

Audiovisual Integration in Speech Processing

Rebecca Woodbury

HST.722: Brain Mechanisms for Hearing and Speech

Student-Selected Topic Proposal

10.31.2007

Importance of Topic and Significance to Speech and Hearing

Visual cues have a broad influence on perceived auditory stimuli, specifically when these visual cues are the lip movements associated with the auditory stimulus of speech. The perception of speech is improved by watching the speaker's lips as they speak, especially in noisy situations, and visual information can help listeners determine the location of sounds or speech (Calvert et al 1997, Macaluso et al 2004). In addition, visual cues can alter the perception of speech to such an extent that phonemes are perceived as different phonemes, as demonstrated in the McGurk effect (McGurk and Macdonald 1976). The effect of visual stimuli on auditory perception is generally thought to be subconscious, although it has been shown that subjects who are forced to divert their attention to other tasks (tactile) are less influenced by visual stimuli (Alsius et al 2007). It has been shown that multisensory information, such as the visual component of speech, can activate unisensory areas such as auditory areas of the brain as well as multisensory areas, even when the corresponding auditory information is removed (e.g., Macaluso et al 2004). The multisensory integration of visual and auditory cues in speech perception is important to understand, as the auditory system clearly does not function alone in understanding speech. In addition, understanding how auditory and visual information is combined in speech perception can help to better understand how both the auditory and language systems operate under broader conditions.

Relation of the Topic to Other Topics Discussed

Multisensory integration occurs in many ways in the brain, including one way that was discussed in the dorsal cochlear nucleus topic in class. In this topic, the multisensory integration of somatosensory, vestibular, and auditory inputs in the dorsal cochlear nucleus was discussed. In the dorsal cochlear nucleus, somatosensory and vestibular inputs are integrated with auditory pinna cues to localize sound sources. In a similar way, visual information is combined with auditory information in the perception of audiovisual speech to improve speech perception, to determine where auditory signals are coming from, and to alter the perception of auditory stimuli based on additional visual input.

Audiovisual integration in speech processing is also closely related to the cortical processing of language. While learning about the cortical processing of language, the centers of the brain that are involved in language processing were discussed as to their specific locations. These locations are the inferior frontal lobe (including Broca's area on the inferior frontal gyrus and the precentral gyrus), the inferior parietal lobe (including the supramarginal gyrus and the angular gyrus), and the temporal lobe (including Heschl's gyrus, the anterior and posterior superior temporal planes, the superior temporal gyrus, the middle temporal gyrus, the superior temporal sulcus, and the insula). We should expect that many of these areas would also be active in audiovisual speech processing. Indeed, the major areas activated by the perception of audiovisual speech include the superior temporal sulcus, the inferior frontal gyrus (Broca's area), the insula, Heschl's gyrus, and the supramarginal and angular gyri, all areas that are involved in language processing (Miller and D'Esposito 2005, Bernstein et al 2007, Campbell et al 2001, Calvert et al 1997). Many of these "language" areas are also activated during silent lip-reading in the absence of auditory input (Calvert et al 1997, Campbell et al 2001).

The subject of audiovisual integration will be briefly explored in the upcoming topic of neuroimaging correlates of human auditory behavior and multisensory integration. One paper for this topic explores the organization with the superior temporal sulcus (a region highly implicated in multisensory integration, including audiovisual integration) using high-resolution fMRI, by determining how the STS responded differentially to audio, visual, and audiovisual stimuli.

Synopsis of Current Knowledge and Key Issues in this Research Area

An early area of research in audiovisual integration was to determine the effects of observing the visual component of speech (lip-reading) as compared to observing similar movements of the face that were not speech-related (gurning). It was determined that observing the visual component of speech without auditory information activates regions of the brain associated with auditory or language function, whereas observing gurning movements does not activate these areas. In one study, lip-reading was shown to activate the primary auditory and the auditory association cortices, the angular gyrus

(associated with language), and the inferoposterior temporal lobe, whereas observing gurning did not activate these areas (Calvert et al 1997). In another similar study that compared the observations of lip-reading and gurning to watching the face at rest, lip-reading activated Broca's area, the superior temporal gyrus and sulcus, and activation extended into the right auditory cortex and into the right and left lateral prefrontal regions. Observing gurning produced less activation than lip-reading, with less activation in the superior temporal cortex (especially in the left hemisphere), and with no activation of Broca's area or the auditory cortex (Campbell et al 2001). The activation of auditory regions with visual cues alone suggests that integration of audiovisual cues occurs before speech sounds are categorized into distinct phonemes in the auditory association cortex, or in other words that the integration is perceptual rather than related to higher-level processing (Calvert et al 1997, Bernstein et al 2007).

Another important aspect of studying audiovisual integration is in exploring the effects of temporal and spatial differences between the auditory and visual cues. One such study (Macaluso et al 2004) presented subjects with audiovisual stimuli that were either synchronous or asynchronous and were presented from either the same location or from different locations in order to determine the effect of temporal and spatial synchrony and asynchrony on brain activation patterns. Using PET, it was determined that the superior temporal sulcus and the visual regions of the ventral occipital cortex were activated in all conditions when the audio and visual stimuli were synchronous, regardless of whether they came from the same location. In contrast, dorsolateral occipital regions were identified that only responded when the stimuli were produced at different locations. As audiovisual integration of speech generally occurs when the audio and visual stimuli are coming from the same speaker (except in the ventriloquist effect), the authors concluded that the superior temporal sulcus is responsible for the multisensory integration of audiovisual stimuli. This view of the superior temporal sulcus as a primary region for multisensory integration is one that is supported by many other studies (e.g., Campbell et al 2001, Miller and D'Esposito 2005).

From studies such as those mentioned previously, a general model has been developed for the pathways involved in audiovisual speech integration. One fMRI study compared the activations in synchronous and asynchronous audiovisual speech stimuli to

develop a model for audiovisual integration in speech processing (Miller and D'Esposito 2005). The authors concluded that the middle superior temporal sulcus is where audiovisual inputs are first combined. This pathway then progresses along the superior temporal sulcus and the superior temporal gyrus. If the visual input to the superior temporal sulcus matches or supports the auditory input, the superior temporal sulcus provides feedback to the auditory cortex, which strengthens the auditory signal. If the audiovisual stimuli are asynchronous, the superior temporal sulcus recruits the intraparietal sulcus to perform temporal transformations in order to achieve a match between the stimuli. In addition, Broca's area is activated if it is necessary to parse the observed speech into intelligible parts.

This model is supported by many studies that implicate the superior temporal sulcus as the major orchestrator and earliest location of audiovisual integration. However, this model has recently been disputed, notably by a study that measured event-related potentials (ERPs) using EEG to resolve the temporal dynamics of audiovisual processing (Bernstein et al 2007). As fMRI does not have the temporal resolution necessary to determine the fine time course of audiovisual processing, EEG was used to test the temporal aspects of the audiovisual processing model. The major findings of the EEG study were that audiovisual integration occurred first in the dorsolateral prefrontal and inferior frontal cortex, with major activations occurring slightly later in the supramarginal and angular gyri of the inferior parietal lobe and in the intraparietal sulcus. The EEG study showed that the superior temporal sulcus was involved and was activated during the processing of audiovisual information, but it was not the first area to be activated and it was not the primary orchestrator of audiovisual integration. Instead, the authors proposed that the major areas of audiovisual integration are the supramarginal and angular gyri of the inferior parietal lobe. Although different models have been suggested and supported by extensive research, the exact locations and mechanisms of audiovisual integration in speech processing are still widely debated.

Recent Technical Developments that Foster New Approaches to the Topic

One important technical development that promises to elucidate the mechanisms of audiovisual integration is the advancement of fMRI. Advances in fMRI over the past decade have allowed for imaging of more regions of the brain with increasingly higher spatial resolution, which allows for more precise identifications of involved brain regions and aids in determining the substructures within these regions.

Proposed Papers for Discussion

- Macaluso, E, et al. "Spatial and Temporal Factors During Processing of Audiovisual Speech: a PET Study." *NeuroImage* 21 (2004): 725-732.

- A PET study which compares activation areas using audio and visual signals that are synchronous or asynchronous and produced at either the same or different locations.

- Miller, Lee M., and Mark D'Esposito. "Perceptual Fusion and Stimulus Coincidence in the Cross-Modal Integration of Speech." *The Journal of Neuroscience* 25 (2005): 5884-5893.

- An fMRI study which uses synchronous and asynchronous audio/visual signals to differentiate between sensory comparison of AV stimuli and perception of unified AV signal and summarizes how AV integration occurs.

- Bernstein, Lynne E., et al. "Spatiotemporal Dynamics of Audiovisual Speech Processing." *NeuroImage* (2007), doi:10.1016/j.neuroimage.2007.08.035

- An EEG study that examined temporal dynamics of AV integration and argues against generally accepted model of AV integration pathway.

Works Cited

Alsius, Agnes, et al. "Attention to touch weakens audiovisual speech integration." *Exp Brain Res* 183 (2007): 399-404.

Beauchamp M.S., et al. "Unraveling multisensory integration: patchy organization within human STS multisensory cortex." *Nat. Neurosci.* 7(2004): 1190-1192.

Bernstein, Lynne E., et al. "Spatiotemporal Dynamics of Audiovisual Speech Processing." *NeuroImage* (2007), doi:10.1016/j.neuroimage.2007.08.035

Calvert, Gemma A., et al. "Activation of Auditory Cortex During Silent Lipreading." *Science* 276 (1997): 593-596.

Campbell, Ruth, et al. "Cortical Substrates for the Perception of Face Actions: an FMRI Study of the Specificity of Activation for Seen Speech and for Meaningless Lower-Face Acts (Gurning)." *Cognitive Brain Research* 12 (2001): 233-243.

Macaluso, E, et al. "Spatial and Temporal Factors During Processing of Audiovisual Speech: a PET Study." *NeuroImage* 21 (2004): 725-732.

McGurk, H, and J Macdonald. "Hearing Lips and Seeing Voices." *Nature* 264 (1976): 746-748.

Miller, Lee M., and Mark D'esposito. "Perceptual Fusion and Stimulus Coincidence in the Cross-Modal Integration of Speech." *The Journal of Neuroscience* 25 (2005): 5884-5893.