Virtual-reality based examination of audiovisual prosody in cochlear implant recipients

Khaled Abdellatif [1] Isa Samira Winter [1] Moritz Wächtler [1] Pascale Sandmann [2] Hartmut Meister [1]

[1] *Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital of Cologne and Jean-Uhrmacher-Institute for Clinical ENT-Research, University of Cologne, Geibelstraße 29-31, D-50931 Cologne, Germany,*

Email: khaled.abdellatif@uk-koeln.de

[2] *Department of Otolaryngology, Head and Neck Surgery, University of Oldenburg, Oldenburg, Germany*

Extended abstract

Prosody plays a vital role in verbal communication. It is important for the expression of emotions, but also carries information on sentence stress or the distinction between questions and statements. It is not only transmitted via the auditory system but also visually. This is partly due to the simple fact that changes in acoustic features are produced by changes in articulatory movements. As a result, stressed syllables are typically associated with larger mouth openings [1]. Additionally, head and facial movements accompany vocal expressions [2]. Prosody perception may be problematic for listeners with hearing impairment, especially in individuals with severe hearing loss using cochlear implants (CI) who are restricted in the use of acoustic prosody cues, particularly in terms of the voice fundamental frequency.

The present work describes procedural and technical aspects for investigating multi-modal features of linguistic prosody. As an example we show the animation of head and eyebrow movements during different sentence stress conditions. The application is based on two stages. The first one considers techniques of motion capture ("mocap") by the use of state-of-the-art face recognition algorithms. The second stage is the application of virtual humans with a realistic appearance. These virtual humans can be controlled by mocap data and – even more importantly – most of the underlying parameters can be modified arbitrarily. In accordance with the modification of acoustic cues this may give the opportunity to manipulate visual prosodic features and to determine the perceptual consequences.

In order to capture natural head and face movements, a dialogue situation between two interlocutors was realized [3]. The task of interlocutor 1 is to provoke natural prosodic expressions of interlocutor 2 during a realistic face-to-face communication. Movements of interlocutor 2 are video-recorded and supply the mocap data. We used phrases such as "Der Mann faehrt ein gelbes Auto." (English: "The man drives a yellow car.") with stress either on the subject ("Mann"), the adjective ("gelbes"), or the object ("Auto"). Such simple sentences of the structure subject, verb, adjective, object were previously applied in a prosody test-battery [4].

The mocap data were then used to operate a virtual human (Reallusion™, iClone V. 7), strength of the facial expression and the head movements can be modified arbitrarily. In order to compare the head and eyebrow movements between the video of the real talker and the virtual human, techniques of automated facial landmark localization were used. We analyzed the head movement – in particular the pitch angle of the vertical head rotation (nodding) – on the basis of a subset of six face landmarks by means of a Perspective-n-Point solving. In addition, we analyzed the eyebrow raise by determining the Euclidean distance between the landmarks at the medial palpebral commissure ("inner corner" of the eye) and the innermost point of the eyebrow of the right eye, as tracking was found to be most robust for these landmarks.

The technical evaluation of the proposed method shows that rigid head and eyebrow movements of the virtual human are qualitatively similar to those of the video-recorded real talker. The critical period of the stressed word is preceded by eyebrow raises and it is accompanied by a pronounced downward head rotation. In terms of the quantitative evaluation it can be shown that eyebrow movements of the virtual human are closely related to those of the real talker.

The poster will describe the technical evaluation of such an approach and discuss the potential application of the method in the framework of examining audiovisual prosody perception in CI users. Preliminary data derived from pilot studies are presented.

Acknowledgements

Supported by the Deutsche Forschungsgemeinschaft (ME 2751-4.1) to HM and (SA 3615/1-1 and SA3615/2-1) to PS.

References

[1] Scarborough, R., Keating, P., Mattys, S. L., Cho, T., & Alwan, A. (2009). Optical phonetics and visual perception of lexical and phrasal stress in English. Language and Speech, 52(2-3), 135-175.

[2] Krahmer, E., Swerts, M. (2009). Audiovisual prosody--introduction to the special issue. Lang Speech, 52(Pt 2-3),129-33.

[3] Cvejic, E., Kim, J., & Davis, C. (2010). Prosody off the top of the head: Prosodic contrasts can bediscriminated by head motion. Speech Communication, 52(6), 555-564.

[4] Meister, H., Landwehr, M., Pyschny, V., Walger, M., & Wedel, H. V. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. International Journalof Audiology, 48(1), 38-48.