Print this Page

Presentation Abstract

Program#/Poster#:	672.19/JJ8
Title:	Audio-visual speech integration: How to bring lips and voice to work together
Location:	Halls B-H
Presentation Time:	Tuesday, Nov 16, 2010, 3:00 PM - 4:00 PM
Authors:	***M. PARENTEAU**[1], S. WOOD[1], J. ROUAT[1], S. MOLOTCHNIKOFF[2]; [1]Computer and Electrical Engin., Univ. De Sherbrooke, Sherbrooke, QC, Canada; [2]Life sciences, Univ. De Montréal, Montreal, QC, Canada

Abstract:

Multimodal integration is still not well understood. There are many mechanisms involved in this integration and it is difficult to identify its important parts. Modeling can help in understanding how this can be done. However, different models can achieve similar results in different ways. Last year, we presented a superadditive model to process multisensory stimuli, more specifically audio-visual speech. This year, we return to roots and explain what and how stimuli are integrated to bind lips to speech, and vice-versa. Our investigation is independent of the model. We based our works on a previous study by Hollich et al., (2005). They showed that 7 month-old infants will pay more attention when lips and speech are in synchrony than when they are not. The same study showed that infants do as well when the lips display is replaced by an oscilloscope display. These results may imply that the binding involves movement rather than actual shape recognition.

To show the plausibility of this hypothesis, we developed simple audio-visual stimuli preprocessing that accentuates the correlation between modalities and allows us to determine if they are bound. The acoustic stimulus (French male speech) is processed through a cochlear filter bank, and then envelopes are extracted for each cochlear channel. The visual stimulus (speaker's lips) is

decomposed as lips' inner width and height. The movement is then computed with instantaneous power variation of the width and height parameters.
We choose a stimulus from a given modality (e.g. auditory) and compute cross-correlation with each stimulus of the complementary modality (e.g. vision). Comparing the results, we identify which stimulus is bound to the chosen one. We show significant identification task results, with over 65% success rate. There is no complex processing, nor any postprocessing of the results. This makes our approach an interesting hypothesis to explain early audio-visual integration mechanism.
We will also discuss the properties of this approach and how it could be used in a more complete multisensory integration system, such as computational or biological models.

Disclosures: **M. Parenteau**, None; **S. Wood**, None; **J. Rouat**, None; **S. Molotchnikoff**, None.

Keyword(s): MULTISENSORY

SPEECH

MODELING

Support: NSERC

FQRNT

[Authors]. [Abstract Title]. Program No. XXX.XX. 2010 Neuroscience Meeting Planner. San Diego, CA: Society for Neuroscience, 2010. Online.