Auditory stream segregation relying on timbre involves left auditory cortex

Susann Deike,CA Birgit Gaschler-Markefski, André Brechmann and Henning Scheich

Leibniz Institute for Neurobiology, Brenneckestr. 6, 39118 Magdeburg, Germany

CACorresponding Author: sdeike@ifn-magdeburg.de

Received 8 April 2004; accepted 22 April 2004

DOI: 10.1097/01.wnr.0000132919.12990.34

An important aspect of auditory scene analysis is sequential grouping of sounds that are similar to one another in preference to sounds that follow one another. This grouping problem is captured by stream segregation tasks with alternating distinct sounds. We examined human auditory cortex activity with low noise fMRI in a stream segregation experiment relying on timbre differences of alternating harmonic tones (organ-like and trumpet-like). We found that stream segregation performance in comparison to monitoring a non-separable control stream increased activation exclusively in left auditory cortex and particularly in posterior areas. Our results suggest that left auditory cortex is selectively involved in this complex sequential task although the available cue for sequential grouping was timbre, usually attributed to right hemisphere analysis. *NeuroReport* 15:1511–1514 © 2004 Lippincott Williams & Wilkins.

Key words: Auditory cortex; Auditory scene analysis; Auditory stream segregation; fMRI; Sequential grouping; Timbre perception

INTRODUCTION

Selective tracking of an instrument in an orchestra or a human voice in a cocktail party requires separation of acoustic components of that instrument/voice from overlapping acoustic objects. In the search for mechanisms of separation [1] the complementary problem has been largely neglected, namely that musical or linguistic phrases become evident only if sequential elements (tones, phonemes) of the same instrument/voice are bound together over time. More specifically, sounds that are similar to one another (e.g. with respect to pitch, timbre, spatial location) must be sequentially grouped in preference to events that follow one another [2]. Experimentally, this sequential grouping process can be captured by stream segregation of alternating distinct sounds (ABAB scheme) which may perceptually split into separate streams [2]. At slow presentation rates, characteristic for speech and music (<10 Hz), sequential grouping is intentional, i.e. requires effort to attend to those sounds that are similar to one another and thus to avoid the percept of alternation. In contrast, at higher presentation rates (>10 Hz) stream segregation occurs quasi automatically. Segregation in these two temporal domains may be based on mechanisms located at different levels of auditory processing. This would be compatible with physiological results in primary auditory cortex of awake untrained monkeys [3] and ERP data in humans [4] which provide insights into the neural basis of stream segregation at fast presentation rates but not at slow rates (<10 Hz). This suggests that an additional cortical mechanism must be invoked when stream segregation is performed at these slow rates. The neural basis of this intentional stream segregation is unknown.

We examined which human auditory cortex areas are specifically involved in stream segregation at slow presentation rates using low noise functional magnetic resonance imaging (fMRI). Stimuli were sequences of harmonic tones with alternating timbre (ABAB; A: organ-like, B: trumpet-like). Perceptual selection of either the organ stream or the trumpet stream was controlled by detection of infrequent targets distributed in both streams. As control the same targets had to be detected in a non-separable stream (organ or trumpet) with the stimulus repetition rate of the double stream.

MATERIALS AND METHODS

Subjects: Nineteen right-handed normal hearing subjects (eight males and 11 females, age range 20–39 years) participated in this study. Handedness was assessed with the Edinburgh Handedness Inventory [5] (mean 92.9). All subjects were native German speakers with no special musical expertise or education. Subjects gave written informed consent to the study which was approved by the Ethics Committee of the University of Magdeburg.

Materials: Stimuli were sequences of digitally synthesised (Sound Forge 4.5) harmonic tones (200 ms duration, 5 Hz presentation rate) with alternating spectral envelopes (timbre: organ-like and trumpet-like). Fundamental frequencies of tones (261, 293, 329, 349 Hz) varied randomly in all streams to avoid habituation. Perceptually, tones with the same timbre had to be grouped into one stream, either organ or trumpet stream. In order to control for perceptual selection subjects had to indicate by right hand key press the occurrence of additional targets (level deviants +7 dB SPL, 10% target proportion) distributed in either stream. Half of the subjects had to detect the targets in the organ stream, the other half in the trumpet stream. As control the same targets had to be detected in a single organ stream or trumpet stream at 5 Hz stimulus presentation.

Stimuli were arranged in blocks of 24 s duration covering one of two conditions: double stream (n=8) and single stream (n=16, 8 for each instrument). Between the randomly distributed stimulus blocks silence blocks of the same duration were presented serving as the resting condition (Fig. 1).

For stimulus presentation and recording of behavioural responses the software Presentation (Neurobehavioral Systems, Inc.) was used. The stimuli were presented via fMRI-compatible electrodynamic headphones integrated into ear-muffs for reduction of residual background scanner noise [6]. The sound level of stimuli was individually adjusted to target audibility.

Scanning: Subjects were scanned in a Bruker 30/60 3 T head scanner. Three contiguous 6 mm slices were oriented parallel to the Sylvian fissure covering the superior temporal gyrus of both hemispheres. Functional volumes (matrix size 64 × 64, 18 cm field of view) were collected using a low-noise FLASH-based gradient echo sequence (TE/TR/flip=30.7 ms/125 ms/15°). A long gradient rise time (2500 μs) reduced the scanner noise to ~54 dB SPL at the ear. With these settings, a single volume required a scan time of ~8 s. The total experiment comprised 147 volumes scanned in 19 min 36 s. In order to obtain anatomical landmarks, functional measurements were followed by high resolution T1-weighted imaging. The subject's head was fixed with a vacuum cushion. During the whole fMRI session the subjects were instructed to keep their eyes closed.

fMRI data preprocessing: First, the subject's head motion was detected using the AIR package [7]. Data with continuous head motion greater than one voxel in at least one direction were excluded from further analysis (one subject). The remaining 18 functional data sets were analysed with the software-package KHORFu [8]. The images were corrected for in-plane head motion using the AIR package. The matrix size was increased to 128 × 128 by pixel replication followed by in-plane smoothing with a Gaussian filter (FWHM=2.8 mm, kernel width=7 mm). For each subsequent scan of the same slice, the mean intensity was computed and then scaled to the mean slice intensity average over all time points. Then each voxel time series was temporally smoothed using a moving average filter with a kernel width of two time points.

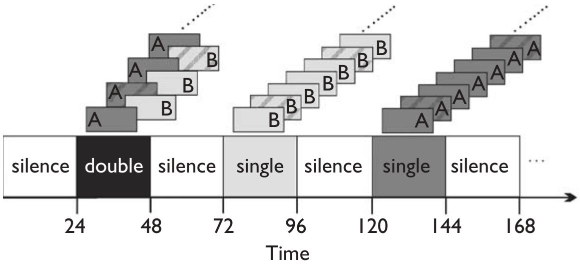

Fig. 1. Schematic diagram illustrating the experimental block design. For each stimulus condition the sequence of harmonic tones is shown. The double stream contains the classical paradigm of alternating A and B tones differing in timbre (A: organ-like, B: trumpet-like). The single stream contains sequences of tones with identical timbre, either organ- (A) or trumpet-like (B). The targets with increased sound pressure level are hatched.

Analysis of activation in Talairach space: Analysis of group data was performed with Brain Voyager 2000. After alignment to the corresponding 3D anatomical data set the fMRI data were transformed into Talairach coordinate space and analysed with the multi-subject general linear model (GLM) using the single stream and double stream condition as the two predictors. The Talairach coordinate of the activated voxel with the highest significance of the linear contrast (double stream=1, single stream=−1) was calculated.

Analysis of activation in individual regions of interest (ROI): This empirical landmark-oriented method systematised across individuals the few separate clusters of activated voxels on the superior temporal plane which are regularly seen with imaging parallel to the Sylvian fissure (for discussion see [9]). It has proven useful for regional comparison because a functional parcellation of human auditory cortex is not yet available and grand averages of brain transforms tend to blur and mislocalise activations in the superior temporal lobe due to large interindividual anatomical variability [10,11].

For each subject functional activation in each slice was analysed by correlation analysis to obtain a statistical parametric map. A trapezoid function served as correlation vector, roughly modelling the expected BOLD response. Thereby the first image of each stimulus and silence block was set to half-maximum values. Pearson's correlation analysis tested the double stream and the single stream condition *vs* rest. Voxels were accepted if they reached the significance level α=0.05 (double stream condition) and α=0.03 (single stream condition) in order to adjust the Pearson's correlation r-value (adjustment for the different number of acquired images). Only those voxels which belonged to a cluster of at least eight significant voxels were accepted.

Using 3D visualisation with Brain Voyager clusters of activation were attributed to one of four landmark-oriented ROI: TA on the planum polare anterior to the first transverse sulcus, T1 on Heschl's gyrus, T2 centred to and following the course of Heschl's sulcus and T3 with several clusters on the posterior planum temporale [9,12]. Since activations were usually not confluent between ROI no boundaries needed to be defined. In case of confluence the lowest z values were used for a separation (Fig. 2b,c). For each ROI the number of activated voxels were multiplied by their average relative BOLD signal intensity resulting in intensity weighted volumes (IWV). Across all subjects the individual difference between the double stream and single stream condition in each ROI and hemisphere was tested (two-tailed t-test, p=0.05). Thus, no comparisons are made which depend on *a priori* differences of ROI size.

RESULTS

Task performance: In each condition all subjects performed the task well above chance (one-sided χ^2-test, p=0.01; u > 2.33). In addition, the sensitivity index (d') showed no significant difference between the double stream and the single stream condition across subjects (two-tailed t-test, p=0.068).

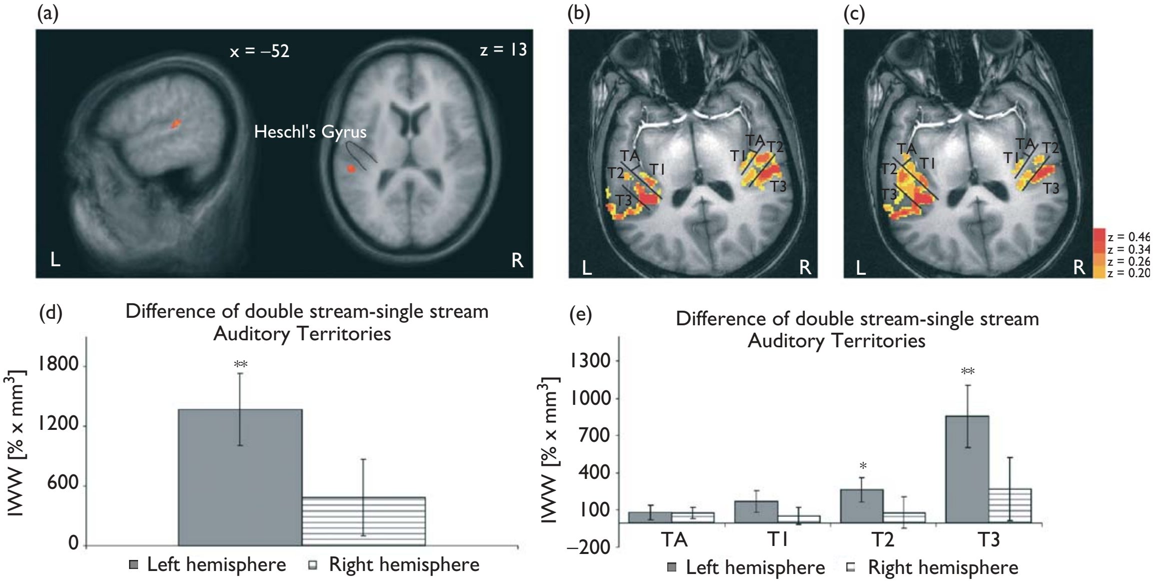

Fig. 2. (**a**) Activation cluster with highest significance of contrast between single and double stream condition resulting from the multi-subject analysis in Talairach space. This activation cluster located on planum temporale posterior to Heschl's gyrus is mapped on the group-average anatomical MR image. (**b**) Pattern of activation in one slice of an individual subject in the single stream condition and (**c**) in the double stream condition. Images show an enhanced activation in left auditory cortex in the double stream condition *vs* rest compared to the single stream condition *vs* rest. In contrast activation in the right auditory cortex is nearly unchanged. Significant activations in territories TA, T1, T2 and T3 are shown on a colour scale (z-value). (**d**) Global activation difference (IWV) in left and right auditory cortex between double stream and single stream condition resulting from the ROI-based analysis. (**e**) Activation differences in the auditory territories. A significant increase of activation was found in the left auditory cortex and particularly in the territories T2 and T3 during the double stream compared to the single stream condition.

Analysis of activation in Talairach space: The activation cluster in auditory cortex with the highest significance (t=3.6541, p=0.000263) directly testing double stream *vs* single stream condition was located in left superior temporal gyrus, posterior to Heschl's sulcus (Talairach coordinates: −52, −29, 13; Fig. 2a). The probability that the peak is located in PAC is outside the probability range of 10% according to the probabilistic map by Rademacher *et al.* [11]. The probabilistic map of the planum temporale by Westbury *et al.* [13] reveals a higher probability (26–45%) of the activation peak to be on planum temporale. This is also suggested by Figure 2a clearly showing that the position of the activation peak is posterior to Heschl's sulcus.

Analysis of activation in individual ROI: The t-test of IWV revealed a significantly stronger global activation of the left auditory cortex (p=0.002) during the double stream condition *vs* rest compared to the single stream condition *vs* rest (Fig. 2d). The global activation of the right auditory cortex did not differ significantly (p=0.368) between the two conditions. Analysis of individual territories (Fig. 2b,c) yielded an increase of activation in the left posterior auditory cortex territories T2 (p=0.02, Cohen's effect size=1.22) and T3 (p=0.005, Cohen's effect size=1.92) during the double stream compared to the single stream condition (Fig. 2e). The effect in left T3 was robust against Bonferroni-correction for multiple testing.

DISCUSSION

Our results show a selective increase in left auditory cortex activation over single stream analysis when the subjects had to group tones in the double stream condition according to timbre similarities. Stream segregation was indeed performed since the reliable detection of small changes in level was only possible by comparing sequential tones belonging to one stream because possible targets were introduced in both streams. Even though mechanistic explanations of this result are not yet available arguments can be put forward why the effect is found in left auditory cortex and why it depends on cognitive (top down) rather than on stimulus driven (bottom up) processing.

The double stream had the same stimulus repetition rate and composition of fundamental frequencies and their harmonics as the single stream but led to stronger responses in left auditory cortex. From a point of view of stimulus properties the only difference was the alternation of spectral envelopes. This alternation could lead to less habituation of neuronal responses than in the single stream i.e. to more maintained stronger responses during the stream segregation, a general habituation difference that we reduced by randomised fundamental frequencies in single and double streams. But even if the timbre alternation led to stronger responses this would not be expected in left auditory cortex. Current hypotheses would argue that timbre differences are spectral properties analysed in right auditory cortex [14–16].

There is numerous evidence that the left auditory cortex is specialised for temporal features of sounds, particularly in connection to speech analysis [17–19]. It has also been demonstrated by ERP experiments with speech and music that this left auditory cortex specialisation is not restricted to speech but also applies to music if temporal and specifically sequential analysis is required [20]. A selective left planum temporale activation with pure tone sequences combined with a sequential task was found with fMRI [21]. The stream segregation in the present experimental design sheds new light on sequential analysis. Even without a temporal cue for segregation of stimuli a sufficiently demanding sequential task (top down) challenged exclusively left hemisphere mechanisms.

Some indications of physiological correlates of stream segregation have been obtained in cortex but not yet for slow stimulation rates [3,4]. In primary auditory cortex of awake untrained monkeys neuronal responses to alternating best frequency and non-best frequency tones exhibited a magnitude contrast throughout the stream. This contrast with reliably higher responses at best frequency disappeared below 10 Hz presentation rate. This disappearance of response contrast may be explained by results of a monkey auditory cortex study in which neuronal responses to two tone sequences showed forward facilitation to second tones when dissimilar tones had large onset asynchronies (> 100 ms) [22]. Thus, a simple activation difference of neurones for different tones may be available for selective grouping but only at high alternation rates. A similar argument can be derived from an ERP-study in which only at fast pace auditory streaming was indicated by the occurrence of MMNs within separated streams [4].

Finally, there is an argument derived from stimulus repetition rate suggesting that the left T3 effect is task-related rather than stimulus-related. When a stream is perceptually singled out from the double stream the perceived stimulus repetition rate (2.5 Hz) is half that of the control stream (5 Hz). BOLD responses in human auditory cortex as a function of stimulus repetition rate were found to be lower for 2.5 Hz than for 5 Hz [23]. Therefore the increase of activation is the converse of what is expected from response magnitudes to stimulus repetition and must be task-dependent.

It is likely that the task of stream segregation at low stimulus rates involve several mechanisms including selective attention, short-term-memory-based comparisons and presumably even working memory. Selective attention was challenged in a similar fashion by target detection in double stream and control streams but may have an additional load from timbre selection in the double stream. On the other hand there is a formal similarity of stream segregation to working memory tasks, namely two-back matching to sample paradigms. But none of these possible components to date seem to provide a sufficient explanation for the selective temporal binding to fuse the events in a segregated stream to a coherent percept.

The second interesting aspect of our results is that the left hemisphere effect is mainly found in posterior areas of auditory cortex. This is in accordance with studies showing that selective tracking of one of two simultaneously playing instruments [24] and sequential comparisons of tones [21] led to selective activation on left posterior STG.

CONCLUSION

The results of the present study suggest that the left auditory cortex is specifically involved in stream segregation of sounds relying on spectral cues (timbre) for sequential grouping. Thus, in spite of the spectral cue the complex sequential analysis may determine left auditory cortex dominance. This provides novel insights into the neural basis of the cocktail party effect and selective listening to orchestral music.

REFERENCES

1. Yost WA. Auditory image perception and analysis: the basis for hearing. *Hear Res* 1991; **56**:8–18.
2. Bregman AS. *Auditory Scene Analysis*. The Perceptual Organization of Sound. Cambridge, MA: MIT Press; 1990.
3. Fishman YI, Reser DH, Arezzo JC and Steinschneider M. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res* 2001; **151**:167–187.
4. Sussman E, Ritter W and Vaughan HG Jr. An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology* 1999; **36**:22–34.
5. Oldfield RC. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 1971; **9**:97–113.
6. Baumgart F, Kaulisch T, Tempelmann C, Gaschler-Markefski B, Tegeler C *et al*. Electrodynamic headphones and woofers for application in magnetic resonance imaging scanners. *Med Phys* 1998; **25**:2068–2070.
7. Woods RP, Grafton ST, Holmes CJ, Cherry SR and Mazziotta JC. Automated image registration: I. General methods and intrasubject, intramodality validation. *J Comput Assist Tomogr* 1998; **22**:139–152.
8. Gaschler B, Schindler F and Scheich H. In: Prat A (ed.). *COMPSTAT 1996: Proceedings in Computational Sstatistics*. 12th symposium held in Barcelona, Spain, 1996. Heidelberg: Physica-Verlag; 1996, pp. 57–58.
9. Brechmann A, Baumgart F and Scheich H. Sound-level-dependent representation of frequency modulations in human auditory cortex: a low-noise fMRI study. *J Neurophysiol* 2002; **87**:423–433.
10. Penhune VB, Zatorre RJ, MacDonald JD and Evans AC. Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebr Cortex* 1996; **6**:661–672.
11. Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C *et al*. Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage* 2001; **13**:669–683.
12. Scheich H, Baumgart F, Gaschler-Markefski B, Tegeler C, Tempelmann C *et al*. Functional magnetic resonance imaging of a human auditory cortex area involved in foreground-background decomposition. *Eur J Neurosci* 1998; **10**:803–809.
13. Westbury CF, Zatorre RJ and Evans AC. Quantifying variability in the planum temporale: a probability map. *Cerebr Cortex* 1999; **9**: 392–405.
14. Zatorre RJ, Belin P and Penhune VB. Structure and function of auditory cortex: music and speech. *Trends Cogn Sci* 2002; **6**:37–46.
15. Koelsch S, Gunter TC, v Cramon DY, Zysset S, Lohmann G *et al*. Bach speaks: a cortical ''language-network'' serves the processing of music. *Neuroimage* 2002; **17**:956–966.
16. Platel H, Price C, Baron JC, Wise R, Lambert J *et al*. The structural components of music perception. A functional anatomical study. *Brain* 1997; **120**:229–243.
17. Liegeois-Chauvel C, de Graaf JB, Laguitton V and Chauvel P. Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cerebr Cortex* 1999; **9**:484–496.
18. Jäncke L, Wüstenberg T, Scheich H and Heinze HJ. Phonetic perception and the temporal cortex. *Neuroimage* 2002; **15**:733–746.
19. Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A *et al*. Lateralization of speech and auditory temporal processing. *J Cogn Neurosci* 1998; **10**:536–540.
20. Besson M and Schon D. Comparison between language and music. *Ann NY Acad Sci* 2001; **930**:232–258.
21. Binder JR, Frost JA, Hammeke TA, Rao SM and Cox RW. Function of the left planum temporale in auditory and linguistic processing. *Brain* 1996; **119**:1239–1247.
22. Brosch M, Schulz A and Scheich H. Processing of sound sequences in macaque auditory cortex: response enhancement. *J Neurophysiol* 1999; **82**:1542–1559.
23. Harms MP, Melcher and Jennifer R. Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. *J Neurophysiol* 2002; **88**:1433–1450.
24. Janata P, Tillmann B and Bharucha JJ. Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cogn Affect Behav Neurosci* 2002; **2**:121–140.

Acknowledgements: This work is integrated in the project 'Cocktail-party effect: mechanisms and plasticity of acoustical signal binding and segregation' supported by the Volkswagen Foundation.