The temporal envelope of speech is represented on multiple time scales

York Neuroimaging Centre
THE UNIVERSITY of York

Rebecca E. Millman[1] and Philip T. Quinlan[2]
[1]York Neuroimaging Centre, The Biocentre, York Science Park, Heslington, York, YO10 5DG.
[2]Department of Psychology, University of York, York, YO10 5DD,
Email: rem@ynic.york.ac.uk

Background

The temporal envelope of speech
The acoustic structure of speech contains relevant information on multiple time scales. Slow fluctuations (<50 Hz) in amplitude over time are described as the speech temporal envelope (Rosen, 1992).

The temporal envelope of speech is sufficient for speech intelligibility, at least in quiet (e.g. Drullman et al., 1994).

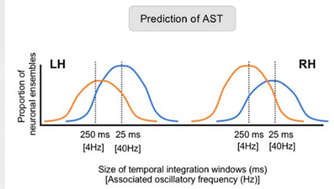
EEG/MEG frequency bands and speech
The speech signal contains more than one time scale relevant to auditory cognition (e.g. syllabic vs. phonemic) (e.g. Poeppel, 2003).

Time scales are consistent with some classical EEG/MEG frequency bands:

"Classical" frequency band	Frequency	Time scale	Role in spoken language comprehension?
delta	1-4 Hz	250-1000 ms	syllabic rate of "clear speech" (e.g. Abrams et al., 2008)
theta	4-8 Hz	125-250 ms	mean syllabic rate (e.g. Poeppel, 2003; Greenberg, 1999) sensory-memory comparison (e.g. Ghitza & Greenberg, 2009)
alpha	8-13 Hz	77-125 ms	mean segmental rate (Greenberg, 1999)
gamma	25-80 Hz	13-40 ms	segmental processing, phonemic representation (e.g. Poeppel, 2003)

Table 1. Functionality of time scales implicated in auditory cognition.

The Asymmetric Sampling in Time (AST) model



Temporal integration is reflected as oscillatory neuronal activity in different frequency bands, namely theta (4-8 Hz) and gamma (28-40 Hz) activity.

AST model predicts that speech drives gamma activity (segmental) in left cortical areas and theta activity (syllabic) in right auditory cortical

Figure 1. Schematic of AST model (Giraud et al., 2007).

Aims

We used noise-vocoded single words as speech stimuli, which are less acoustically complex than natural speech. In noise-vocoded stimuli, the speech signal is not distorted but the original temporal envelope is presented with a reduced amount of spectral information.

MEG beamformers were used to determine **WHERE** in the brain the temporal envelope of speech is represented and **HOW** the speech temporal envelope is represented in terms of power changes in different frequency bands. In conjunction with the beamforming analyses, we also used virtual electrode analyses to show the **TIMING** of changes in power in locations identified by the beamformer.

Predictions

• Lateralisation of representation of speech temporal envelope? To which hemisphere? Effects of attention on lateralisation?
Intelligible speech is encoded by the phase angle of the theta (4-8 Hz) frequency band, and this mechanism is lateralised to the right hemisphere (Luo & Poeppel, 2007). An alternative model of auditory processing predicts that the left hemisphere is specialised for temporal processing and the right hemisphere specialised for spectral processing (e.g. Zatorre et al., 2002). The processing of the noise-vocoded words (i.e. only temporal cues) is slightly lateralised to the left hemisphere (Obleser et al., 2008). Moreover, attention may lateralise sound processing to the left hemisphere (e.g. Poeppel et al., 1996).

• How is the temporal envelope of speech represented?
Low temporal envelope modulation frequencies (< 16 Hz) are most important for speech intelligibility in quiet (e.g. Drullman et al., 1994). Previous studies (e.g. Ahissar et al., 2001; Abrams et al., 2008) show that human auditory cortex can phase-lock to the temporal envelope of speech when the same/similar speech material is presented repeatedly. In this study each noise-vocoded word has a *different* temporal envelope. Is there phase-locking to the average of many different temporal envelopes? What are the contributions on both phase-locked and non-phase-locked activity to temporal envelope processing?

• The mechanism/s underlying comprehension of single words and discrimination of sentences (Luo & Poeppel, 2007) may not be same.
Luo & Poeppel (2007) reported that the phase angle of the theta frequency band could track and discriminate sentences. However, this phase angle tracking took ~2000 ms to build up relative to the onset of the stimulus. In the present study, single monosyllabic words were used as speech stimuli (mean duration of ~750 ms).

Methods

Participants
• 14 participants (right-handed and native speakers of English).

Stimuli
• Monosyllabic noise-vocoded single words were used as speech stimuli.
• Noise-vocoded words were created with a 16-channel vocoder (intelligibility ~70 % determined from our own behavourial experiments) using Matlab-based software published by Smith et al. (2002).
• Control stimulus was 16-channel vocoded noise i.e. the control stimulus contained same spectral information as the 16-channel noise-vocoded word stimuli.
• A subset of noise-vocoded word stimuli was randomly assigned to each participant from a list of 400 words. No noise-vocoded word was presented more than once within a subset.
• Each participant was presented with 120 noise-vocoded words and 120 presentations of vocoded noise.

Experimental Design
• Participants listened to stimuli presented through Etymotics insert earphones.
• Stimuli were presented in a random order whilst the participant fixated on a fixation cross.
• 12 randomly presented "catch trials", which required a button-press response from the participant, were used to encourage participants to attend to the stimuli. Following the auditory presentation of the stimulus in a catch trial, two words were presented visually on the screen on the left- and right-hand side of the fixation cross. Participants had to indicate by means of the response button which of the two words had been presented using the index finger of their left hand.

Beamformer analyses
• MEG beamformer analyses were used to localise neural sources to the speech temporal envelope. Sources were localised with a vectorised linearly-constrained minimum-variance beamformer (VLCMV) (e.g. Huang et al., 2004) using a 5-mm grid.
• The output of a beamformer spatial filter generates the ***total power*** (i.e. both phase-locked and non-phase-locked activity) at the target location over a given temporal window and within a given frequency band.
• Both the "active" (noise-vocoded words) and "control" (vocoded noise) beamformer windows were 700-ms in duration (starting 50 ms post-stimulus onset).
• Beamformer constrasts were carried out using the outputs of bandpass filters corresponding to classical frequency bands or frequency bands implicated in auditory cognition (see Table 1).
delta: 1- 4 Hz; theta: 3-6 Hz (Giraud et al., 2007); theta2: 4-8 Hz (Luo & Poeppel, 2007);
alpha: 8-13 Hz; beta: 13-25 Hz; gamma: 25-40 Hz; mid gamma: 60-80 Hz; high gamma: 80-100 Hz.
• Non-parametric *t*-tests (e.g. Nichols and Holmes, 2002) were used to determine souces with significant ($p \leq 0.05$) changes in power between the vocoded noise and the noise-vocoded words for each voxel.

Virtual electrode (VE) analyses
VEs were used to reconstruct the source activity at voxels of interest identified by the beamforming analyses.
Co-ordinates where significant changes in power were identified by the beamformer in the group image were converted back into the slice numbers for each individual i.e. VEs were calculated from equivalent locations for each participant.
Stockwell transforms (time-frequency plots) were used to determine how the power of a virtual electrode time series varies over time within the given frequency bands.
Statistical comparisons between the VEs for the noise-vocoded words and vocoded noise conditions were made using PROC MIXED in SAS (SAS Institute Inc., North Carolina, US) to compute a generalised linear mixed model (GLMM) using methods described in Cornelissen et al. (2009).

Virtual Electrode analyses

Virtual electrode analyses offer a way to take advantage of the > 1 ms temporal resolution of MEG. Our aim was to investigate where in time, relative to stimulus presentation, the power changes identified by the beamformer occured.

Beamformer localisations were determined based on total power (i.e. phase-locked and non-phase-locked activity) changes. Virtual electrode analyses enable us to carry out statistical analyses [see Cornelissen et al. (2009)] on both the phase-locked and non-phase-locked activity at each location identified by the beamformer.

The time series of virtual electrodes were reconstructed and used to generate both the phase-locked (evoked) and non-phase-locked (induced) Stockwell Transforms [http://www.cora.nwra.com/~stockwel (only one 'l'!)].

Virtual electrodes were reconstructed using the same time window as used in the beamformer analyses (50–750 ms post-stimulus presentation). However, the flexibility of virtual electrodes allowed us to reconstruct virtual electrode time series using different filter bandwidths to those that were used in the beamforming analyses.

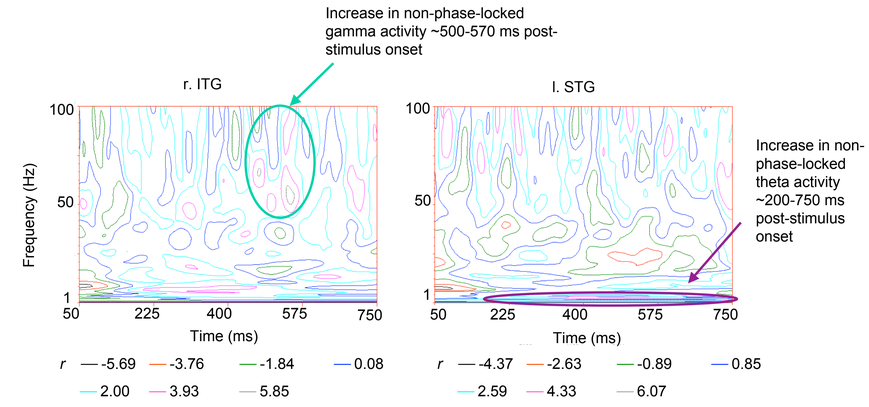


Figure 3. Non-phase-locked (induced) Stockwell Transforms generated from areas r. ITG and l. STG identified in the beamforming analyses. These Stockwell Transforms show differences in non-phase-locked activity generated by noise-vocoded words and vocoded noise.

Results

Frequency band	FSL (slice #)	TAL (mm)	Location (TAL)	*t*-value ($p \leq 0.05$)
delta	70,80,38	-47,30,9	l. IFG	10.87
(1-4 Hz)	10,40,35	64,-44,0	r. MTG	6.67
	75,53,33	-60,-21,-5	l. MTG	6.14
	68,58,20	-44,-10,-32	l. ITG	6.13
	75,58,45	-60,-10,18	l. PoCG	5.96
	18,68,55	54,10,38	r. MFG	5.92
	73,63,38	-56,1,4	l. ant. STG	5.89
theta	70,53,35	-47,-19,0	l. mid STG	6.83
(3-6 Hz)	78,53,25	-61,-18,-17	l. ITG	6.51
high gamma (80-100 Hz)	18,50,23	50,-26,-21	r. ITG	5.46

l. left; r. right; ant., anterior; IFG, inferior frontal gyrus; MTG, middle temporal gyrus; ITG, inferior temporal gyrus; PoCG, post central gyrus; MFG, middle frontal gyrus; STG, superior temporal gyrus

Table 2. Overview of locations identified by the beamformer contrast (noise-vocoded words – vocoded noise) based on significant ($p \leq 0.05$) changes in power.

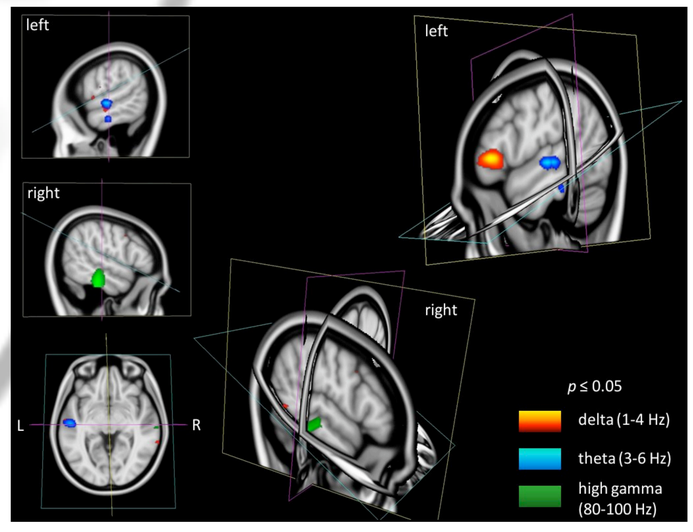


Figure 2. Group beamforming results for the beamformer contrasts (noise-vocoded words - vocoded noise).

Discussion

We found evidence that complex temporal envelopes, in this case noise-vocoded words, are represented on multiple time scales in the human brain when participants are required to attend to speech stimuli.

The beamformer localisations (based on total power changes) show that the temporal envelope of speech is represented bilaterally in temporal areas. In extra-temporal areas, there is evidence of left hemisphere lateralisation of speech temporal envelope processing.

Increases in power in the **delta** frequency band were found bilaterally in both temporal and extra-temporal areas. Increases in power in the **theta** frequency band were found in the left temporal lobe and increases in **gamma** power were found in the right inferior temporal lobe. However, the functional asymmetry in the localisations based on theta and gamma power are inconsistent with the AST model (e.g. Poeppel, 2003).

Virtual electrode analyses suggest that the beamformer localisations were based on changes in non-phase-locked (induced) activity. The VE analyses also suggest that although beamforming localised brain areas based on the power in a given frequency band, there may also be activity in other frequency bands that did not result in significant power changes in the beamformer analyses.

Previous EEG (e.g. Abrams et al., 2008) and MEG (Ahissar et al., 2001) studies showed phase-locking to the temporal envelope of speech by repeatedly presenting the same/similar sentence material. In the present study participants were presented with 120 *different* noise-vocoded words. The average of 120 *different* speech temporal envelopes would result in essentially a flat temporal envelope. Therefore it is unlikely that there would be phase-locked activity to the speech temporal envelope given the experimental design used in this study.

Phase-locking is an important mechanism for processing sounds that change over time. However, the results from the present study suggest that non-phase-locked activity also contributes to the processing of the temporal envelope of speech.

REFERENCES

Ahissar, E. et al. (**2001**). PNAS **98**, 13367-13372.
Abrams, D. A. et al. (**2008**). J. Neurosci. **28**, 3958-3965.
Cornelissen, P. L. et al. (**2009**). PLoS ONE **4**, e5359.
Drullman, R. et al. (**1994**). J. Acoust. Soc. Am. 95, 1053-1064.
Giraud, A. L. et al. (**2007**). Neuron **56**, 1127-1134.
Ghitza, O. & Greenberg, S. (**2009**). Phonetica **66**, 1-14.
Greenberg, S. (**1999**). Speech Commun. **29**, 159-176.
Huang, M-X. et al. (**2004**). Brain Topography **16**, 139-158.
Luo, H. & Poeppel, D. (**2007**) Neuron **54**, 1001-1010.
Nichols, T. E. & Holmes, A. P. (**2002**). Human Brain Mapping **15**, 1-25.
Obleser, J. et al. (**2008**). J. Neurosci. **28**, 8116-8124.
Poeppel, D. (**2003**). Speech Commun. **41**, 245-255.
Poeppel, D. (**1996**). Cog. Brain Res. **4**, 231-242.
Rosen, S. (**1992**). Philos. Trans. R. Soc. Lond. B B Biol. Sci. **336**, 367-373.
Smith, Z.M. et al. (**2002**) Nature **416**, 87-90.
Zatorre et al. (**2002**). Trends Cogn. Sci. **6**, 37-46.

ACKNOWLEDGMENT: ***This research was supported by a IRPF and a departmental grant***