Orofacial signals beyond sight: A study of expressive faces and whispered voices in German

Nasim Mahdinazhad Sardhaei[1], Marzena Zygis[2], Hamid Sharifzadeh[3], Annika Blietz[4]

[1] *Leibniz-Zentrum für Allgemeine Sprachwissenschaft (Germany)*, [2]*Humboldt Universität (Germany)*, [3]*Unitec Institute of Technology (New Zealand)*, [4]*University of Potsdam (Germany)*

A simple look at an interaction between a pair of speakers can reveal the fact that human communication involves a multimodal system in which gestures play an integral role. In this domain, one of the intriguing questions is about the nature of the relationship between gesture and speech. Closely related to this, there are two influential hypotheses. One possible conjecture is that there is a trade-off relation between gesture and speech in terms of the communicative load [1,2,3]. Another alternative account is a hand-in-hand hypothesis viewing the relation between gestures and speech in parallel rather than compensatory [4,5]. These two hypotheses largely depend on type of gesture as well as the communicative settings [6,7]. In this study, we focus on measuring the orofacial expressions including eyebrow movements, eye opening, and lip aperture in two different prosodic conditions, i.e., polar questions with rising intonation vs. statements with falling intonation. The varying intonation can enable us to find out whether and to what extent speech with varying prosody interacts with the oro-facial expressions. Taking (semi-)whispered speech" and "invisibility" of speakers as two communicative difficulties into account, we investigate what happens to speech and gesture when speakers (semi-)whisper and do not see each other.

To this end, we ran an experiment in which 15 native speakers of German were audio and video recorded while producing 20 pairs of statement and questions. The content of questions vs. statements was identical with the only difference in the punctuation mark. Each sentence was composed of 4 content words. The target word, which was the focus of our study, appeared at the sentence's final position. All the target words were bisyllabic with the stress falling on the initial syllable. The stressed syllables had CVC structure containing one of the bilabial stops /p/, /b/, /m/ followed by an unrounded vowel of /e/, /a/, /i/. The experiment took part in the interaction between a confederate and a participant. The confederate who was the same speaker during the whole experiment, generated either a question or a statement and the participants were supposed to respond the question by converting it into a statement or ask a question in response to the statement by altering its intonation (see appendix). The data were double checked with respect to intonation perceptually by a native speaker of German. The experiment consisted of four stimulus blocks linking two conditions, i.e., speech mode [*normal, semi-whispered, and whispered speech*] and visibility [*visible vs invisible mode*]. The orofacial expressions were measured using Openface2 facial landmark detector [8] by which each video was iterated, and 68 landmarks were mapped into the key regions of the face.

Based on the results of linear mixed effect models, the three-way interaction between *Speech Mode*(In)visibility*Sentence Type* for both the right eyebrow (t= -2.237, p<.05, see Figure 1) and left eyebrow (t=-2.045, p<.05, see Figure 2) was significant indicating that speakers raise their eyebrows the highest when they produce questions in whispered speech and when they are visible. For the lip aperture, a difference between three speech modes was observed with the largest lip aperture in whispered speech mode (t= 2.481, p<.05).The results also revealed a significant effect between *Speech Mode*(In)visibility* for opening of both eyes (t= -2.348, p < 0.01 for the left eye, and t= -3.350, p < .001 for the right eye) with the largest effect for the whispered speech mode in visible condition. Also, the interaction between *Speech Mode* Sentence Type* indicated more opened eyes when questions were produced in whispered speech mode (t= -1.967, p < 0.05 for the left eye, and t= -1.245, p < .05 for the right eye).

Overall, the results reveal more pronounced oro-facial expressions in a communicatively marked situation, i.e. when speakers whisper. Also, more pronounced orofacial gestures are produced when speakers see each other. We will discuss these findings in terms of trade-off and hand-in-hand hypothesis.
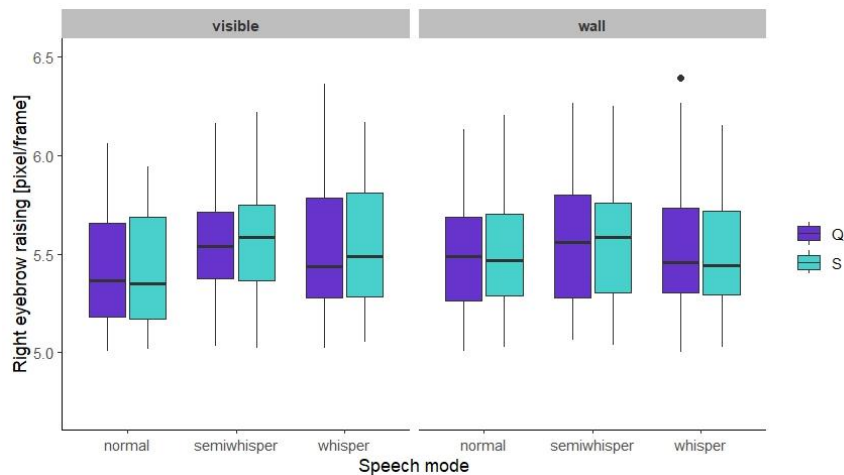
Fig. 1. Right eyebrow raising in the sentence-final word: interaction between **speech mode**, **(in)visibility**, and **sentence type**
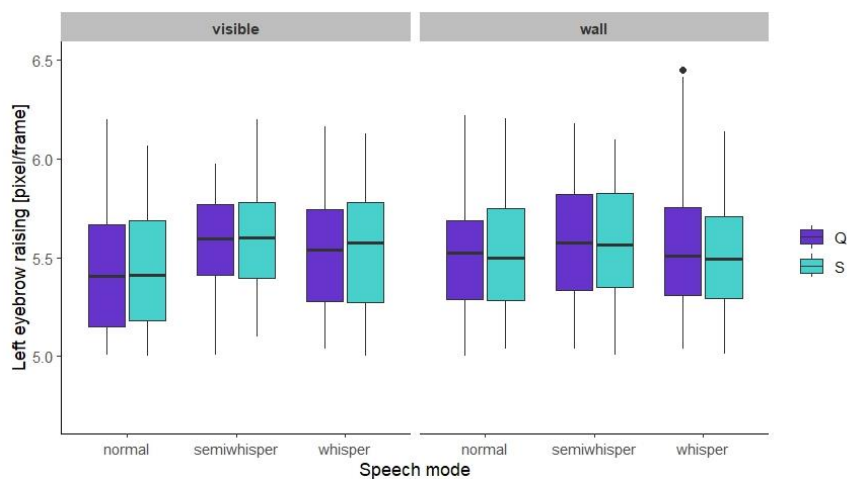

Fig. 2. Left eyebrow raising in the sentence-final word: interaction between **speech mode**, **(in)visibility**, and **sentence type.**

References
[1] Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science, 15*(6), 415-419.
[2] Melinger, A., & Levelt, W. J. (2004). Gesture and the communicative intention of the speaker. *Gesture, 4*(2), 119-141.
[3] De Ruiter, J. P. (2006). Can gesticulation help aphasic people speak, or rather, communicate?. *Advances in Speech Language Pathology, 8*(2), 124-127.
[4] Goldin-Meadow, S. (2009). How gesture promotes learning throughout childhood. *Child Development Perspectives, 3*(2), 106-111.
[5] So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science, 33*(1), 115-125.
[6] Bavelas, J. (1994). Gestures as part of speech: Methodological implications. *Research in Language and Social Interaction, 27*, 201–221.
[7] Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495–520.
[8] Baltrušaitis, T., Zadeh, A., Lim, Y. Ch., Morency L.-Ph. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. *IEEE International Conference on Automatic Face and Gesture Recognition.* https://par.nsf.gov/servlets/purl/10099460.