

There is more to speech perception than meets the ear!

Speech perception in face-to-face communication integrates linguistic auditory and visual information. Visual speech, more commonly known as lip-reading, refers to the ability humans have to perceive speech sound visually. For example, a salient visual feature which influences the perception of the phoneme /m/ is the closure of the lips which is clearly distinct from the movement associated with /v/ where the upper teeth come in contact with the lower lip.

The 1976 hallmark study by the cognitive psychologists McGurk & MacDonald, strongly suggests that the two sensory modalities involved in audio-visual (AV) perception are not simply juxtaposed but rather integrated. The McGurk illusion is often cited as proof positive of the effect of AV speech integration, as observed when the auditory and visual signals are incongruent. For example, the syllable /ba/ uttered over a face mouthing /ga/, is often perceived as /da/, a percept influenced by combined features of /ba/ and /ga/.

While evidence about the robustness of AV speech integration is plentiful, there is no agreement yet on how the fusion of the two modalities takes place. This article examines the account on AV speech integration of two main theories of speech perception, MTSP and FLMP. A good understanding of AV integration in speech perception may facilitate the teaching of new phonemes in a foreign language class. More importantly, it may inform new methods for treating hearing impaired patients who tend to rely more heavily on visual speech perception.

The basic claim of Liberman's (1985) MTSP or the revised Motor Theory of Speech Perception is that speech perception is mediated by phonetic gestures. Perceivers in the babbling stages of infancy develop abstract representations of speech gestures and map them onto their resulting sounds. Perceivers are therefore able to recover from the acoustic signal information about the vocal tract movements that leads them to the corresponding percept. Based on this model, the auditory and visual channels send out to the brain gestural information independent of modality, and integration of the linguistic information occurs in the early stages of perception prior to phonetic categorization (Ojanen, 2005).

On the other hand, Massaro's FLMP or the Fuzzy Logic Model of Perception (1987) claims that listeners receive information independently from each sensory source and weigh each modality's contribution to perception. First, sources of information or features are evaluated in terms of how closely they match prototypes in memory. Second, features are integrated and assigned a goodness-of-match value; third, the perceiver reaches a decision based on the alternative judged to be the closest match to the prototype. In this three-step behavioral model, phonemic features are transmitted via separate sensory pathways and integrated at a later stage after phonemes classification.

In summary, the question under investigation is whether the audio and visual modalities of speech perception are integrated pre-categorically in a modality independent module, or whether each modality is processed independently and categorized before getting integrated.

Perception of non-native sounds may shed light on the question. The Arabic emphatic /T/ for example, is characterized by a primary articulation similar to the regular /t/ and a secondary articulation involving the retraction of the tongue in the pharynx accompanied by lip rounding. One acoustic consequence is a deep sounding collocated vowel. The problem in MTSP is that the novice listener cannot accurately produce this new sound and has no gestural representation of its production.

FLMP on the other hand processes the visual cues –lip rounding in this case, and auditory cues such as vowel quality, independently weighing the contribution of each modality, integrating visual and auditory cues and evaluating them against stored prototypes. FLMP also runs into a problem: the perceiver has no stored prototype of the newly learned sound.

Perception of non-native sounds may have unveiled a weak point in both theories; neither fully explains how audio and visual information get integrated in the perception of new sounds. At this point, the issue of audio and visual integration remains unresolved.

## References:

- Liberman, A.M., Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition* 21, 1-36.
- Massaro, D. W. (1987). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. Hillsdale, NJ: Lawrence Erlbaum,.
- McGurk, H & MacDonald, J (1976). "Hearing lips and seeing voices." *Nature* 264 (5588), 746–748.
- Ojanen V, Mottonen R, Pekkola J, Jaaskelainen IP, Joensuu R, Autti T, Sams M. (2005). Processing of audiovisual speech in Broca's area. *NeuroImage*, 25:333–8.