

Perception of Nonverbal Gestures of Prominence in Visual Speech Animation

Samer Al Moubayed and Jonas Beskow*
KTH Speech, Music and Hearing, 10044 Stockholm, Sweden



Figure 1: Gaze tracking results for ten subjects, showing gaze plot and heat map for two animation conditions (with and without gestures)

Keywords: Prominence, Gesture, Facial Animation, Gaze

1 Introduction

It has long been recognized that visual speech information is important for speech perception [McGurk and MacDonald 1976] [Summerfield 1992]. Recently there has been an increasing interest in the verbal and non-verbal interaction between the visual and the acoustic modalities from production and perception perspectives. One of the prosodic phenomena which attracts much focus is prominence. Prominence is defined as when a linguistic segment is made salient in its context.

Recent studies have focused on the relation between the visual modality (the face) and acoustic prominence. In [Beskow et al. 2006], results on Swedish showed that in all expressive production modes, words which are produced with a focal accent exhibit greater variation in the facial parameters movements (articulators, eyebrows, head, etc.). In addition to that, visualizing eyebrow movements and head nods on a talking head is found to be a powerful cue to enforce the perception of prominence. This paper explores, in addition to the known nonverbal effects of prominence gestures, whether they have any effects of speech comprehension and facial reading.

2 Audio-Visual Comprehension

The experimental setup in this study used a lip synced avatar. A speech intelligibility experiment was conducted, where speech quality was acoustically degraded and the fundamental frequency was removed from the signal. The speech was then presented to 12 subjects through a lip synchronized talking head performing head nodding and eyebrow raising gestures, which were synchronized with the auditorily prominent syllables. The experiment showed that perceiving visual prominence as gestures significantly increases speech intelligibility compared to when these gestures are randomly added to speech. For a full description, please see [Al Moubayed and Beskow 2009]. This result opens the possibility for talking heads to use visual correlates to prominence to support visual speech perception and aid the communication of prominence through the facial modality.

3 Gaze during Face Reading

The question this study targets is whether prominence facial gestures would increase the naturalness of the resulting avatar. A 15 minutes audio book (novel) was audio-visual presented to moderately hearing impaired subjects using a lip-synced avatar, and prominent syllables were automatically detected on the syllable level using a method explained in [Al Moubayed and Beskow 2010]. Using gaze tracking technology, the gaze tracking results, presented briefly in Figure 1, show that the facial reading pattern of subjects is more similar to this of reading a human face, in terms of fixation time on the eyes and mouth, and in terms of saccades in a triangular fashion between the eyes and the mouth. The questionnaires results show that the face with gestures is significantly more natural, helpful and understandable than this with no gestures. This study is described in more depth in [Al Moubayed et al. In press].

References

- AL MOUBAYED, S., AND BESKOW, J. 2009. Effects of visual prominence cues on speech intelligibility. In *International Conference on Auditory-Visual Speech Processing (AVSP)*.
- AL MOUBAYED, S., AND BESKOW, J. 2010. Prominence detection in swedish using syllable correlates. In *Interspeech'10*.
- AL MOUBAYED, S., BESKOW, J., AND GRANSTRM, B. In press. Auditory-visual prominence: From intelligibility to behavior. *Journal on Multimodal User Interfaces*.
- BESKOW, J., GRANSTRM, B., AND HOUSE, D. 2006. Visual correlates to prominence in several expressive modes. In *Proc of the Ninth International Conference on Spoken Language Processing*.
- MCGURK, H., AND MACDONALD, J. 1976. Hearing lips and seeing voices. *nature*.
- SUMMERFIELD, Q. 1992. Lipreading and audio-visual speech perception. *Philosophical Transactions: Biological Sciences* 335, 1273, 71–78.

*e-mail: {sameram,beskow}@kth.se