Current Biology 19, 1028–1033, June 23, 2009 ©2009 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2009.04.054

Report

Decoding of Emotional Information in Voice-Sensitive Cortices

Thomas Ethofer,[1,2,5,*] Dimitri Van De Ville,[6] Klaus Scherer,[2,3] and Patrik Vuilleumier[1,2,4]
[1]Laboratory for Behavioral Neurology & Imaging of Cognition
Department of Neuroscience & Clinic of Neurology
Medical School
[2]Swiss Center of Affective Sciences
[3]Department of Psychology
[4]Neuroscience Center
University of Geneva
Geneva
Switzerland
[5]Clinic for Psychiatry and Psychotherapy
University of Tuebingen
Germany
[6]Biomedical Imaging Group
Ecole Polytechnique Fédérale de Lausanne
Switzerland

Summary

The ability to correctly interpret emotional signals from others is crucial for successful social interaction. Previous neuroimaging studies showed that voice-sensitive auditory areas [1–3] activate to a broad spectrum of vocally expressed emotions more than to neutral speech melody (prosody). However, this enhanced response occurs irrespective of the specific emotion category, making it impossible to distinguish different vocal emotions with conventional analyses [4–8]. Here, we presented pseudowords spoken in five prosodic categories (anger, sadness, neutral, relief, joy) during event-related functional magnetic resonance imaging (fMRI), then employed multivariate pattern analysis [9, 10] to discriminate between these categories on the basis of the spatial response pattern within the auditory cortex. Our results demonstrate successful decoding of vocal emotions from fMRI responses in bilateral voice-sensitive areas, which could not be obtained by using averaged response amplitudes only. Pairwise comparisons showed that each category could be classified against all other alternatives, indicating for each emotion a specific spatial signature that generalized across speakers. These results demonstrate for the first time that emotional information is represented by distinct spatial patterns that can be decoded from brain activity in modality-specific cortical areas.

Results and Discussion

We tested whether emotions expressed by speech melody (prosody) can be decoded from neural activity in voice-sensitive regions [1, 3] of the human auditory cortex (AC). To this aim, we used functional magnetic resonance imaging (fMRI) and multivariate pattern analysis (MVPA [9, 10]) based on a linear support vector machine (SVM) to determine whether the spatial response distribution in voice-sensitive regions encodes distinctive features of vocal emotions. This methodology exploits distributed information in activation patterns, as opposed to conventional approaches that are based on differences obtained at each voxel in isolation. MVPA has been successfully used to distinguish speech content and speaker identity [11]. However, it is unknown whether vocal emotions are likewise spatially encoded and whether it is possible to decrypt this code with MVPA.

Previous fMRI studies relying on standard data analyses have shown that the middle part of the superior temporal gyrus (STG) reacts more strongly to various vocal emotions [4–8] than to neutral prosody. However, because the increase is similar for all emotions, particular emotional categories could not be distinguished by conventional approaches. Similarly, electrophysiological findings [12] demonstrated that early event-related potentials differ between emotional and neutral prosody but failed to identify differences between emotions. These findings suggest that processing of emotional voices within the AC might primarily reflect a discrimination between emotional and neutral stimuli only, whereas categorization of emotions might occur at later stages; e.g., within the frontal cortex [13, 14]. However, conventional approaches have important limitations for determining how information is represented within cortical areas [9].

Here, we asked whether vocal emotions might be represented by specific spatial distributions in voice-processing modules. Participants listened to pseudowords spoken in five emotions and performed a gender discrimination task during event-related fMRI. Participants correctly classified the gender for 94 ± 1% of trials, indicating reliable performance throughout the experiment.

On the basis of previous observations showing a strong overlap between emotion- and voice-sensitive regions [15], we defined the most voice-sensitive voxels within the AC (i.e., the STG and Heschl's gyrus) for each subject by using an fMRI "voice localizer" [1]. As expected, a conventional analysis of this localizer revealed bilateral activation within the mid STG and Heschl's gyrus (Figure S1, available online). We then performed MVPA on these voxels and systematically varied their number (from 25 to 1800) to determine the optimal scale for decoding.

Results showed a reliable discrimination between the five categories in both hemispheres. Decoding accuracy improved with increasing voxel number but then leveled out (at ~400 voxels) for both hemispheres (Figures 1A and 1B, dashed lines). Optimal decoding was obtained with 1000 voxels (28.6 ± 1.4%) and 600 voxels (30.3 ± 1.2%) for the right and the left side, respectively. Discrimination was also significantly better ($p < 0.01$) when data from the bilateral AC, rather than the unilateral AC, were used (Figure 1C, dashed lines). These findings accord with neuropsychological studies reporting only mild deficits in vocal emotion recognition after unilateral damage of the STG but devastating impairments for prosody comprehension after bilateral lesions [16].

To clarify whether decoding benefited from the inclusion of voice-sensitive voxels or more general auditory responses,

*Correspondence: thomas.ethofer@med.uni-tuebingen.de

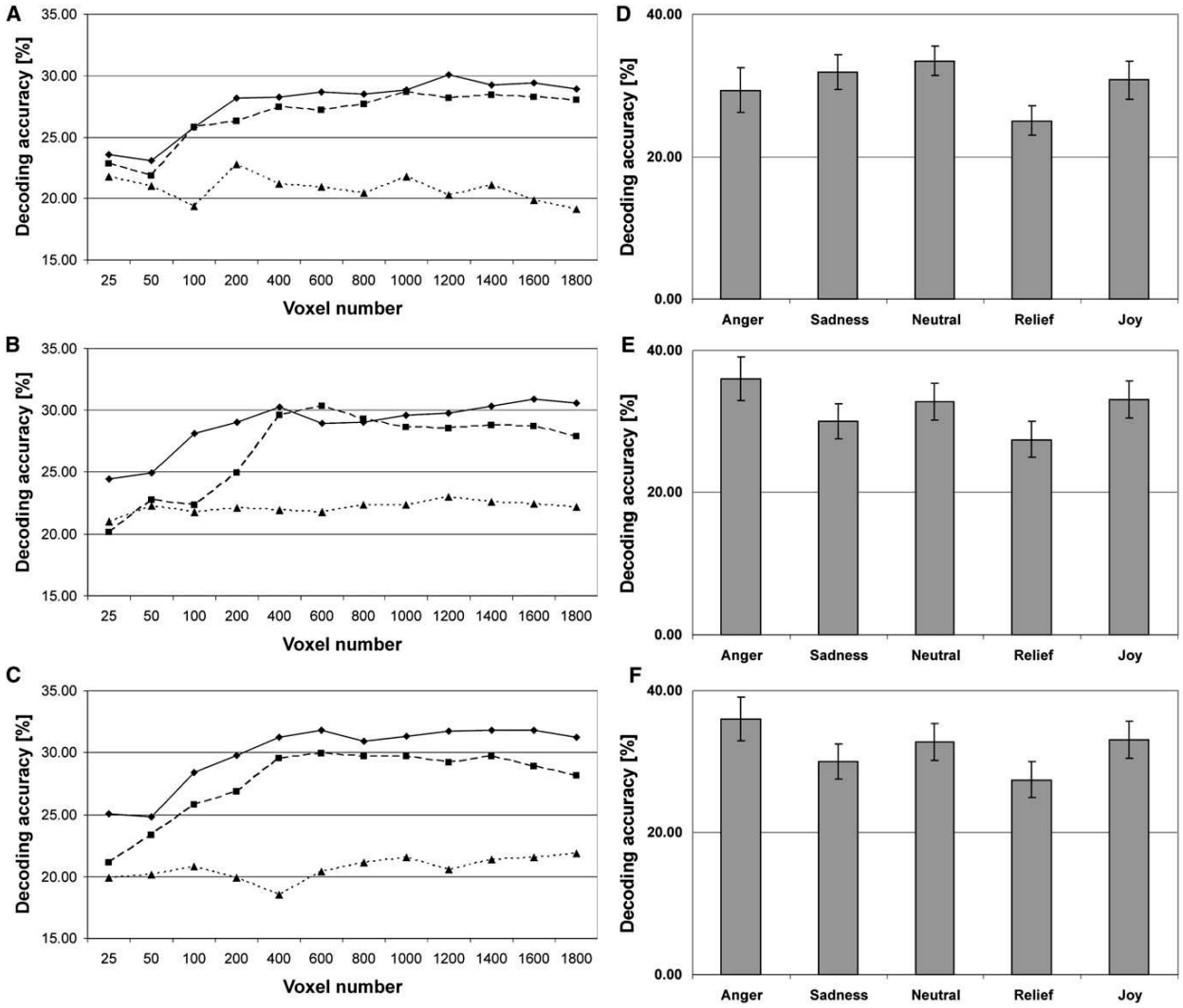

Figure 1. Accuracies for Decoding of Emotional Prosody

Decoding accuracies obtained for smoothed data (solid lines), unsmoothed data (dashed lines), and average amplitude data (pooled across voxels, dotted lines) via the right AC (A), the left AC (B), and the bilateral AC (C). Decoding accuracies (mean ± standard error) for each of the five emotional categories obtained at the optimal number of voxels with 10-mm-smoothed data for the right AC (D), the left AC (E), and the bilateral AC (F). An accuracy of 20% denotes chance level (for discriminating one out of five possible categories).

we performed an analogous analysis, using bilateral AC voxels that were the least versus the most voice-sensitive. Irrespective of the number of voxels, accuracy rates were always lower when the least voice-sensitive voxels were used ($p < 0.001$) and approached the results obtained with the most voice-sensitive responses only when nearly all voxels in the AC were included in the analysis (Figure S2), suggesting a preponderant role for voice-specific activity in successful decoding.

Previous studies using MVPA to discriminate between perceptual or cognitive states used either no spatial smoothing kernels [17–21] or very small spatial smoothing kernels [22] on the basis of the assumption that the information used for decoding is represented by local response differences of nearby voxels. Here, we examined whether the information employed for decoding of vocal emotions was represented at such a fine-grained scale (i.e., by subtle differences between adjacent points on cortical surface) or represented at a larger scale (i.e., involving more distant cortical points corresponding to segregated subregions). Smoothing of fMRI data should have opposite effects for these two encoding schemes: decoding from large-scale representations could benefit from improved signal-to-noise ratio within subregions, whereas decoding from fine-scale representations would suffer from a loss of contrast between nearby sites. In our case, MVPA after smoothing yielded accuracy rates that were 1%–1.5% higher ($p < 0.001$) than those obtained with unsmoothed data (Figures 1A–1C, solid lines). This suggests that the relevant information conveying emotion from prosody is likely to be encoded at a relatively coarse scale, possibly by the interrelationship of several subregions, rather than by more millimetric patterns expressed at the voxel level. Again, similar accuracies were obtained for both hemispheres, the highest accuracies being obtained for 1200 voxels (30.1 ± 1.4%) and 1600 voxels (30.9 ± 1.1%) in the right

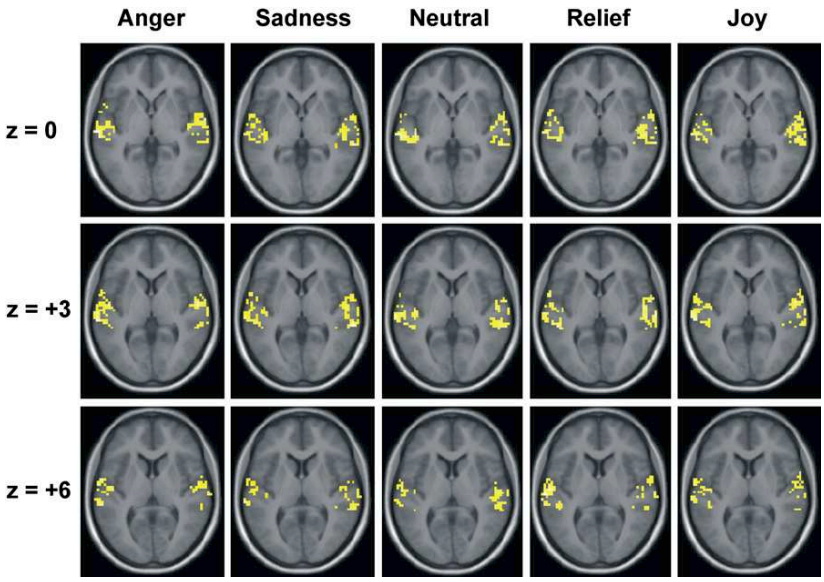

Figure 2. Mapping of the Most Informative Voxels within the Auditory Cortex

Distribution of the 400 most informative voxels for each of the five emotions, as determined by the SVM weights rendered on transversal slices (z = 0, z = 3, and z = 6) of the average normalized brain of the study participants.

and the left AC, respectively, and accuracies were significantly ($p < 0.005$) higher when the bilateral AC was used, the best discrimination being obtained for 1400 voxels (31.8 ± 1.4%).

Most importantly, for voxel numbers greater than 200, accuracy was significantly higher than chance level (20%) for all five categories (all $p < 0.01$; Figures 2D–2F), indicating that distinctive sensory information of vocally expressed emotions is represented by specific patterns of spatial activation. In addition, to rule out the possibility that decoding of vocal emotions was driven by the spatial pattern evoked by the identities of actors expressing this emotion, we trained the classifier on the stimuli of nine speakers and assessed performance for the tenth speaker. Resulting decoding accuracies were only minimally lower (30.3 ± 1.4%), indicating that emotion classification generalized across speakers.

Having established that a spatial code can be employed to decode emotional prosody, we examined whether the average response magnitude in the most voice-sensitive voxels might similarly be used for decoding. Accuracy rates obtained with the average magnitude were much lower than those obtained by decoding of spatial patterns and did not differ from chance level (18.5%–22.9%), irrespective of the number of included voxels (Figures 1A–1C, dotted lines). Furthermore, discrimination accuracy was never significantly higher than chance level for more than one category, indicating that averaging across voxels, as done in conventional fMRI analyses, degrades crucial information that is necessary for the decryption of vocal emotions.

Finally, to determine whether distinct spatial signatures for each of the five prosody categories might be used, we trained the SVM for subsequent pairwise decoding. All pairwise comparisons based on spatial patterns in the bilateral AC at the optimal number of features (1400 voxels) yielded accuracies that were significantly higher than chance level (Table 1), indicating that each category was represented by a characteristic spatial pattern. Given that decoding with various numbers of voxels revealed that accuracy reached a plateau around 400 voxels (Figure 1), we extracted the SVM weights that characterize the importance for classification across voxels and then mapped the 400 most important voxels on the mean normalized brain of our participants (Figure 2). This analysis revealed that, for all five categories, the most informative voxels were widely distributed. On average, these maps showed an overlap with each other for approximately 50% of the voxels, and about 25% of these voxels were included in all five maps. These common voxels were mostly situated in the mid STG, confirming the key role of this region in processing emotion in voices [4–8]. Remarkably, categories that were either both high arousing (i.e., anger and joy) or both low arousing (i.e., sadness and relief) exhibited a stronger overlap (55.5% of voxels for anger compared to joy, and 63.3% for sadness compared to relief) than did emotional categories that differed in arousal (46.5%–47.5%) or comparisons between individual emotional categories and neutral prosody (44.5%–52.5%). Likewise, pairwise comparisons between categories (Table 2) showed the greatest confusion between emotions with similar arousal (sad versus relief, joy versus anger) but good discrimination between emotions with a similar negative valence (anger versus sad) or a similar positive valence (joy versus relief). These findings concur with psychological [23] and neural [24] accounts postulating that arousal is a key dimension defining different emotion categories.

It must be noted that vocally expressed emotions differ in several acoustic parameters [25]. In particular, fundamental frequency (*F0*) is an important parameter for expression of emotional arousal [26], and consequently, anger and joy were characterized by a higher *F0* than were other categories. Better discrimination rates between emotions that strongly differ in *F0* converge with previous MVPA results demonstrating that the distinctiveness of activation patterns correlates with differences in speaker *F0* [11] and suggests that *F0* might encode several voice features potentially reflected in spatial activation patterns. However, the fact that we could differentiate between emotions with similar *F0* (e.g., anger versus joy) indicates that decoding did not depend solely on *F0*. Recent work [27] demonstrates that *F0* is only one of the

Table 1. Decoding Accuracies for Pairwise Comparisons

	Sadness	Neutral	Relief	Joy
Anger	65.8 ± 2.6 % p<0.001	69.0 ± 2.9 % p < 0.001	63.4 ± 3.1 % p < 0.001	55.6 ± 2.4 % p < 0.05
Sadness		55.8 ± 2.4 % p < 0.05	56.3 ± 2.1 % p < 0.01	64.4 ± 3.1 % p < 0.001
Neutral			59.4 ± 2.3 % p < 0.001	70.0 ± 2.5 % p < 0.001
Relief				60.1 ± 2.5 % p < 0.001

All values represent mean ± standard error. p values were calculated by random-effects analyses against chance level (50 %).

Table 2. Valence, Arousal, and Acoustic Parameters of the Stimuli

	Valence	Arousal	Mean *I* [a.u.]	Mean *F0* [Hz]	Duration [sec]
Anger	−1.64 ± 0.26	3.74 ± 0.24	70.14 ± 0.03	271.62 ± 66.28	1.89 ± 0.42
Sadness	−1.38 ± 0.29	0.61 ± 0.33	69.96 ± 0.39	156.10 ± 44.12	1.88 ± 0.39
Neutral	0.09 ± 0.29	1.57 ± 0.35	70.09 ± 0.05	155.20 ± 39.87	1.88 ± 0.57
Relief	0.91 ± 0.35	1.59 ± 0.40	70.07 ± 0.04	188.07 ± 32.62	2.03 ± 0.49
Joy	1.69 ± 0.19	3.36 ± 0.40	70.14 ± 0.04	298.78 ± 51.51	1.89 ± 0.45

All values represent mean ± standard deviation. a.u.: arbitrary units, Hz: Hertz.

important parameters of denoting a specific type of emotion, but other features such as timbre may play an equal function. Moreover, previous fMRI results [8] showed that the activation of STG was driven mainly by intensity and duration of stimuli, more than by their *F0*, although this study did not employ MVPA. We used natural stimuli because artificial manipulations of acoustic parameters would change or even abolish the original emotional signal. Nevertheless, future studies using systematically manipulated stimuli might help to address the question of which parameters (or combinations thereof) are most important for recognizing a particular emotion at both behavioral and neural levels.

To our knowledge, this is the first study demonstrating that vocal emotions are spatially encoded in the human AC and that such patterns can be decoded by using MVPA. Although conventional neuroimaging studies [4–8] showed stronger response amplitudes to emotional prosody in the right AC as compared to the left AC, our data demonstrate that relevant information is bilaterally represented, consistent with our interpretation that it reflects auditory cues useful for emotion recognition, rather than emotional categories per se. The wide distribution of the informative voxels is also in agreement with a recent model on the processing of vocal emotions [13] suggesting that various subregions of the STG subserve extraction and representation of suprasegmental information.

Comprehension of emotional prosody is crucial for social functioning [28] and compromised in various psychiatric disorders, including schizophrenia (deficits for anger and sadness), [29], bipolar affective disorder (deficits for fear and surprise) [30], and depression (deficits for surprise) [31]. Future research might apply an approach similar to ours to clarify whether these deficits are paralleled by activity changes blurring emotions at the level of the AC or are due to disrupted patterns within frontal regions [32, 33] reflecting biased interpretation of emotional signals.

Our new findings also open exciting avenues for emotion research in other modalities (e.g., vision, smell), as well as between sensory modalities (i.e., in supramodal brain areas [34, 35]). Thus, intriguing issues to be addressed in future studies include whether emotions perceived in the visual modality, such as facial or body expressions, are similarly represented in a distributed manner within the network of face-sensitive [36] and body-sensitive regions [37] or as a fine-grained pattern within specialized areas, such as the fusiform face area [38] and the extrastriate body area [39].

Experimental Procedures

Subjects, Stimulus Material, and Experimental Design

Twenty-two right-handed healthy subjects (13 females; 26.3 ± 7.7 years) participated in the fMRI experiment. The study was approved by the ethical committee of the University of Geneva.

Ten actors pronounced the pseudosentence "Ne kalibam sout molem" in five different categories (anger, sadness, neutrality, relief, joy). These recordings were normalized to the same mean sound intensity (*I*) and evaluated by 24 subjects (12 females; 28.5 ± 4.5 years) to ensure that the intended emotion was recognized by at least 70% of the subjects. Furthermore, 14 subjects (7 females; 28.6 ± 4.6 years) rated valence and arousal expressed by prosody. For each stimulus, the mean *I* and mean *F0* were determined with Praat software (http://www.praat.org [40]). Table 2 shows valence and arousal ratings, in addition to acoustic parameters of the stimuli. Example stimuli are provided in the Supplemental Data.

All stimuli were presented twice in pseudorandomized order and jittered relative to scanning in steps of 850 ms (intertrial interval: 6.8–10.2 s). Subjects were instructed to classify the gender of the speaker as accurately and quickly as possible.

A voice localizer was run in each participant, in a passive-listening block design with 32 stimulation and 16 silent epochs (each 8 s), as validated in previous research ([1] and http://vnl.psy.gla.ac.uk/). Stimuli included 16 blocks with human voices (HV; e.g., speech, sighs, laughs), eight blocks with animal sounds (AS; cries of various animals), and eight blocks with environmental sounds (ES; e.g., doors, telephones, cars).

Image Acquisition

Structural T_1-weighted images (TR = 1900 ms, TE = 2.32 ms, TI = 900 ms, voxel size: 0.9 × 0.9 × 0.9 mm^3) and functional images (30 axial slices, slice thickness 4 mm + 1 mm gap, TR = 1.7 s, TE = 30 ms, voxel size: 3 × 3 × 5 mm^3) were acquired with a 3T scanner (Siemens TRIO, Erlangen, Germany). Time series consisted of 509 images for the main experiment and 242 images for the voice localizer. For correction of image distortions, a field map (36 slices, slice thickness 3 mm + 1 mm gap, TR = 400 ms, TE[1] = 5.19 ms, TE[2] = 7.65 ms, voxel size: 3 × 3 × 4 mm^3) was acquired.

Conventional fMRI Analysis

Images were analyzed with statistical parametric mapping software (SPM5, Wellcome Department of Imaging Neuroscience, London, UK). Preprocessing comprised realignment, unwarping [41], slice time correction, and normalization into MNI space (Montreal Neurological Institute [42], resampled voxel size: 3 × 3 × 3 mm^3). Images were additionally smoothed with a Gaussian filter (10 mm full width at half maximum). Statistical analysis was based on a general linear model [43]. Events with missed responses (<1% of trials) were excluded from analysis. To test for the effect of smoothing on decoding, we estimated an additional statistical model by using unsmoothed data after otherwise identical preprocessing.

Pattern Classification

To select the voxels as feature vectors for MVPA, we used the voice localizer to define voice-sensitive voxels, by contrasting responses during one's perception of HV with those during one's listening to AS and ES. Voxels within Heschl's gyrus and the STG were defined by the automatic anatomic labeling toolbox [44], then ordered on the basis of their *t* values, and the most significant ones were then selected as features. The features' values were obtained from single-trial beta images estimated by conventional analysis of fMRI data. The SPIDER toolbox, available at http://www.kyb.tuebingen.mpg.de/bs/people/spider, was employed, and a linear SVM was trained with the use of all trials except for one—that is, the stimulus to be classified (leave-one-out procedure). The multiple classes were dealt with by the standard SVM voting mechanism. So that the classification algorithm was not biased by differences in overall activation between conditions, the mean beta estimates of activity within the selected voxels were subtracted for each category before the features were submitted to the SVM. Pairwise comparisons were used to test whether each category could be successfully identified against all four alternatives. In order to visualize the most informative voxels for each category, we mapped the absolute SVM weights (averaged across trials and subjects) back to the MNI brain

anatomy. To obtain voxels that were the most important for decoding a certain category against all alternatives, we calculated minimum SVM weights for the five categories against their respective four alternatives and displayed the 400 most informative voxels.

Supplemental Data

Supplemental Data include two figures and can be found with this article online at http://www.cell.com/current-biology/supplemental/S0960-9822(09)01053-7.

Acknowledgments

This work was supported in part by the Centre d'Imagerie Biomédicale (CIBM), the Société Académique de Genève, and a grant from the Swiss National Science Foundation (51NF40-104897) to the National Center of Competence in Research (NCCR) for Affective Sciences. The authors thank Tanja Bänzinger for recording of the stimuli and Anne Boesch for help in preparing the stimuli.

Received: December 19, 2008
Revised: April 13, 2009
Accepted: April 14, 2009
Published online: May 14, 2009

References

1. Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. Nature *403*, 309–312.
2. Petkov, C.I., Kayser, C., Steudel, T., and Whittingstall, K. (2008). A voice region in the monkey brain. Nat. Neurosci. *11*, 367–374.
3. Zähle, T., Geiser, E., Alter, K., Jancke, L., and Meyer, M. (2008). Segmental processing in the human auditory dorsal stream. Brain Res. *1220*, 367–374.
4. Kotz, S.A., Meyer, M., Alter, K., Besson, M., von Cramon, D.Y., and Friederici, A.D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. Brain Lang. *68*, 366–376.
5. Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., and Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. Nat. Neurosci. *8*, 145–146.
6. Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., Grodd, W., and Wildgruber, D. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. Neuroreport *17*, 249–253.
7. Ethofer, T., Wiethoff, S., Anders, S., Kreifelts, B., Grodd, W., and Wildgruber, D. (2007). The voices of seduction: Cross-gender effects in processing erotic prosody. Soc Cogn Affect Neurosci. *2*, 334–337.
8. Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., and Ethofer, T. (2008). Cerebral processing of emotional prosody–influence of acoustic parameters and arousal. Neuroimage *39*, 885–893.
9. Haynes, J.D., and Rees, G. (2006). Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. *7*, 523–534.
10. Norman, K.A., Polyn, S.M., Detre, G.J., and Haxby, J.V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. *10*, 424–430.
11. Formisano, E., De Martino, F., Bonte, M., and Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. Science *322*, 970–973.
12. Paulmann, S., and Kotz, S.A. (2008). Early emotional prosody perception based on different speaker voices. Neuroreport *19*, 209–213.
13. Wildgruber, D., Ackermann, H., Kreifelts, B., and Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. Prog. Brain Res. *156*, 249–268.
14. Schirmer, A., and Kotz, S.A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. Trends Cogn. Sci. *10*, 24–30.
15. Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., and Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. J. Cogn. Neurosci., *21*, 1255-1268.
16. Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., and Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. Brain *117*, 1283–1301.
17. Cox, D.D., and Savoy, R.L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage *19*, 261–270.
18. Haynes, J.D., and Rees, G. (2005). Predicting the stream of consciousness from activity in human visual cortex. Curr. Biol. *15*, 1301–1307.
19. Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. Nat. Neurosci. *8*, 679–685.
20. Haynes, J.D., Sakai, K., Rees, G., Gilbert, S., Frith, C., and Passingham, R.E. (2007). Reading hidden intentions in the human brain. Curr. Biol. *17*, 323–328.
21. Soon, C.S., Brass, M., Heinze, H.J., and Haynes, J.D. (2008). Unconscious determinants of free decisions in the human brain. Nat. Neurosci. *11*, 543–545.
22. Polyn, S.M., Natu, V.S., Cohen, J.D., and Norman, K.A. (2005). Category-specific cortical activity precedes retrieval during memory search. Science *310*, 1963–1966.
23. Russel, J.A. (1980). A circumplex model of affect. J. Pers. Soc. Psychol. *39*, 1161–1178.
24. Anderson, A.K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D.G., Glover, G., Gabrieli, J.D., and Sobel, N. (2003). Dissociated neural representations of intensity and valence in human olfaction. Nat. Neurosci. *6*, 196–202.
25. Banse, R., and Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. J. Pers. Soc. Psychol. *70*, 614–636.
26. Scherer, K.R. (2003). Vocal communication of emotion: A review of research paradigms. Speech Commun. *40*, 227–256.
27. Hammerschmidt, K., and Jürgens, U. (2007). Acoustic correlates of affective prosody. J. Voice *21*, 531–540.
28. Poole, J.H., Tobias, F.C., and Vinogradov, S. (2000). The functional relevance of affect recognition errors in schizophrenia. J. Int. Neuropsychol. Soc. *6*, 649–658.
29. Bozikas, V.P., Kosmidis, M.H., Anezoulaki, D., Giannakou, M., Andreou, C., and Karavatos, A. (2006). Impaired perception of affective prosody in schizophrenia. J. Neuropsychiatry Clin. Neurosci. *18*, 81–85.
30. Bozikas, V.P., Kosmidis, M.H., Tonia, T., Andreou, C., Focas, K., and Karavatos, A. (2007). Impaired perception of affective prosody in remitted patients with bipolar disorder. J. Neuropsychiatry Clin. Neurosci. *19*, 436–440.
31. Kan, Y., Mimura, M., Kamijima, K., and Kawamura, M. (2004). Recognition of emotion from moving facial and prosodic stimuli in depressed patients. J. Neurol. Neurosurg. Psychiatry *75*, 1667–1671.
32. Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., Grodd, W., and Wildgruber, D. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. Neuroimage *30*, 580–587.
33. Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., and Ackermann, H. (2005). Identification of emotional intonation evaluated by fMRI. Neuroimage *24*, 1233–1241.
34. von Kriegstein, K., and Giraud, A.L. (2006). Implicit multisensory associations influence voice recognition. PLoS Biol. *4*, e326.
35. Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., and Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. Neuroimage *37*, 1445–1456.
36. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425–2430.
37. Peelen, M.V., and Downing, P.E. (2007). The neural basis of visual body perception. Nat. Rev. Neurosci. *8*, 636–648.
38. Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. J. Neurosci. *17*, 4302–4311.
39. Downing, P.E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. Science *293*, 2470–2473.
40. Boersma, P. (2001). Praat, a system for doing phonetics by computer. Glot. Internation. *5*, 341–345.
41. Andersson, J.L., Hutton, C., Ashburner, J., Turner, R., and Friston, K.J. (2001). Modeling geometric deformations in EPI time series. Neuroimage *13*, 903–919.
42. Collins, D.L., Neelin, P., Peters, T.M., and Evans, A.C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. *18*, 192–205.

43. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., and Frackowiak, R.S.J. (1994). Statistical parametric maps in neuroimaging: A general linear approach. Hum. Brain Mapp. *2*, 189–210.
44. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage *15*, 273–289.