Cortical Processing of Spoken Vowel-Consonant and Consonant-Vowel Syllables

UNIVERSITY OF CAMBRIDGE

Centre for the Neural Basis of Hearing

Alexis G Hervais-Adelman[1], Roy D Patterson[1], D Timothy Ives[1], Stefan Uppenkamp[2], Ingrid Johnsrude[3], Dennis Norris[4] and William Marslen-Wilson[4]

[1]Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3EG
[2]Medical Physics Section, Carl von Ossietzky-Universität Oldenburg, D-26111, Oldenburg, Germany
[3]Department of Psychology, Queen's University, 62 Arch Street, Kingston, Ontario, Canada, K7L 3N6
[4]MRC Cognition and Brain Sciences Unit, 15 Chaucer road, Cambridge, CB2 7EF, UK

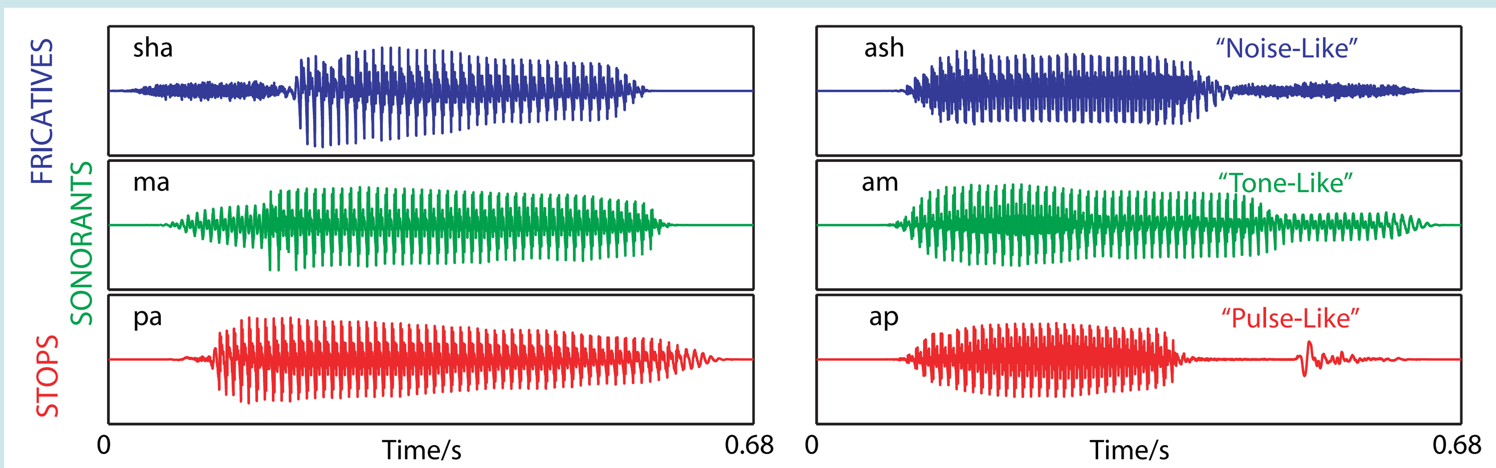
Introduction

A recent study by Uppenkamp *et al*[1] contrasted cortical responses to natural and synthetic vowels with spectrally-matched nonspeech stimuli. They showed that the initial stage of human speech processing is located bilaterally, inferior to Heschl's Gyrus (HG), in superior temporal sulcus (STS).

This region may be an initial stage of speech processing, but it is unclear whether it is specific to vowels or general to all speech.
In the present study, we extend this investigation by studying cortical responses to consonant-vowel (CV) and vowel-consonant (VC) pairs.
By comparing activations elicited by CV and VC stimuli with those produced by vowels alone, we hoped to reveal the cortical locus or loci of consonant processing.

Stimuli

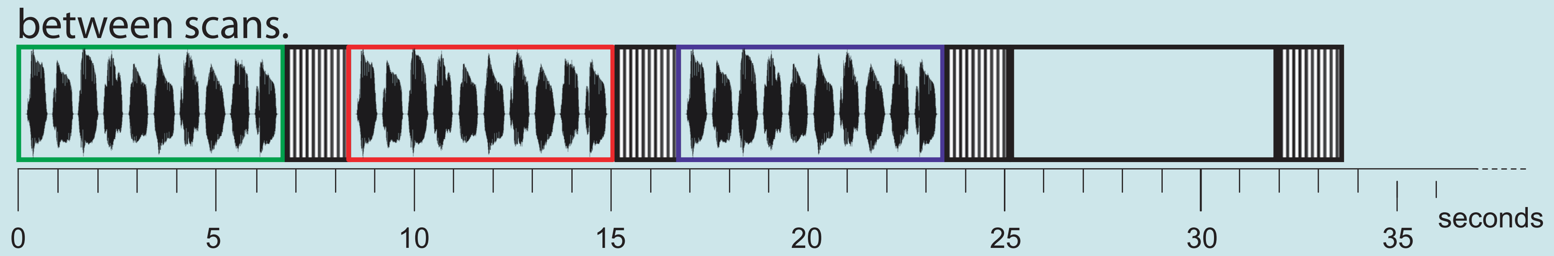
CV and VC pairs can be considered as vowels with an additional acoustic event prepended or appended. We used three classes of consonant: Fricatives (F), Sonorants (So) and Stops (St), which can be thought of as noise-like, tone-like and pulse-like, respectively. These were combined with the 5 canonical English vowels /a/, /e/, /i/, /o/, and /u/



CV, VC and Vowel (V) stimuli were taken from a syllable database[2] recorded by a male speaker.
We included a nonspeech condition: musical rain (MR) and silence, yielding a total of 9 conditions: CVF, CVSo, CVSt, VCF, VCSo, VCSt, V, MR and silence.

Procedure

Sparse imaging[3], block design. TR: 8400 ms; TA: 1600 ms. 3T Bruker Medspec Scanner.
Slice thickness: 4 mm, interslice gap: 1mm. 21 axial slices, angled away from the eyes.
12 right-handed, native English-speaking participants.
Each stimulus consisted of 10 randomly-assorted items from one condition. 24 stimulus were produced per condition. These were presented in the silent intervals between scans.



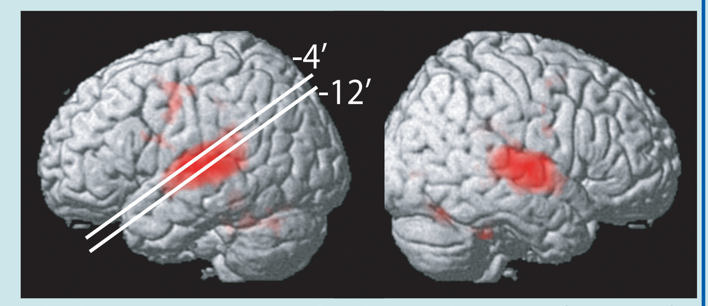
3 scanning sessions, each consisting of 8 blocks from each condition, 72 blocks per session, presented in a randomised order. Session order was randomly ascribed to participants.

Task: Button press in response to a quieter item in the sequence. 3 blocks of each sound condition had one of items 5-9 attenuated by 10dB. These trials were randomly distributed throughout the experiment.

Results

2-step analysis in SPM5 (fixed effects analyses on individual participant's data, followed by random effects analysis). Random effects (2nd level) analyses significant at both FDR-corrected $p<0.05$ and uncorrected $p<0.001$ are presented below.
Results are shown projected onto a canonical structural image.

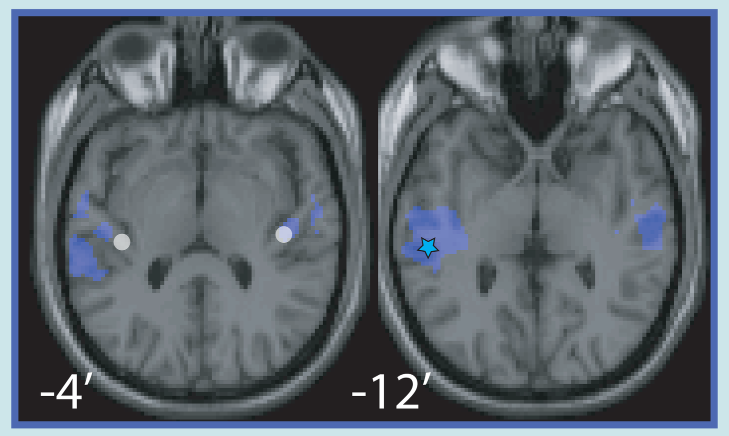
All Sound - Silence



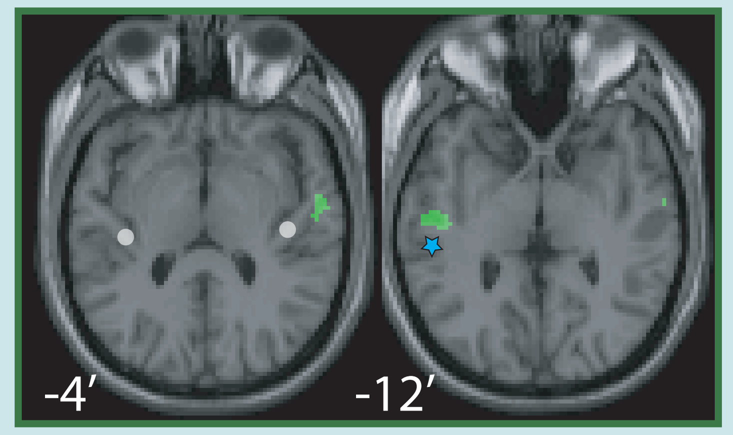
A control contrast between all of the sound conditions and the silence condition (right) showed the typical pattern of bilateral activation in superior temporal gyrus (STG) about the primary region in medial Heschl's gyrus (mHG).

The activation is largely confined to two quasi-axial slices, one just below the surface of the temporal lobe (marked - 4' in the figures) and the other lower down in the superior temporal sulcus (STS) (marked -12' in the figures). There was no overall significant difference between activation produced by CV and VC syllables, so the data from the CV and VC conditions were combined.

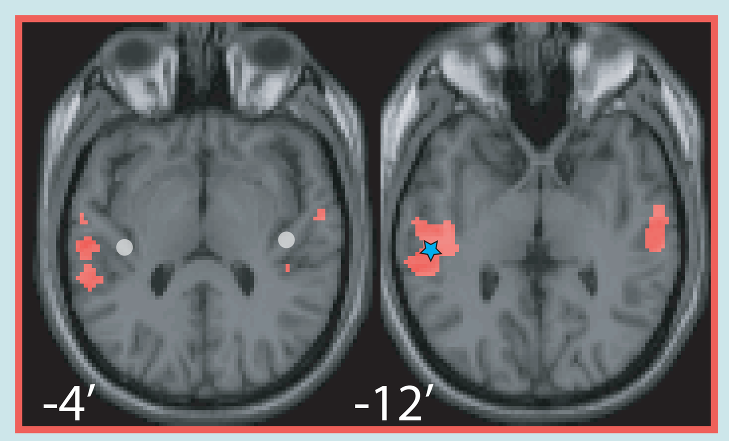
Fricative Syllables - Vowels



Sonorant Syllables - Vowels



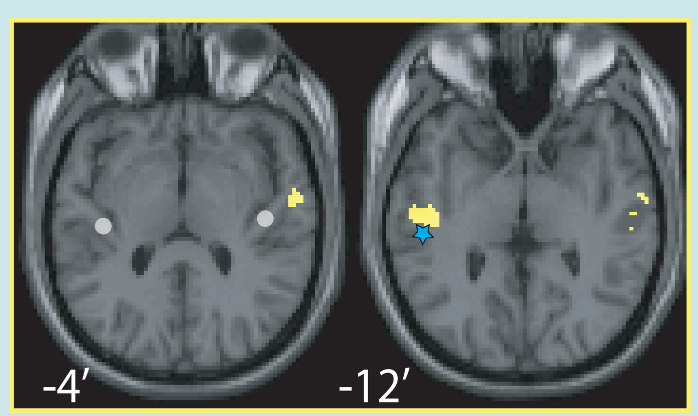
Stop Syllables - Vowels

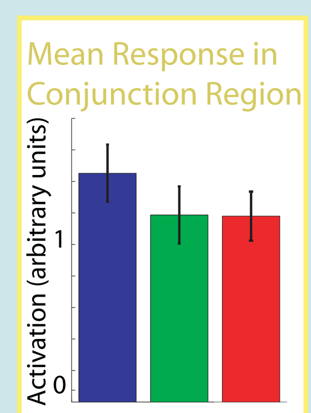


In general, the syllables produced more activation than the vowels, and all three of the syllable types produced more activation than vowels in regions along the surface of the temporal lobe (slice -4') which includes primary auditory cortex (white circles) and in regions of the STS (slice -12') which includes the region where vowels produce more activity than musical rain (blue star).

The activation in the upper slice probably represents the processing of acoustic properties of the consonants, i.e., whether the non-vowel part of the sound is noise-like, tone-like or pulse-like. It is the activation in the lower slice that is more likely to represent the identification of the sound as a syllable, or the recognition that it is a specific class of syllable.

Conjunction of Consonants

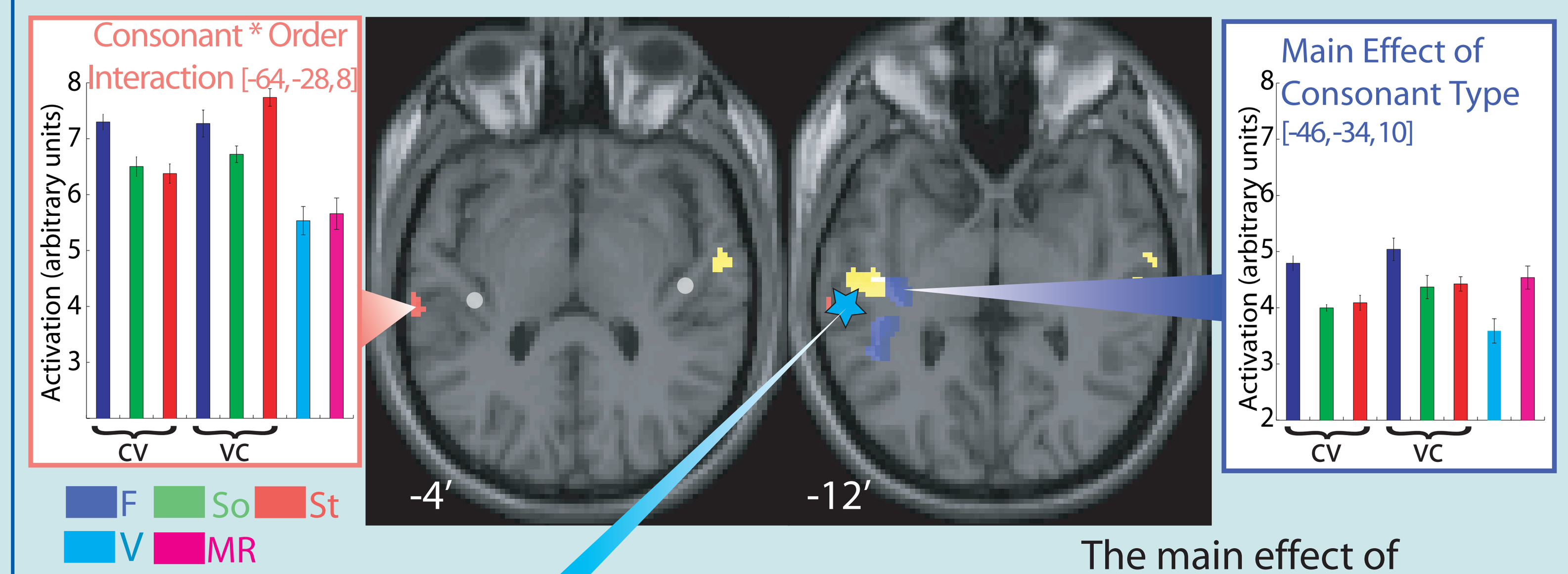




There is a region of overlap between the 3 consonant types, which does not show significantly different activation in response to different consonants

Results (cont.)

An ANOVA was carried out to test for a main effect of consonant type, order (CV vs VC) and the interaction between the two. There was no main effect of order. There is a significant main effect of consonant type and a consonant by order interaction. These are shown below alongside the conjunction of the 3 consonants.



The interaction is driven by a greater response to VCSt syllables than CVSt. VCSt have a brief silent interval between V and C. This region may respond to acoustic onsets and hence be more active in this condition.

Vowels - MR. There is significantly more activation in response to vowels than MR in mid STS [peak voxel: -62,-28,4]. Uppenkamp *et al*[1] interpreted this contrast to mean that phonological processing begins at, or after, this point in the auditory pathway.

The main effect of consonant type is driven by the significantly greater response to fricatives than to stops or sonorants, along deep STS. In contrast to stops and sonorants, fricatives are a form of broadband noise; this region may be involved in determining when a noise-burst is a phonological cue.

The consonant processing examined in this study occurs outside the "vowel region" reported by Uppenkamp *et al*[1] and replicated here. This suggests that the initial stages of vowel and consonant processing occur in distinguishable cortical regions.

Conclusions

These results demonstrate that the initial stages of consonant processing occur in regions separate from the regions involved in the initial stages of vowel processing. The regions involved in consonant processing on the surface of the temporal lobe are probably processing the acoustic properties of the consonants; that is, whether they are noise-like, tone-like or click-like. The consonant regions in STS are more likely to be the site of the initial stages of phonological processing, that is, identifying that the sound is a syllable of speech and what its class is. The results suggest that linguistic processing has probably not begun by this stage in the system.

References:
[1]Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., & Patterson, R. D. (2006). Locating the initial stages of speech-sound processing in human temporal cortex. Neuroimage, 31(3), 1284-1296.
[2]Ives, D. T., Smith, D. R., & Patterson, R. D. (2005). Discrimination of speaker size from syllable phrases. J Acoust Soc Am, 118(6), 3816-3822.
[3]Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al. (1999). "Sparse" temporal sampling in auditory fMRI. Hum Brain Mapp, 7(3), 213-223

Acknowledgements:
Research supported by the UK Medical Research Council (G0500221, G9900369).