Hearing versus Understanding Speech in Noisy Environments

Lee M. Miller

Center for Mind & Brain, and Section of Neurobiology, Physiology, & Behavior, University of California, Davis

INTRODUCTION

Listeners with moderate hearing loss readily hear speech but have great difficulty understanding what is said, particularly in noisy environments. This study identifies the neural networks that distinguish mere Hearing from Understanding speech, in both the auditory and audiovisual domains.

EXPERIMENTAL DESIGN

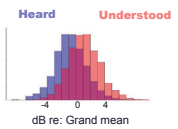
In a mixed event-related/blocked fMRI design, subjects with healthy hearing discerned utterances within a noisy background. Target stimuli were vowel-consonant-vowel utterances, with the vowel 'ah' and one of four stop consonants (/d/,/t/,/g/,/k/), filtered to approximate sloping high-frequency hearing loss and embedded in a background of 16-talker babble. All auditory stimuli were delivered ~15dB over constant scanner noise, and presented in three conditions:

1) auditory only
2) congruent audiovisual
3) incongruent audiovisual
where the video was always a single generic, average across all utterances.

Subjects indicated with a button press whether they:
i) heard an utterance but were unable to identify it ("Heard"), or
ii) heard and could identify the utterance ("Understood")
The intensity of the target utterance, and thus signal-to-noise level, was adaptively roved near each subject's Heard/Understood threshold.

This allowed us to identify brain activity related to understanding, independent of stimulus properties.

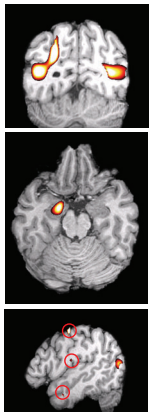
BEHAVIOR



Signal-to-noise distributions of all Heard and all Understood utterances are highly overlapping: virtually ***identical stimuli yet categorically different percept.***

Congruent audiovisual speech improves understanding relative to auditory-only or incongruent audiovisual speech.

HEARING WITH UNDERSTANDING



Left IPS

Bilateral posterior MTG

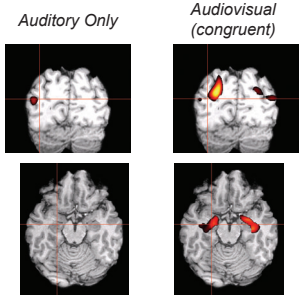
Left Hippocampus

Left anterior STS, dorsal & ventral motor/premotor

(FDR p<0.05)

Across all three conditions (auditory only, audiovisual congruent, and audiovisual incongruent) these areas showed greater activity when utterances were **understood** versus heard without understanding.

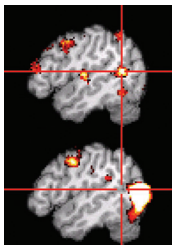
Understanding auditory-only speech compared to audiovisual (congruent) speech suggests a common neural substrate in posterior MTG / angular gyrus, but different mechanisms in IPS and hippocampus.



RELATION TO AREA MT+

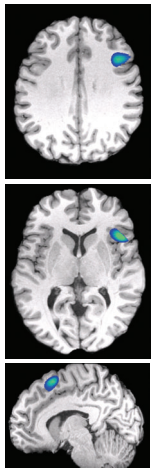
To determine whether the left middle-temporal/lateral-occipital region sensitive to understanding was visual motion Area MT+, a subset of subjects performed an MT localizer task. In this representative subject, the area reflecting understanding is adjacent but anterior to MT+.

Understood > Heard



Visual Motion

HEARING WITHOUT UNDERSTANDING



Right Precentral Sulcus

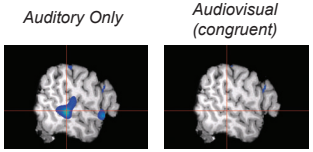
Right Frontal Operculum & Insula

Right Pre-SMA

(FDR p<0.05)

Across all conditions, these areas showed greater activity when utterances were **heard without understanding**, versus understood.

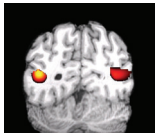
Hearing auditory-only compared to audiovisual (congruent) speech suggests similar neural mechanisms. One notable difference: heard auditory-only speech engages relatively more right IFG and right STS than audiovisual speech.

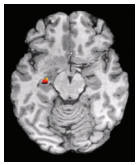


VISUAL CONTRIBUTION TO UNDERSTANDING

To identify areas that support improved intelligibility from visual speech, we conjoined Understanding > Hearing areas with those having greater mean activity for **audiovisual congruent** versus **audiovisual incongruent**.

Bilateral posterior MTG

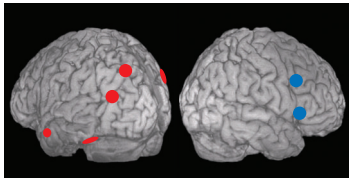




Left Hippocampus

(both contrasts FDR p<0.05)

SYNTHESIS & CONCLUSIONS



Understanding rather than merely hearing speech recruits a network of occipito-temporal and posterior parietal areas, with a left hemisphere dominance. The core of the network lies in posterior MTG/angular gyrus, an area known to play a role in comprehending language of several modalities.

Hearing but failing to understand speech engages right-hemisphere areas, notably in and near the inferior frontal gyrus. This may reflect the challenge of picking out a known speaker's voice using cues such as pitch.

Supported by the National Institutes of Health: National Institute on Deafness and other Communication Disorders (NIDCD) RO1-DC008171.

Correspondence: leemiller@ucdavis.edu

SCANNING & ANALYSIS

Twenty-five healthy subjects with normal hearing participated. Blood-oxygenation level-dependent (BOLD) activity was measured with functional MRI on a 3 Tesla Siemens TRIO scanner with 8-channel phased-array headcoil: EPIs 26 slices, 4mm thick,TR 1.5s, TE=25ms, FOV 220x220mm, flip angle 90 degrees. There were 7 functional runs, appx. 7.5 min each: 6 runs of the experimental task followed by a non-speech control task. Images were corrected for subject motion and slice-timing offset. BOLD activity was compared across conditions with a modified general linear model, using SPM2 (Matlab). All results reported are masked by an F-test conjoined across all three conditions (FDR p<0.01).