INTRACRANIAL RECORDINGS OF HUMAN SPEECH PERCEPTION

Adam Burch

PI: Michael S. Beauchamp, Ph.D.

**Abstract**

 Speech perception in humans is an important and computationally difficult problem. To study the neural mechanisms underlying speech perception, we utilized intracranial electrodes implanted on the cortical surface in patients with intractable epilepsy. We recorded the summed local-field potential (LFP) from the neurons under each electrode while the subject watched audio and visual recordings of single syllable words. We then analyzed this data by examining the gamma band oscillations (35-150Hz). The first method we used to analyze the LFPs was the gamma band response (GBR) method developed by Lachaux and colleagues (Lachaux et al. 2010). The GBR is calculated by executing a Wilcoxon Z-test on a power spectra obtained from a multitaper frequency transformation. The second method was used was an estimate of broadband power, developed by Miller and colleagues (Miller et al. 2009). Broadband power is computed by taking z-scores of power values obtained by carrying out a Hilbert transform on the bandpass-filtered data. Both methods were implemented in Matlab and tested in over 200 electrodes from four patients. Using these techniques, we found differences in the neural responses to congruent words (auditory and visual matching) and incongruent words (auditory and visual mismatch). Because direct recording of neural activity has high spatial and temporal precision, LFPs will reveal new insight into the neural mechanisms for multisensory integration and human speech perception. This work was supported by NSF REU award DMS-0755294, NSF Cognitive Neuroscience Initiative 0642532 and NIH R01 NS065395.

**Introduction**

 Speech perception is an important and computationally difficult problem. The brain processes speech using both audio and visual cues. People can comprehend speech more easily when they can see the mouth of the speaker. Sometimes, speech can be understood entirely without auditory information, as with lip reading. In our research, we are concerned with how the mind integrates the information from the visual and audio pathways from the brain as speech is perceived. The location for multisensory integration has been found to be in the Superior Temporal Sulcus (STS). Our work focuses on imaging this area as well as the visual and auditory cortices. In particular, we examined these areas’ responses to speech recordings when the auditory and visual recordings did not correspond.

 Many imaging techniques have been applied to study the brain’s response to speech. In our work, we used intracranial electrodes implanted on the cortical surface of a subject’s brain. Many laboratories, including our own, have used Magnetic Resonance Imaging (MRI) to study speech perception. This has distinct advantages, owing to its good spatial resolution, and the fact that it can image the entire brain at one time. However, MRI has a very poor temporal resolution, on the order of seconds. With intracranial recordings, we forego the advantage of whole-brain imaging in favor of excellent temporal resolution: our equipment samples the electrodes potential at 2000 Hertz. However, using implanted intracranial electrodes does have huge drawbacks; namely, the extremely invasive nature of the procedure.

 Because of the invasive nature of intracranial electrodes, we can only utilize electrodes that have already been implanted in patients for clinical reasons. The most common use of intracranial electrodes is to locate the seizure focus in patients with pharmacologically intractable epilepsy. These patients suffer frequent debilitating seizures which interfere with their daily lives, and drugs do not help them. In order to stop the seizures, a neurosurgeon can remove the section of the brain which is causing the seizures. In order to locate the seizure focus, scalp electroencephalography (EEG) is first used. If these electrodes do not provide a conclusive picture, the patient may elect to have intracranial electrodes implanted on his cortex (and in some cases inside it) in order to locate the seizure focus. After the electrodes are implanted, the patient spends several days within an Epilepsy Monitoring Unit (EMU). During this time, the electrodes detect seizures and then analyze where the focus is located. Before a patient undergoes the surgery, we ask if he or she would like to volunteer for our study. If the patient consents, he or she watches various videos and images in the EMU while the electrodes record the Local Field Potential (LFP). Because of the clinical use of the electrodes, we as researchers have no input on where the electrodes are located.

 We then utilize several methods to analyze the LFP, including Gamma Band Response (GBR) and Broadband power (BB). These methods focus on the gamma band of oscillations, from 35 Hertz to 150 Hertz. This band of frequencies has been shown to be crucial to cognition. As will be shown, the gamma band increases dramatically in power when stimuli are presented.

**Experiment**

 Between 62 and 96 intracranial electrodes each were implanted on the cortical surface of four patients with intractable epilepsy. During time in the EMU, these subjects viewed audiovisual recordings of four different single syllable words: drive, meant, last, and known. Four different categories of recordings were used: audio only, audiovisual congruent (audio matches video), audiovisual incongruent (audio does not match video), and visual only. The visual component of the videos was simply a face saying the word. Local-field potentials (LFPs) were recorded from each electrode as the subjects watched the recordings.

 The recordings were randomly ordered, with an inter-stimulus interval (ISI) of about two seconds. Stimulus length was about one second. Each recording was presented a total of twenty times, for a total of 320 trials (20 trials per image X 4 images per word X 4 words). The resulting data was organized by category (audio only, audiovisual congruent, audiovisual incongruent, and visual only).



**Methods**

 The LFP signal is recorded from 2.2 millimeter diameter extracellular electrodes. This signal is then low pass filtered at 300 Hertz to remove any spike train data. Hence, the LFP represents background electric field activity of a conglomeration of cells around the electrode. The LFPs are then averaged for each stimulus type (audio only, audiovisual congruent, audiovisual incongruent, visual only).

 The gamma band oscillations were examined more closely using a method called the Gamma Band Response (GBR). Power spectral analysis is performed on the LFP data using a multitaper frequency transformation on each trial using frequency bins of two Hertz from 30 to 150 Hertz and a time window of 10 milliseconds. A baseline level is computed by averaging the power spectra across frequencies and then taking the median across the pre-stimulus time, retaining different values for each trial. A test level is computed similarly at each time point in the series, and then these two groups are compared using the Wilcoxon rank-sum Z-test.

 Another method to analyze gamma band oscillations is broadband power. The LFP data is bandpass filtered from 30 to 150 Hertz, and the Hilbert transform is then taken. A standard z-score is taken of the logarithm of the Hilbert transform. This data is then smoothed by convolution. The z-score is taken again and then re-exponentiated to obtain the final result.

 These two methods have various advantages and disadvantages. The GBR is more straightforward to apply and easier to understand how it functions. In addition, the GBR shows more fine details; however these details can be attributed to mere noise. On the other hand, the broadband power is much smoother, and erases all of the fine details that the GBR preserves. Its method of implementation is much more complex and harder to justify. For more information on these methods, see the work of JP Lachaux and KJ Miller.

**Conclusion**

After comparing audiovisual congruent and audiovisual incongruent stimuli, we found that differences existed in the responses to these stimuli in the Superior Temporal Sulcus. These differences include differences in latency and magnitude of the GBR and the broadband power. These differences have yet to be quantized, and further work on analyzing the data continues. Further work will also involve using these methods to analyze data from other experiments, including a similar experiment involving the sound quality of recordings, and other similar variations.

**Acknowledgements**

 I would like to thank Dr. Michael Beauchamp, my faculty mentor on this project, for his help and support, and Sarah Baum, the graduate student with whom I worked extensively. I must thank Dr. Kai Miller, who was key in developing the broadband power method. I would also like to thank Dr. Steve Cox, who organized the NSF REU program in Theoretical and Computational Neuroscience. Finally, I would like to thank all of the people in the Beauchamp Lab, Dr. Siavash Pasalar, Eswen Fava, and Audrey Nath, for the assistance during this project. This work was supported by NSF REU award DMS-0755294, NSF Cognitive Neuroscience Initiative 0642532 and NIH R01 NS065395.