Primacy of Multimodal Speech Perception
Lawrence D. Rosenblum

Aswathy Velutharambath

Institut für Maschinelle Sprachverarbeitung

12.05.2016

Outline

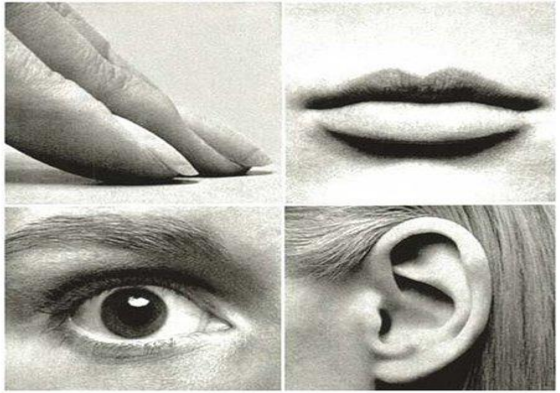
- Introduction
- Proposal
- Supporting Evidences
- Speculations
- Conclusion

Introduction

- **General Impression:** Speech perception is primarily an auditory function.
  - Easiest way to comprehend spoken language.
  - Based on technological artifacts(telephone and radios).
  - Other modalities are considered supplementary.

Speech as a multimodal function

- **Modalities of Speech Perception:**
  - Hearing
  - Lip reading
  - Gestures
  - Haptic apprehension



- Proposal : **Multimodal Speech is the primary mode of speech perception.**

Primacy of multimodal speech perception

- **Supporting evidences:**
  - The ubiquity and automaticity of multimodal speech
  - Extremely early speech integration
  - The neurophysiological primacy of multimodal speech
  - Modality – neutral speech information
- **Speculations :**
  - Multimodal basis of the evolution of spoken language

The Ubiquity and Automaticity of Multimodal Speech

- **Visual speech** : for enhancing speech degraded by
  - Cochlear implants for hearing impaired
  - Background noise
  - Heavy foreign accent
- For enhancing clear auditory speech while conveying complicated content
- Infant's language development

The Ubiquity and Automaticity of Multimodal Speech

- **McGurk Effect:**

  A perceptual phenomenon that demonstrates an interaction between hearing and vision in speech perception. The illusion occurs when the auditory component of one sound is paired with the visual component of another sound, leading to the perception of a third sound.

- Demo

Early Speech Integration

- **Integration of information from different sensory modalities.**
- Proposed stages at which integration occurs:
  - At the informational input (Green,1998; Rosenblum & Gordon,2011)
  - Before feature extraction (Summerfield, 1987)
  - After feature extraction (Massaro 1987)
  - After segment, or even word recognition (Berndtein et al 2004)

Early Speech Integration

- Multimodal speech is integrated at an early stage of the process, at least before phonetic categorization.
- **Summerfield (1987)**
  - Speech perception system takes in all auditorily and visually specified linguistic dimensions, integrates them and then performs phonetic categorization.
  - Visually perceived rate of articulation can influence auditory voice onset time (VOT) feature.
  - Support for the proposal that integration occurs pre-categorically.

Early Speech Integration

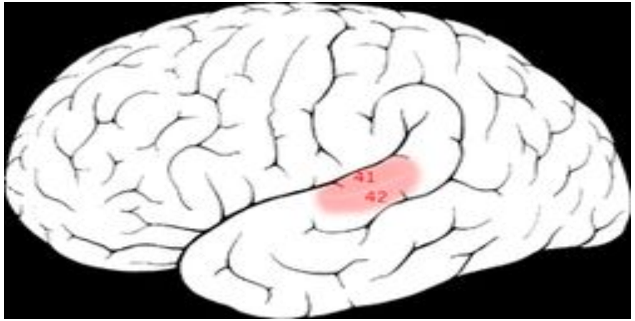
- **Green (1998)**
  - Cross-modal influences work at feature level.
  - Visually influenced co-articulatory information can affect perception of adjacent segments
  - Co-articulatory context established solely in one modality can influence segments in the other modality.
- Both theorists suggest a modality independent metric at the point of integration.
- True unimodal perception is rare.

Early Speech Integration

- Multimodal speech is integrated at an early stage of the process, at least before phonetic categorization.
- **Summerfield (1987)** – Speech perception system takes in all auditorily and visually specified linguistic dimensions, integrates them and then performs phonetic categorization.
- **Green (1998)** – Cross-modal influences work at feature level.
- Both theorists suggest a modality independent metric at the point of integration.
- True uni-modal perception is rare.

Neurophysiological Primacy

- ➢ Auditory Cortex - is the part of the temporal lobe that processes auditory information in humans and other vertebrates. It is a part of the auditory system, performing basic and higher functions in hearing.



- from Wikipedia

Neurophysiological Primacy

- Sams et al. (1991) showed that ***changes in visual speech information*** can change auditory cortex activity during audio visual integration.
- Callan et al. (2001) supported an early speech perception mechanism that is ***sensitive to multimodal speech.***
- FMRI research by Calvert and MacSweeny report evidence that ***silent lip reading*** can induce primary cortex activity similar to that of auditory speech.

Neurophysiological Primacy

➢ Evidences outside speech perception literature

- ❖ Numerous brain regions and neuronal sites are specifically tuned to multimodal input.
- ❖ Brain regions once thought to be sensitive to unimodal input are now known to be modulated to multimodal inputs.
- ❖ Research on neuroplasicity and neurodevelopment.

Modality-Neutral Speech Information

- Modality is invisible to perception function.
- **Summerfield –** Speech information is considered to be composed of higher-order, time-varying patters of energy(light, sound).
- All relevant speech regardless of the modality is defined by higher order gestural structure.
- “Cross-modal” integration occurs in and as a property of information itself.

Modality-Neutral Speech Information

- Example of modality neutral speech information (**Summerfield**)
  - Repeat the syllable /ma/ with a regular frequency.
  - Acoustic structure : influence overall amplitude and formant structure.
  - Optical structure : visible lip and jaw opening trajectories, structure light in a way specific to articulatory rate.
  - Higher order information for frequency of oscillation could be considered modality independent.

Modality-Neutral Speech Information

- Example of modality neutral speech information (**Vatikiotis-Bateson**)
  - 3D kinematic tracking of facial and interior articulatory movement (vocal tract).
  - The estimation of speech acoustics from facial kinematics was better than from internal vocal tract measures.
  - High correlation observed between visible facial kinematics and acoustics dimensions.
  - When applied to noise source, the parameters estimated from facial kinematics can produce intelligible auditory speech.

Modality-Neutral Speech Information

- **Informational similitude in audio and visual speech**
- Isolated, time - varying dimensions of signals can provide useful speech information.

Modality-Neutral Speech Information

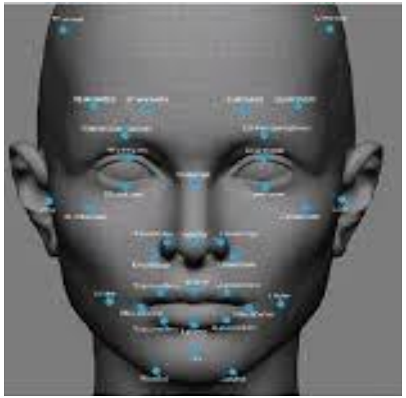
Auditory speech:

- ➢ Signals which do not involve the traditional cues of formants, transitions and noise bursts can still be understood as speech.
- ➢ This sine wave speech can be understood and transcribed by listeners.
- ➢ Portions of the signal that are least changing (like vowel nuclei) are less informative.
- ➢ For example , much of the vowel nucleus of a CVC syllable can be deleted without hindering vowel identification.

Modality-Neutral Speech Information

Visual speech:

- Point light technique



- Even though it does not contain any facial features, they provide visual

Modality-Neutral Speech Information

- Indexical dimension of speech - speaker specific properties of speech
- Auditory Speech
  - Sine wave speech retains speaker specific properties (Remez et al.)
  - This facilitates both speech and speaker identification.
- Visual Speech
  - Speaker's articulatory movements/gestures can facilitate visual speech perception.
  - Upright facial distortion can disrupt visual and audiovisual speech perception.

Visible speech and Evolution of spoken language

- If the primary mode of speech perception is multimodal, evidences should be found in evolution of language.
- **MacNeilage's frame/content theory of language evolution**
  - The "frame" of spoken language is constructed from components of ingestive mastication.
  - Assignment of ingestive gestures with communicative potential
  - Eg : teeth chattering, lip-smacks, tongue smacks etc in non human primates.

Visible speech and Evolution of spoken language

- ➢ Other theorists : Evolution of language occurred in a visual medium.
- ➢ Corballis (2002)
  - ❖ Continued use of gestural language by great apes.
  - ❖ Broca's area was well developed before vocal tracts were ready for speech.
  - ❖ Evolution of language : Manual → Facial → Vocal
  - ❖ Importance of vocal aspects : usefulness in dark and distances, free hands.

Conclusion

Argument : Primary mode of speech perception is multimodal.

- Evidences for ubiquity and automaticity as well as behavioral and neurophysiological findings.
- Conceptualization of modality-neutral information.
- Informational similitude across modalities.
- Speculation on multimodal influence on evolution of language.

Thanks