The neural implementation of contextual influences in speech perception

(623072 – Speech in context)

Summary report:
In everyday life, we perceive speech under various listening conditions and from a wide range of speakers. Although we have, most of the time, little problems understanding the words we hear, our perceptual system is actually performing a tremendously complex task. That is so because in spite of the intuition that understanding speech is simple, in real life, words are acoustically quite difficult to reliably distinguish. This is an important reason why automatic speech recognition has been such a hard problem to solve. The main reason for this difficulty is variance: The word "boot" spoken by person A may be acoustically quite dissimilar from the word "boot" spoken by person B, and may in fact be acoustically more similar to an instance of the word "boat" spoken by person B. Add to that the acoustic influences of background noise, and it becomes clear that perceiving speech is in fact a very complex problem. So why is it, then, that human speech perception is mostly effortless? Why don't we struggle with these ambiguities more often? We manage to solve this problem because our brains strongly rely on context. But how our brains manage to do this has been largely unknown.

The project "Speech in Context" relied on novel and very detailed measures of brain activity to investigate how our perceptual system uses context to reduce perceptual ambiguities. The main method that we used was electrocorticography (ECoG). In ECoG, an invasive procedure, electrodes are placed directly on the surface of the brain. These electrodes are placed as part of a procedure to treat severe cases of epilepsy or brain tumors. While patients were hospitalized and had electrode arrays implanted for treatment, we asked them whether they were willing to participate in fundamental research. The reason we asked these patients is because ECoG recordings provides high spatial (millimetre scale) and temporal (millisecond scale) resolution and is much less prone to

Figure 1: Example brain with electrode grid. All electrodes that measure brain activity from brain regions that re-spond to speech are color coded (red = low actovity; yellow = strong activity). Regions with small black dots do not respond to speech. The white dashed outline marks the auditory cortex.

artefacts of muscle activity (compared to, for instance, EEG) than other neuroimaging methods. The intracranial recording procedure is only performed under rare clinical conditions and therefore patients are relatively few in number. Nevertheless, research with ECoG has been starting to provide important new insights about the neural processing characteristics of speech perception.

The project Speech in Context has revolved around a number of subprojects. The main subproject investigated how listeners can tune in to the voice properties of different speakers by relying on context. In this experiment, while brain activity was being recorded, participants listened to speech sounds that were ambiguous between the sounds /u/ (the vowel in "boot") and /o/ (the vowel in "boat"). Importantly, these target sounds were preceded by speech from speaker that has a "dark" voice (a tall speaker) or a speaker with "light" voice (a short speaker). The results demonstrated that participants indicated that a single ambiguous /u-o/ sound was more similar to /o/ when it was preceded b speech from the speaker with the "light" voice, and more similar to /u/ when preceded by speech from the speaker with the "dark"
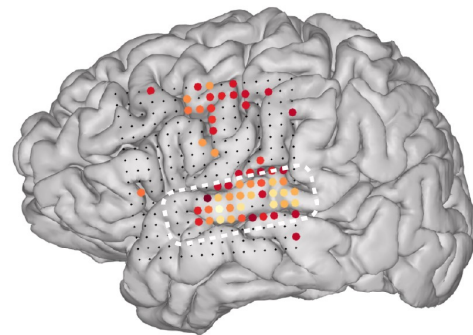
voice. The context sentence thus caused a behavioral shift in perception (listeners normalized the category boundary to the speaker). We then investigated how brain activity could cause this effect. We found parts of the auditory cortex that strongly respond to /u/ sounds, and other parts of the auditory cortex that strongly respond to /o/ sounds. Importantly, we observed that those brain regions responded in different ways depending on context as well, and in the same direction as observed in behavior. This demonstrates that the brain already tunes in to different speakers at very low levels of processing (i.e., already in auditory cortex).

A number of other, closely related projects have investigated other aspects of contextual influences. One project investigated how speaker information (as described above) affects speech perception involving different cues. Cantonese, for example, not only uses "vowel color" to distinguish vowels, but also the pitch of a speaker's voice (ie., in canotone the word ma has a different meaning depending on whether the pitch is falling or rising). In that project we compared normalization for vowels and for tone, and observed that they operate in different ways. A further project investigated to what extent listeners need to pay attention to speech in order to use context. This project demonstrated that contextual influences operate even when participants are highly distracted. In a further project we investigated how listeners can use lexical context to overcome the influences of background noise. For example, when you hear someone saying "automo#ile", where "#" indicates a cough in the background, then listener often report hearing the actual "b" even if it was completely missing in the signal. In this project we also used ECoG to demonstrate that the human auditory cortex fills-in the missing acoustic information at a very low level of processing.

To summarize, we have delivered a number of research projects, all of which show that listeners use acoustic context to resolve various problems when perceiving speech. The auditory cortex, a region that has previously been believed to be involved in only relatively basic acoustic processing, plays a critical role in such contextual influences, and it performs highly complex and context dependent operations. The research project has sparked a number of additional research questions which we look forward to pursuing.

Societal impact:
Humans are vastly better at recognizing speech under varying conditions than computers. Demonstrating what mechanisms the human the brain uses when integrating contextual information will provide important directions for future solutions in automatic speech recognition. Moreover, a better, basic understanding of how our brains manage to understand speech effortlessly will help in increasing our understanding of situations in which speech perception does become difficult. For example in cases of background noise, but also for people who suffer from conditions such as dyslexia.

Exploitable foreground and plans for exploitation:
The project was not aimed at creating exploitable foreground.

Corresponding author:
Matthias Sjerps
Email: m.j.sjerps@gmail.com