

Searching for phonological processing in the human brain using functional MR imaging

Stefan Uppenkamp¹, Roy D. Patterson², Ingrid Johnsrude³, Dennis Norris³

¹Medizinische Physik, Universität Oldenburg, Germany, Email: stefan.uppenkamp@uni-oldenburg.de

²CNBH, Department of Physiology, University of Cambridge, U.K.

³MRC Cognition and Brain Sciences Unit, Cambridge, U.K.

Introduction

Functional magnetic resonance imaging (fMRI) allows the visualization of neuronal processes in the brain with a spatial resolution of several millimetres and a temporal resolution of a few seconds. The combination of anatomical and functional images from the same listeners is a unique tool to directly relate physiological processes to anatomical structures in the brain regions of interest. The main problem with fMRI in the auditory modality is the distinction between controlled acoustical stimuli and unwanted noise, that is produced by the MRI scanner during data acquisition. One way to solve this problem is "sparse temporal sampling", that is, acoustical stimuli and scanner noise are separated in time [1, 2].

During this study, functional MRI in a sparse imaging paradigm was used to search for phonological processing in the human brain. It is commonly assumed that, in the cochlea and the brainstem, the auditory system processes speech sounds without differentiating them from any other sounds. At some stage however, it must treat speech and non-speech sounds differently. The purpose of this study is to try and delimit the location of this stage in the auditory pathway by means of functional MRI. We assume that this is the point where the sound is found to match a specific phonological category, well before lexical or semantic processing begins.

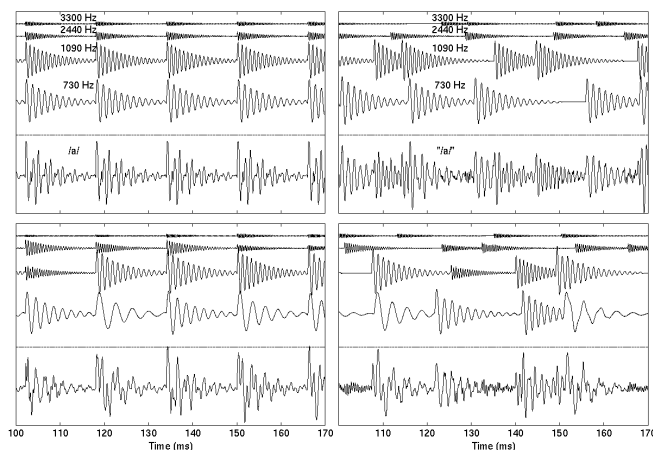


Figure 1: Four classes of stimuli constructed from sets of isolated formants (damped sinusoids). Top left: damped vowel, regular envelope onsets and fixed formant frequencies. Top right: "pathological" vowel, fixed formants but randomised envelope onsets. Bottom left: regular envelopes (produces a pitch), but randomised carrier frequencies from cycle to cycle. Bottom right: both carriers and envelope onsets randomised ("musical rain").

Stimulus generation

We had previously defined and evaluated a set of vowel and non-vowel sounds that were matched for acoustical features but which differed markedly in their similarity to speech [3]. Four damped sinusoids [4] with formant frequencies were used to produce artificial vowels; they were repeated with a period in the vocal pitch range, and had half lives that produced proper formant bandwidths. These artificial vowels automatically activate the phonological system, producing a clear speech perception. Non-speech control sounds with similar long-term distributions of energy over frequency and time are generated by randomising the envelope onsets and the carrier frequencies within each of four formant tracks. Four classes of these sounds are shown in Figure 1: speech-like with pitch or no pitch, and non-speech like with pitch or no pitch.

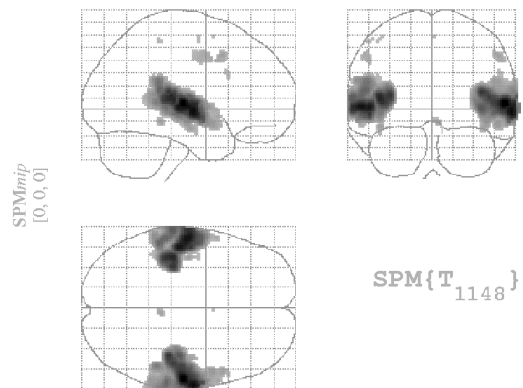


Figure 2: Group analysis of activation in response to sound. All sound conditions together were contrasted with silence, ("glass brain" view, threshold for significance $p < 0.05$, corrected for multiple comparisons). Activation is largely confined to the temporal lobes bilaterally. The main peaks are in line with the transverse temporal gyrus (Heschl's), but note additional peaks posterior to Heschl's gyrus.

fMRI data acquisition

Sparse imaging with a TR of 10 s was used to separate scanner noise and acoustic stimuli in time. Whole-brain volumes were acquired on a 3T MRI system (Bruker). In addition to the four sounds from Figure 1, there were two more conditions included as controls; these were silence and vowels recorded from a real speaker. The sounds were presented as sequences of 16 stimuli (vowels or non-vowels) of 400 ms duration each with inter-stimulus intervals between 70 and 900 ms. There was no speech-specific task, but occasionally two single sounds in the sequence were presented at

an attenuated level. The subjects were asked to press a button when this occurred to maintain their attention. Nine right-handed listeners participated in this experiment; each of the six conditions was repeated 32 times for each listener, giving 1728 scans in total.

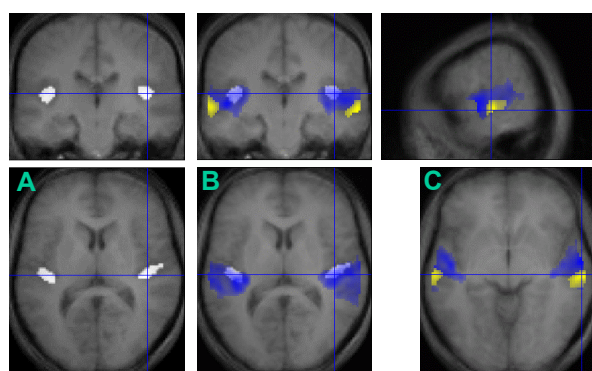
Activation in response to sound

Figure 2 shows a group analysis for all listeners of the activation in response to sound, when the five sound conditions together are contrasted with silence. Each of the five sounds, when contrasted separately with silence, produces very similar patterns of activation, that are largely confined to the temporal lobes bilaterally. The main peaks are very consistent across conditions. Nevertheless, the vowel sounds lead to additional peaks in an area posterior to the line connecting the main peaks of activation which is assumed to be Heschl's gyrus.

Activation specific to vowel and non-vowel sounds

Contrasts between vowel sounds

There was no significant activation specific to natural vowels, when contrasted with damped vowels. There was also no significant activation specific to the fixed pitch in damped vowels when contrasted with "pathological" vowels. There was, however, one area in the left hemisphere with slightly more activation for natural vowels than for "pathological" vowels. This might indicate an area sensitive to the voice-like sound quality of spoken language, rather than to the speech-like fundamental pitch.



Crosshair at: x=45, y=-22, z=10 mm x=65, y=-22, z=-5 mm

Figure 3: Activation in response to sound (vs silence) in blue, and speech-specific activation in yellow. It is rendered on the average structural MRI scan (group analysis for seven listeners). The threshold for significance was set to $p < 0.05$, corrected for multiple comparisons. The average location of Heschl's gyrus based on inspection of the individual anatomical scans is painted in white.

Contrasts between vowels and non-vowels

Similar to synthesized vowels, there was no significant effect of the presence of a fundamental pitch in the musical rain conditions. There was, however, significantly more activation in all speech-like conditions when contrasted with non-speech conditions, bilaterally, in an area below and posterior to Heschl's gyrus on the superior temporal gyrus.

Location of the activation specific to speech sounds

Figure 3 shows two sections of the mean structural scan for the nine subjects in this study. Each individual anatomical scan was inspected for the positions of Heschl's gyri (HG) by three independent judges, and the results were combined into a map of the most likely location in the group (painted in white). Figures 3B and 3C show activation maps of the two main contrasts rendered on this brain. Activation specific to vowel sounds is clearly outside HG, posterior and inferior to the main area of activity in response to sound. Whereas this activation appears bilateral and symmetric in the group, single-subject analyses revealed that some listeners show more activation on the right while others show more on the left.

Discussion

The stimuli used in the current study were designed to cover a wide range of perceived sound qualities from speech-like to completely non-speech-like, while their long-term acoustical properties were closely matched. The non-speech sounds emerge from the synthesized vowels by means of simple manipulations in the time and frequency domains. The design includes variation along two perceptual dimensions: speech-like pitch, and speech-like formant frequencies. It appears that phonological category, as determined by the presence or absence of formants, was much more influential than pitch in producing activation. The areas posterior to Heschl's gyrus that were specifically sensitive to speech sounds can be considered as candidate regions for the beginning of phonological processing. This is, however, not proof of a phonology-specific area in the brain, since similar areas are activated when contrasting sequences of musical notes with fixed pitch and moving pitch [2]. The activation may simply reflect the process of feature extraction from sounds that meet certain (unspecified) criteria.

Acknowledgment

This work was supported by the Medical Research Council (G9901257). Functional MRI was carried out at the Wolfson Brain Imaging Centre in Cambridge and the data were analysed using SPM 99 (<http://www.fil.ion.ucl.ac.uk/spm>).

References

- [1] Hall DA, MP Haggard, MA Akeroyd, AR Palmer, AQ Summerfield, MR Elliott, EM Gurney, RW Bowtell (1999) "Sparse" temporal sampling in auditory fMRI. *Hum. Brain Mapping* 7, 213-223.
- [2] Griffiths TD, S Uppenkamp, I Johnsrude, O Josephs, RD Patterson (2001) Encoding of the temporal regularity of sound in the human brainstem. *Nature Neurosci* 4, 633-637.
- [3] Uppenkamp S, A Kothari, J Bailes, RD Patterson (2001) Synthetische Vokale als Summe von modulierten Sinusschwingungen. In: *Fortschritte der Akustik - DAGA 2001*, pp. 106-107, DEGA e.V., Oldenburg.
- [4] Patterson RD (1994) The sound of a sinusoid: time-interval models. *J Acoust Soc Am* 96, 1419-1428.