members of the LILRB gene family might also play a role in regeneration. The blockade of PirB/LILRB activity, either with antibodies or by other means, provides an important target for therapeutic interventions to stimulate axonal regeneration.

References and Notes

1. A. E. Fournier, T. GrandPre, S. M. Strittmatter, *Nature* **409**, 341 (2001).
2. M. Domeniconi *et al.*, *Neuron* **35**, 283 (2002).
3. B. P. Liu, A. Fournier, T. GrandPre, S. M. Strittmatter, *Science* **297**, 1190 (2002); published online 27 June 2002 (10.1126/science.1073031).

4. K. C. Wang *et al.*, *Nature* **417**, 941 (2002).
5. J. K. Huang *et al.*, *Science* **310**, 1813 (2005); published online 16 November 2005 (10.1126/science.1118313).
6. J. E. Kim, B. P. Liu, J. H. Park, S. M. Strittmatter, *Neuron* **44**, 439 (2004).
7. O. Chivatakarn, S. Kaneko, Z. He, M. Tessier-Lavigne, R. J. Giger, *J. Neurosci.* **27**, 7117 (2007).
8. B. Zheng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1205 (2005).
9. Materials and methods are available as supporting material on Science Online.
10. A. M. Martin, J. K. Kulski, C. Witt, P. Pontarotti, F. T. Christiansen, *Trends Immunol.* **23**, 81 (2002).
11. H. Kubagawa, P. D. Burrows, M. D. Cooper, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5261 (1997).
12. K. C. Wang, J. A. Kim, R. Sivasankaran, R. Segal, Z. He, *Nature* **420**, 74 (2002).
13. J. Syken, T. GrandPre, P. O. Kanold, C. J. Shatz, *Science* **313**, 1795 (2006); published online 16 August 2006 (10.1126/science.1128232).
14. A. E. Fournier, G. C. Gould, B. P. Liu, S. M. Strittmatter, *J. Neurosci.* **22**, 8876 (2002).
15. X. L. He *et al.*, *Neuron* **38**, 177 (2003).
16. W. C. Liang *et al.*, *J. Mol. Biol.* **366**, 815 (2007).
17. A. R. Walmsley, A. K. Mir, *Curr. Pharm. Des.* **13**, 2470 (2007).
18. N. R. Mehta, P. H. Lopez, A. A. Vyas, R. L. Schnaar, *J. Biol. Chem.* **282**, 27875 (2007).
19. M. Colonna *et al.*, *J. Exp. Med.* **186**, 1809 (1997).
20. A. Nakamura, E. Kobayashi, T. Takai, *Nat. Immunol.* **5**, 623 (2004).
21. S. Li *et al.*, *J. Neurosci.* **24**, 10511 (2004).
22. A. W. McGee, Y. Yang, Q. S. Fischer, N. W. Daw, S. M. Strittmatter, *Science* **309**, 2222 (2005).
23. S. Pereira, H. Zhang, T. Takai, C. A. Lowell, *J. Immunol.* **173**, 5757 (2004).
24. B. Chen, L. Hammonds-Odie, J. Perron, B. A. Masters, J. L. Bixby, *Dev. Biol.* **252**, 170 (2002).
25. We thank J. Brady and M. Yan for assistance with expression cloning and members of the Tessier-Lavigne and Shatz laboratories for helpful suggestions and discussions. This work was supported by Genentech.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/967/DC1
Materials and Methods
Figs. S1 to S5
References

30 May 2008; accepted 11 September 2008
10.1126/science.1161151

“Who” Is Saying “What”? Brain-Based Decoding of Human Voice and Speech

Elia Formisano,* Federico De Martino, Milene Bonte, Rainer Goebel

Can we decipher speech content (“what” is being said) and speaker identity (“who” is saying it) from observations of brain activity of a listener? Here, we combine functional magnetic resonance imaging with a data-mining algorithm and retrieve what and whom a person is listening to from the neural fingerprints that speech and voice signals elicit in the listener’s auditory cortex. These cortical fingerprints are spatially distributed and insensitive to acoustic variations of the input so as to permit the brain-based recognition of learned speech from unknown speakers and of learned voices from previously unheard utterances. Our findings unravel the detailed cortical layout and computational properties of the neural populations at the basis of human speech recognition and speaker identification.

In everyday life, we automatically and effortlessly decode speech into language independently of who speaks. Similarly, we recognize a speaker’s voice independently of what she or

he says. Cognitive and connectionist models postulate that this efficiency depends on the ability of our speech perception and speaker identification systems to extract relevant features from the sen-

sory input and to form efficient abstract representations (1–3). These representations are invariant to changes of the acoustic input, which ensures efficient processing and confers a high robustness to noise or to signal distortion. Relevant psycholinguist models consider abstract entities such as phonemes as the building blocks of the computational chain that transforms an acoustic waveform into a meaningful concept (2, 3). There is also psychoacoustic evidence that the identification of a speaker relies on the extraction of invariant paralinguistic features of his/her voice, such as fundamental frequency (1).

Numerous functional neuroimaging studies have provided important insights on the cortical organization of speech (4–11) and voice (12, 13) processing. However, the subtraction-based experimental logic and the limited neuroanatomical detail only allow for partial and indirect inferences on what distinguishes the auditory cortical representations of two natural speech or vocal sounds. Furthermore, it remains unclear how a speech sound is transformed into the more abstract entity of “phoneme” or “speaker” identity. Beyond subtraction, results from functional magnetic resonance adaptation suggest the involvement in voice identification of a specialized region in the right anterior superior temporal sulcus (STS) (14). For speech processing, a hierarchical fractioning of cortical regions for sound-based and a more abstract higher-level processing has been suggested [(15) and supporting online text].

In the present study, we investigate speech and voice recognition and abstraction at the level of representation and processing of individual sounds. By combining multivariate statistical pattern recognition with single-trial functional magnetic resonance imaging (fMRI) (16–20), we estimate and decode the distinct activation patterns elicited by different speech sounds and directly assess the invariance of the estimated neural representations to acoustic variations of the sensory input.

High spatial resolution (1.5 mm × 1.5 mm × 2 mm) functional images of the auditory cortex were collected while participants ($n = 7$) listened to speech sounds consisting of three Dutch vowels (/a/, /i/, /u/) recorded from three native Dutch speakers (Fig. 1) (21). Consistent with previous studies (4–6, 8, 10, 12, 14, 15, 22), all sounds evoked significant fMRI responses in a wide expanse of the superior temporal cortex, including early auditory areas (Heschl’s gyrus) and multiple regions in the planum temporale (PT), along the superior temporal gyrus, the STS, and the middle temporal gyrus. Univariate statistical contrasts, however, yielded only weak response differences (below significance) or no differences between conditions (fig. S2)

After this initial analysis, we asked whether the estimation of a multivoxel activation fingerprint of a sound would allow deciphering its content and the identity of the speaker. With a method based on a machine learning classification algorithm (support vector machine) and recursive feature elimination (23, 24), we performed two complementary analyses. We labeled the stimuli and corresponding response patterns either according to the vowel dimension irrespective of the speaker dimension (“vowel learning”) or according to the speaker dimension irrespective of the vowel dimension (“speaker learning”). This led to the grouping of stimuli and responses in the three conditions: /a/, /i/, and /u/ and sp1, sp2, and sp3, respectively. We then examined whether our algorithm, after being trained with a subset of labeled responses (50 trials), would correctly classify the remaining unlabeled responses (10 trials). In all subjects and in all possible pairwise comparisons, the al-

gorithm successfully learned the functional relation between sounds and corresponding spatial patterns and classified correctly the unlabeled sound-evoked patterns, both in the case of vowel learning [a/ versus i/ = 0.65 (mean correctness), $P = 6 \times 10^{-5}$; a/ versus u/ = 0.69, $P = 2 \times 10^{-5}$; i/ versus u/ = 0.63, $P = 4 \times 10^{-4}$ (Fig. 2A)]; and speaker learning, [sp1 versus sp2 = 0.70, $P = 3 \times 10^{-5}$; sp1 versus sp3 = 0.67, $P = 8 \times 10^{-5}$; sp2 versus sp3 = 0.62, $P = 2 \times 10^{-5}$ (Fig. 2B)].

To investigate layout and consistency across subjects of the spatial patterns that make this decoding possible, we generated group discriminative maps (Fig. 3 and fig. S3), i.e., maps of the cortical locations that contribute most to the discrimination of conditions. Single-subject reliability maps are reported in fig. S4. Discriminative patterns for vowels [red color (Fig. 3B and fig. S3)] were widely distributed bilaterally in the superior temporal cortex and included regions in

Fig. 1. Experimental design and stimuli.

(A) Example of spectrograms of the stimuli from the nine conditions (three vowels × three speakers). Stimuli were presented during the silent intervals of fMRI measurements and were three natural Dutch vowels (/a/, /i/, and /u/) spoken by three native Dutch speakers (sp1: female, sp2: male, and sp3: male). (B) Representation of the vowels based on the first two formants (F1, F2). Each of the conditions was formed by grouping three different utterances from the same speaker. The insert indicates mean value and standard deviation of the fundamental frequency (F0) for each of the speakers.

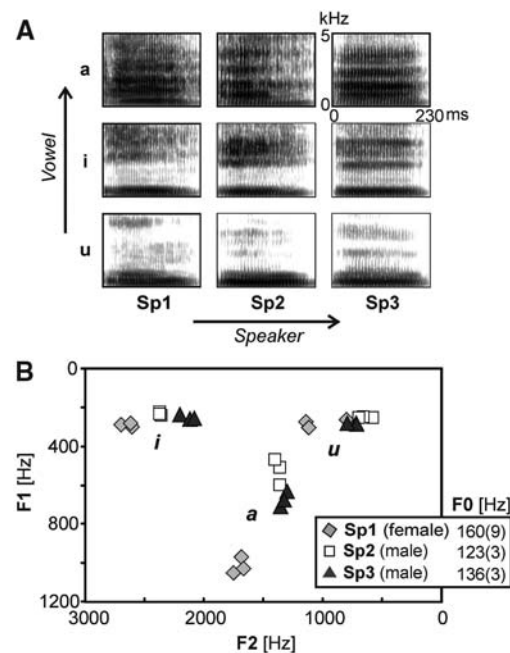
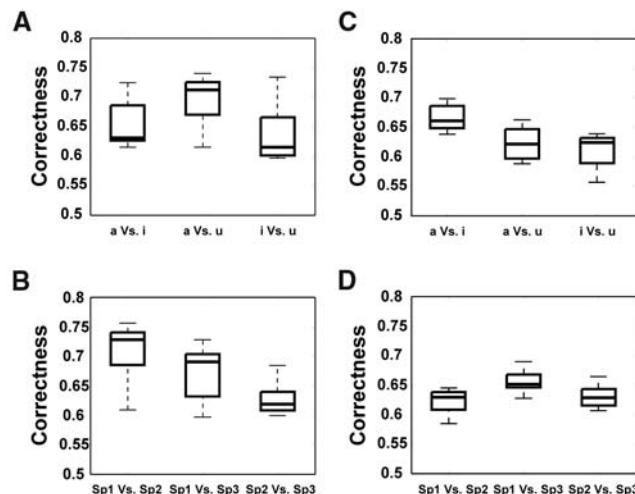


Fig. 2. Performance of the brain-based decoding of vowels and speakers. Correctness (median value and distribution) of all pairwise classifications when training and testing of the algorithm were based on different subsets of responses to the same stimuli [(A) vowels; (B) speakers] or when training and testing were based on responses to different speakers (C) and vowels (D) (chance level is 0.5).



Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, University of Maastricht, 6200 MD Maastricht, Netherlands.

*To whom correspondence should be addressed. E-mail: e.formisano@psychology.unimaas.nl

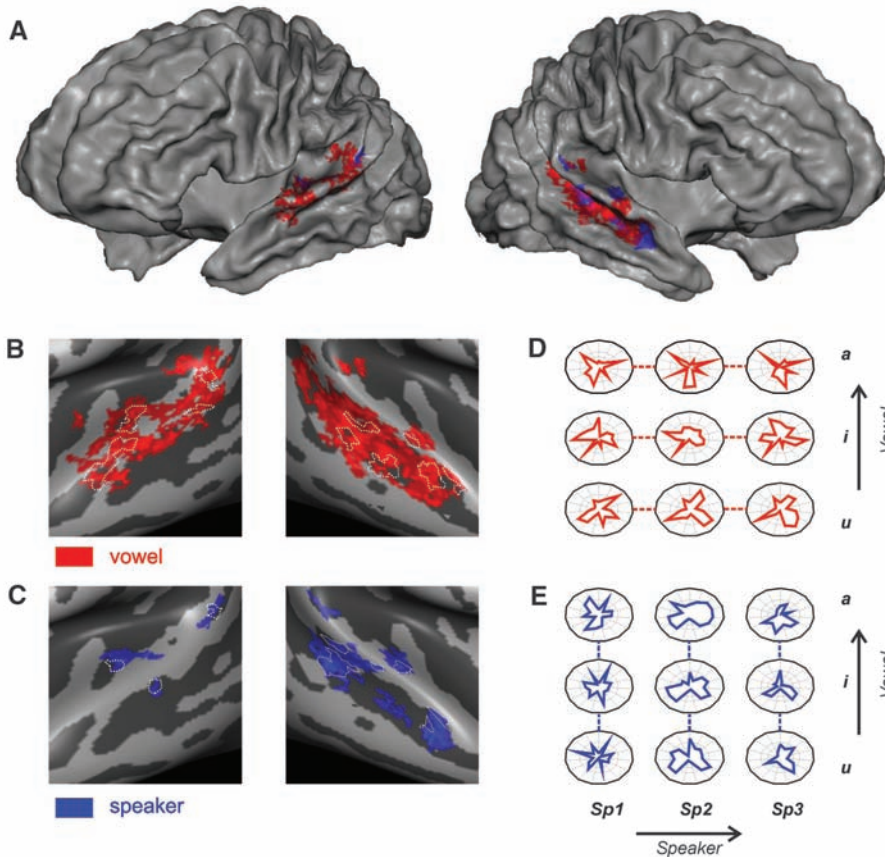


Fig. 3. Cortical discriminative maps and activation fingerprints for decoding of vowels and speakers. (A to C) Group discriminative maps obtained from cortex-based realignment of individual maps. Maps are visualized on the folded (A) or inflated representation of the cortex [auditory cortex detail in (B and C); light gray, gyri; dark gray, sulci] resulting from the realignment of the cortices of the seven participants. A location was color-coded (vowels, red; speakers, blue) if it was present on the individual maps of at least four of the seven subjects. This corresponds to a false discovery rate–corrected threshold of $q = 6 \times 10^{-4}$ for vowels and $q = 9 \times 10^{-4}$ for speakers (21). Outlined regions in (B) and (C) indicate cortical regions that were also included in the group maps obtained with the generalization analysis. (D and E) Activation fingerprints of the sounds created from the 15 most discriminative voxels for decoding of vowels (D) and speakers (E) (single-subject data, subject 1). Each axis of the polar plot forming a fingerprint displays the normalized activation level in a voxel. Note the similarity among the fingerprints of the same vowel [horizontal direction in (D)] or speaker [vertical direction in (E)].

the anterior-lateral portion of Heschl’s gyrus–Heschl’s sulcus, in the PT (mainly in the left hemisphere), and extended portions of the STS/STG (both hemispheres). Discriminative patterns for speakers [blue color (Fig. 3C and fig. S3)] were more confined and right-lateralized than those obtained for vowel discrimination. These patterns included the lateral portion of Heschl’s gyrus–Heschl’s sulcus, located in the posterior adjacency to a similar region described for vowel discrimination and three clustered regions along the anterior-posterior axis of the right STS, also interspersed with vowels regions (fig. S3). These findings indicate a spatially distributed model for both the representation of vowel and speaker identity (see supporting online text).

Encouraged by these results, we tested the capability of our algorithm to decipher the brain activity into speech content and speaker identity also in the case of completely novel stimuli (i.e.,

stimuli not used during the training). We trained the algorithm in discriminating vowels with samples from one speaker (e.g., /a/ versus /i/ for sp1) or one vowel (e.g., sp1 versus sp2 for /a/) and tested the correctness of this discrimination in the other speakers (e.g., sp2 and sp3) or in the other vowels (e.g., /i/ and /u/). With this strategy, stimuli used for training and for testing differ in many acoustical dimensions. An accurate decoding of activation patterns associated with the test stimuli would thus indicate that the learned functional relation between a cortical activation pattern and a vowel (or a speaker) entails information on that vowel (or speaker) beyond the contingent mapping of its acoustic properties. Despite the small number of training samples (20 trials), classification of novel stimuli was accurate in all subjects and in all possible pairwise comparisons, both in the case of vowels [a/ versus /i/ = 0.66 (mean accuracy), $P = 1 \times 10^{-6}$; a/ versus /u/ =

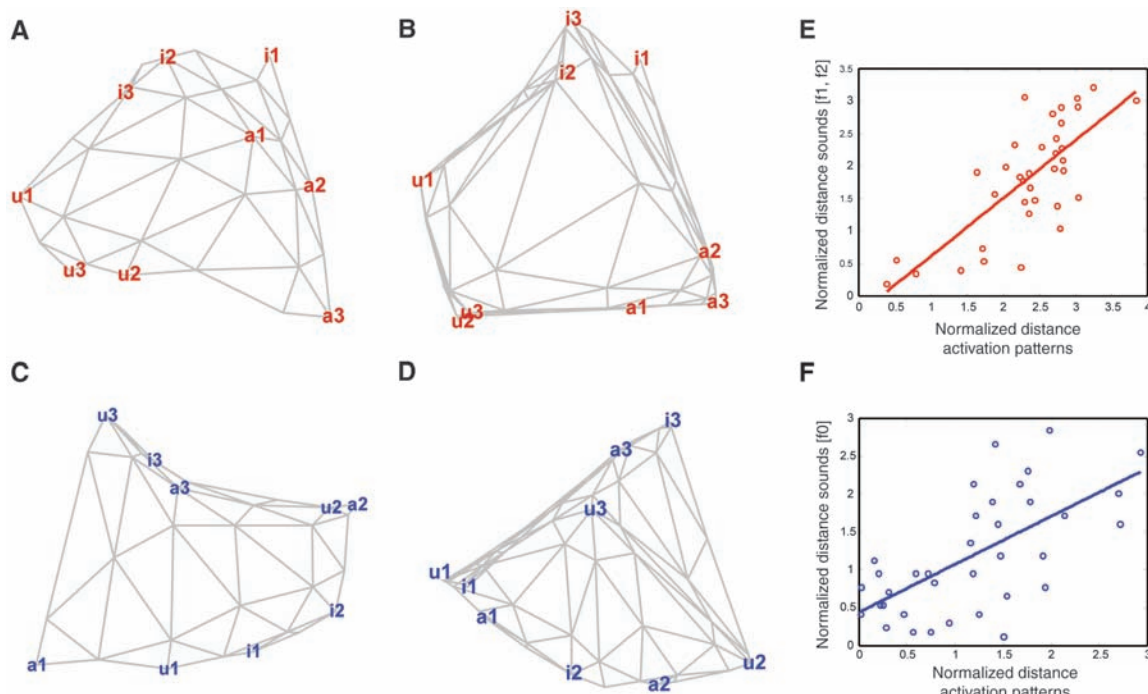
0.62, $P = 3 \times 10^{-5}$; /i/ versus /u/ = 0.60, $P = 7 \times 10^{-5}$ (Fig. 2C)] and in the case of speakers [sp1 versus sp2 = 0.62 (mean accuracy), $P = 6 \times 10^{-6}$; sp1 versus sp3 = 0.65, $P = 8 \times 10^{-7}$; sp2 versus sp3 = 0.63, $P = 2 \times 10^{-6}$ (Fig. 2D)]. Although sparser, the corresponding discriminative maps included a subset of the locations highlighted by the previous analyses (outlined regions in Fig. 3, B and C).

The abstract nature of the estimated cortical representations is illustrated in Figs. 3, D and E, and 4. First, we visualized the speaker-invariant cortical representation of vowels and the vowel-invariant representation of speakers for which we used multidimensional displays [fingerprints (Fig. 3, D and E)]. Second, we visualized the relation between the discriminative patterns of activations for the nine conditions using self-organizing maps (SOMs) (21), which convert complex relations between high-dimensional items into simple geometric relations. The spatial proximity and grouping of the conditions in the SOM-based two-dimensional display thus reflects the level of abstraction and categorical information entailed in the discriminative patterns of vowels (Fig. 4, A and B) and speakers (Fig. 4, C and D). To investigate which acoustic features in the original sounds drive this neural abstraction, we examined the relative distance between the brain-based representations of the stimuli and their description in terms of typical acoustic features (formants, Fig. 1B). We found that the distances between the cortical representations of the sounds correlated best with a description of the stimulus based on the first two formants (F1, F2) in the case of vowels [$r = 0.75$, $P = 2 \times 10^{-7}$ (Fig. 4E and fig. S5)] and on the fundamental frequency (F0) in the case of speakers [$r = 0.64$, $p = 2 \times 10^{-5}$ (Fig. 4F and fig. S5)]. These results provide empirical support for cognitive models of speech and voice processing postulating the existence of intermediate computational entities resulting from the transformation of relevant acoustic features [the (F1, F2) pair for vowels and (F0) for speakers] and the suppression of the irrelevant ones.

Our findings demonstrate that an abstract representation of a vowel or speaker emerges from the joint encoding of information occurring not only in specialized higher-level regions but also in auditory regions, which—because of their anatomical connectivity and response properties—have been associated with early stages of sound processing. This is in agreement with recent neurophysiological findings indicating that neurons in early auditory regions may exhibit complex spectrotemporal receptive fields and may participate in high-level encoding of auditory objects (25–29), e.g., via local feedback loops and reentrant processing. Taken together, these results prompt a revision of models of phoneme and voice abstraction, which assumes that a hierarchy of processing steps is “mapped” into a functional hierarchy of specialized neural modules.

In conclusion, we demonstrated the feasibility of decoding speech content and speaker

Fig. 4. Visualization of the brain-based representation of the sounds and relation with acoustical features. (A to D) SOM-based display of the discriminative patterns in a single-subject (subject 1, A and C) and in the group of seven subjects (B and D) for vowel (A and B) and speaker learning (C and D). (E and F) Relation between normalized distances of the multi-dimensional auditory cortical activation patterns and normalized distances of the vowels in the (F1, F2) space of formants (E) and of the speakers in the space of fundamental frequency (FO) (F).



identity from observation of auditory cortical activation patterns of the listeners. Our analyses provided a detailed empirical demonstration of how the human brain forms computationally efficient representations required for speech comprehension and speaker identification. Our experimental settings, however, were restricted to three vowels and three speakers; furthermore, all sounds were presented in isolation to obtain distinct fMRI activation patterns. Extension of these results to identify a word or concatenation of words in streams of longer speech segments, provides a compelling challenge and will contribute to create a general brain-based decoder of sounds in the context of real-life situations.

References and Notes

- P. Belin, S. Fecteau, C. Bedard, *Trends Cogn. Sci.* **8**, 129 (2004).
- J. L. McClelland, J. L. Elman, *Cognit. Psychol.* **18**, 1 (1986).
- D. Norris, J. M. McQueen, *Psychol. Rev.* **115**, 357 (2008).
- J. R. Binder et al., *Cereb. Cortex* **10**, 512 (2000).
- G. Hickok, D. Poeppel, *Nat. Rev. Neurosci.* **8**, 393 (2007).
- E. Liebenthal, J. R. Binder, S. M. Spitzer, E. T. Possing, D. A. Medler, *Cereb. Cortex* **15**, 1621 (2005).
- R. Näätänen et al., *Nature* **385**, 432 (1997).
- J. Obleser et al., *Hum. Brain Mapp.* **27**, 562 (2006).
- J. Obleser, A. Lahiri, C. Eulitz, *J. Cogn. Neurosci.* **16**, 31 (2004).
- J. Obleser, J. Zimmermann, J. Van Meter, J. P. Rauschecker, *Cereb. Cortex* **17**, 2251 (2007).
- S. K. Scott, I. S. Johnsrude, *Trends Neurosci.* **26**, 100 (2003).
- P. Belin, R. J. Zatorre, P. Lafaille, P. Ahad, B. Pike, *Nature* **403**, 309 (2000).
- J. D. Warren, S. K. Scott, C. J. Price, T. D. Griffiths, *Neuroimage* **31**, 1389 (2006).
- P. Belin, R. J. Zatorre, *Neuroreport* **14**, 2105 (2003).
- M. H. Davis, I. S. Johnsrude, *J. Neurosci.* **23**, 3423 (2003).
- J. V. Haxby et al., *Science* **293**, 2425 (2001).
- J. D. Haynes, G. Rees, *Nat. Neurosci.* **8**, 686 (2005).
- Y. Kamitani, F. Tong, *Nat. Neurosci.* **8**, 679 (2005).
- K. N. Kay, T. Naselaris, R. J. Prenger, J. L. Gallant, *Nature* **452**, 352 (2008).
- N. Kriegeskorte, R. Goebel, P. Bandettini, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3863 (2006).
- Material and Methods and supporting data are available on Science Online
- R. Desai, E. Liebenthal, E. Waldron, J. R. Binder, *J. Cogn. Neurosci.* **20**, 1174 (2008).
- F. De Martino et al., *Neuroimage* **43**, 44 (2008).
- E. Formisano, F. De Martino, G. Valente, *Magn. Reson. Imaging* **26**, 921 (2008).
- C. T. Engineer et al., *Nat. Neurosci.* **11**, 603 (2008).
- N. Mesgarani, S. V. David, J. B. Fritz, S. A. Shamma, *J. Acoust. Soc. Am.* **123**, 899 (2008).
- I. Nelken, *Curr. Opin. Neurobiol.* **14**, 474 (2004).
- F. W. Ohl, H. Scheich, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9440 (1997).
- X. Wang, T. Lu, R. K. Snider, L. Liang, *Nature* **435**, 341 (2005).
- Work supported by Maastricht University and the Netherlands Organization for Scientific Research (NWO), Innovational Research Incentives Scheme Vidi grant 452-04-337 (E.F.) and Veni grant 451-07-002 (M.B.). We thank J. Obleser, U. Hasson, J.-D. Haynes, N. Kriegeskorte, J. Ashburner, and G. Valente for comments and discussions.

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5903/970/DC1
Materials and Methods
Supporting Text
Figs. S1 to S5
References

6 August 2008; accepted 9 October 2008
10.1126/science.1164318