

ORO-FACIAL EXPRESSIONS AND ACOUSTIC CUES IN GERMAN QUESTIONS

Marzena Żygis^{1,2}, Nasim Sarhaei¹, Manfred Krifka^{1,2}

¹Leibniz-ZAS, ²Humboldt Universität zu Berlin
zygis@leibniz-zas.de, sardhaei@leibniz-zas.de, krifka@leibniz-zas.de

ABSTRACT

The present paper investigates acoustic cues and oro-facial expressions produced during the production of word-identical German sentences that represent eight different pragmatic uses (assertion, exclamation, and six varieties of questions). Our results based on 19 native speakers of German reveal that facial movements are sensitive to different sentence types and fine-grained pragmatic differences. Whereas exclamatives and assertions were produced with the highest raised eyebrows, incredulity, echo, and guessing questions showed the most intense eyebrow furrowing. In other question types (confirmation-seeking, relevance, and justification) the oro-facial expressions were less extreme. The data are also subject to huge inter-speaker variation.

Keywords: oro-facial expressions, acoustics, questions, German

1. INTRODUCTION

Questions and responses can be considered as fundamental building blocks in the structure of a conversation [1, 2, 3]. It has been frequently noticed that assertive and interrogative meanings are produced with specific auditory cues [4, 5, 6, 7]. On the other hand, audiovisual speech research has revealed the relative importance of facial cues with respect to auditory cues for signaling the prosodic functions. The role of visual information integrated with auditory information has been frequently addressed in studies on the production and perception of different sentence types. For instance, some studies have observed that eyebrow movements tend to co-occur with questions. To name a few, [8] demonstrated that pronounced eyebrow raising and head tilt enhanced the perception of echo questions in American English. [9] also reported that eyebrow raising and head movements tended to occur in French polar questions rather than assertions, and [10] showed that incredulity questions are characterized by some degree of eyebrow furrowing and eyelid closure in Dutch and Catalan. Similarly, lowering of eyebrows plus head yaw (turning left and right) was employed to mark the production of wh-questions in Brazilian

Portuguese [5]. As another investigated visual cue, [11] observed that Dutch and Catalan speakers directed their gaze at their interlocutors more frequently in information-seeking questions than in assertions. Analyzing face-to-face conversation, [12] noticed that eyebrow furrows were used as non-manual grammatical markers of content questions of requests such as ‘who?’, and eyebrow raises as markers of polar questions to serve as the restricted offers of repair such as ‘John Smith?’.

All these studies show that the production of specific facial gestures may vary not only according to the sentence type, but also according to subtle pragmatic meanings of the specific sentence, as well as language-specific differences. Building on these insights, the aim of this work is to compare the use of visual cues in the production of various sentence types, in particular questions, in German. In addition, we will also complement the study by acoustic investigations.

The present contribution is broader in its scope than previous works. Whereas they concentrated on just one or a few sentence types, we take into account six potentially distinct question types and compare them to assertions and exclamatives. Furthermore, previous studies concentrated on the movements of the eyebrows and the head, leaving aside the eye and lip opening. Hence, the aim of this work is to extend both the repertoire of questions under investigations and the spectrum of oro-facial expressions accompanying the questions.

2. EXPERIMENTAL EVIDENCE

2.1. Experimental design

To meet our research goal we conducted simultaneous video recordings of participants’ faces and their voices while they were producing sentences displayed on a computer screen, see details below. The video recordings were made by means of a camera (Canon EOS1300 D) and for the acoustic recordings we used a microphone (Sennheiser ME 64) to obtain acoustic data of high quality. The camera was positioned ca. one meter in front of a participant face, and the microphone was positioned

about 30 cm from their mouth. Video data were recorded with a sampling frequency of 25 Hz and acoustic data with a sampling rate of 44 100 Hz. The recordings took place in a sound-proof laboratory at Leibniz-ZAS in Berlin.

Our material consisted of six question types, as presented in (1), as well as assertions and exclamatives. All questions were in the form of so-called non-interrogative [13] or declarative questions [14] that do not differ in their morphosyntactic marking from assertions, thus enabling a direct comparison between the sentences. Please note that the American English corpus of [15] declarative questions are frequent in conversation; they make up 68% of the polar questions, which themselves make up 70% of the questions. They are frequent in German as well, see [16], [17] for their prosodic realization.

(1) Question types (all declarative questions)

- Echo questions
- Incredulity questions
- Confirmation seeking questions
- Guessing
- Relevance
- Justification

In addition, two other sentence types, i.e. assertion and exclamative, were included to create a contrast with the question types listed in (1).

Each sentence type complemented a different scenario, but the sentence itself was identical across scenarios. The only difference was the punctuation, i.e., questions (as in (1)) were presented with a question mark, the assertion with a period and the exclamative with an exclamation mark. We created six short SVO sentences that appeared in all scenarios, see (2).

(2) Sentences produced in distinct scenarios

- (i) *Maria fährt nach Algerien*
'Maria goes to Algeria'
- (ii) *Julia bekommt einen Labrador*
'Julia gets a Labrador'
- (iii) *Michaela hat einen Segelflugschein*
'Michaela has a glider pilot's license'
- (iv) *Thomas kauft einen Cadillac*
'Thomas buys a Cadillac'
- (v) *Martin bringt eine Rhabarbersuppe.*
'Martin brings a rhubarb soup'
- (vi) *Lydia spielt Bandoneon*
'Lydia plays bandoneon'

An example of a scenario eliciting an incredulity question is presented in (3).

(3) Scenario

Maria works in real estate and has a lot on her plate at any given time. She has decided to take a vacation soon. In the cafeteria, her colleague Steffi tells her office neighbor Peter that Maria wants to take some time off for vacation in Algeria. Peter is very surprised, because Maria usually goes on vacation to Sylt. He says:

- **“Maria is going to Algeria?”**

Steffi says: “Yes, I was also very surprised”.

Participants were asked to first read the full text, imagine the situation and then pronounce the sentence provided in bold in the way how it would be uttered in the provided scenario. A Latin square design was used so that each participant produced each sentence, in one of eight scenarios. There were no repetitions of identical sentence and scenario sets.

2.2. Participants

19 native speakers of German (12 female and 7 male, aged 22-43, mean age 26.4, sd 5.3) took part in the experiment. 17 participants were right- and two left-handed.

2.3. Measurements

To measure facial movements we used OpenFace 2.0 toolkit [18,19], an implementation of face recognition with deep neural networks, see also [20].

Figure 1 presents the positions of 68 markers mapped on participants' faces.

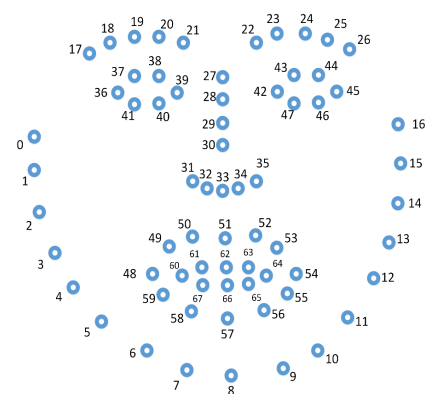


Figure 1: The markers chart used in Open Face ([19]).

Based on the mapping of the markers we measured distances between different markers corresponding to various oro-facial movements, as indicated in (4). In order to normalize the distances, we divided them by the so-called nose normalisation distance placed between markers 32 and 34. The normalisation

procedure provided a reference for distance measurements in (4) and mitigated against possible forward/back head movements.

(4) Measured distances:

(i)	right eyebrow rising	33 - 19
(ii)	left eyebrow rising	33 - 24
(iii)	eyebrow furrow	21 - 22
(iv)	lip opening	51 - 57
(v)	left eye opening	43 - 47
(vi)	right eye opening	38 - 40

In addition, we also performed an acoustic analysis in PRAAT (version 6.1.12 [21]), excerpting duration, RMS amplitude and max and mean pitch in ERB at the level of segments, syllables and words. In addition, each sentence was divided into seven intervals to normalize for different length. The present paper limits the presentation of acoustic results to mean pitch and to the domain of sentence.

2.4. Statistics

The statistical analysis of the data was conducted in the R Studio software (version 4.0.0, [22]) by using the lmer4 ([23]) and the emmeans packages ([24]).

Linear mixed effect models were employed for assessing the influence of SENTENCE TYPE [echo, incredulity, confirmation seeking, guessing, relevance, justification, assertion, echo, exclamative] on LEFT and RIGHT EYEBROW, EYEBROW FURROW, LIP OPENING, LEFT and RIGHT EYE OPENING. In addition, PARTICIPANT and WORD were included as random intercepts. SENTENCE TYPE was added as by-PARTICIPANT slope but due to the non-convergence had to be removed. The mean values over the intervals were fitted to the model. Finally, we corrected for multiple testing by using the Tukey correction from the emmeans package ([24]).

2.5. Results

Figure 2 presents results for mean F0 excerpted for seven equal intervals of each sentence type. It is evident that exclamations and assertions are produced with lowered F0 in the sentence-final position (target word) and all question sentences are produced with a continuously rising F0 in the same position. The statistical modelling shows no significant differences in F0 between exclamations and assertions ($t=0.809$, n.s.) but significant differences between exclamations and the six questions types: confirmation ($t=-5.23$, $p<.001$), echo ($t=-4.45$, $p<.001$), guessing ($t=-3.93$, $p<.0.01$), incredulity ($t=-5.73$, $p<.001$), justification

($t=-3.54$, $p<.05$), and relevance ($t=-3.86$, $p<.01$). Similarly, assertions were also produced with a significantly lower F0 in comparison to the six question types: confirmation ($t=-6.49$, $p<.001$), echo ($t=-5.62$, $p<.001$), guessing ($t=-5.06$, $p<.001$), incredulity ($t=-7.06$, $p<.001$), justification ($t=-4.56$, $p<.001$), and relevance ($t=-4.99$, $p<.001$). All comparisons between different question types (confirmation, echo, guessing, incredulity, justification, relevance) turned out to be not significant.

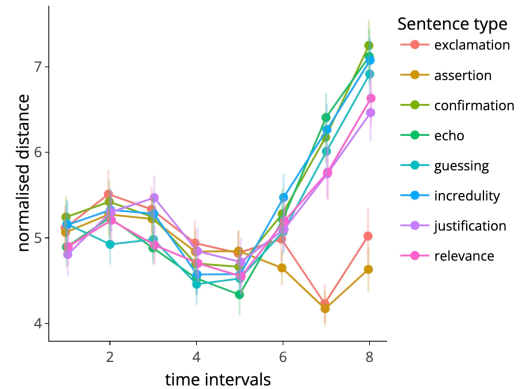


Figure 2: Mean F0 across different sentence types

Figure 3 presents the results obtained for the left and right eyebrow position during the entire sentence across all sentence types.

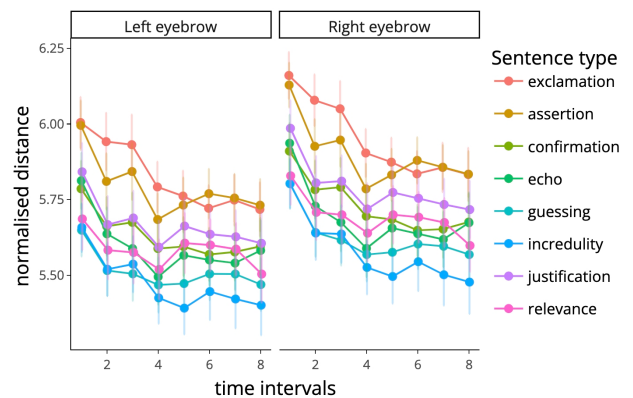


Figure 3: Left and right eyebrow positions across different sentence types

It appears that although the movement of the eyebrows shows a similar pattern across all sentence types, the right eyebrow was generally higher in comparison to the left eyebrow. Furthermore, exclamations and assertions showed the highest positions and incredulity questions the lowest position of the eyebrows. Statistical modelling reveals that several question types were produced with a significantly lower left and right eyebrow in comparison to exclamations and assertions. In fact, in the case of exclamations all comparisons referring to

the right eye-brow, besides the one with justification ($t=2.39$, n.s.), were significant: confirmation ($t=3.16$, $p<.05$), echo ($t=3.80$, $p<.01$), guessing ($t=4.20$, $p<.001$), incredulity ($t=5.35$, $p<.001$), and relevance ($t=3.62$, $p<.01$). As for assertions, they were produced with a significantly higher right eye-brow as compared to echo ($t=3.47$, $p<.05$), guessing ($t=3.91$, $p<.01$), incredulity ($t=5.11$, $p<.001$), and relevance ($t=3.29$, $p<.05$). Similar effects were found for the left-eyebrow. The difference between exclamations and assertions was neither significant for the left eyebrow ($t=0.042$, n.s.) nor for the right eyebrow ($t=0.491$, n.s.).

Furthermore, as shown in Figure 4 the greatest distance between the eyebrows, i.e., the least eyebrow furrowing, is found in exclamations and assertions and the most intense eyebrow furrowing is produced in echo, guessing and incredulity. The differences are significant: exclamation vs. echo ($t=3.74$, $p<.01$), guessing ($t=3.78$, $p<.01$), incredulity ($t=4.61$, $p<.001$); assertion vs. echo ($t=3.45$, $p<.05$), guessing ($t=3.52$, $p<.05$), incredulity ($t=4.38$, $p<.001$); The differences between exclamations and assertions ($t=0.44$, n.s) and other questions types remain not significant.

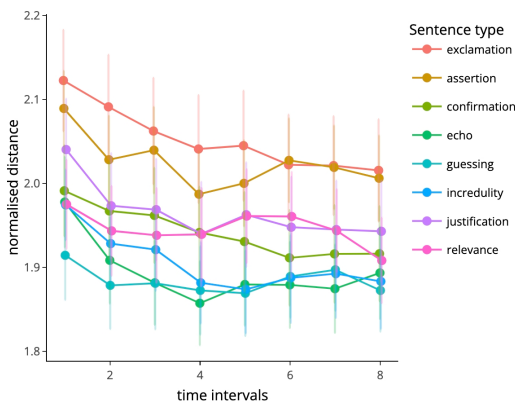


Figure 4: Eyebrow furrow across different sentence types

Our results also show that the exclamations were produced with the highest lip opening, especially in the initial and middle part of the sentence, but the differences turned out to be not significant.

Finally, regarding the left and right eye opening, it appears that it was larger at the beginning and especially at the end of sentences with exclamations and assertions showing again more eye opening, see Figure 5. However, the only significant difference was found for the left eye opening between exclamation and echo question being higher for the former ($t=3.13$, $p<.05$). No significant differences were found for the right eye opening.

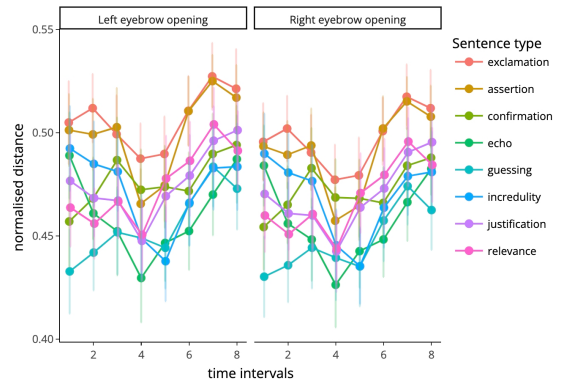


Figure 5: Left and right eye opening across different sentence types

3. DISCUSSION AND CONCLUSIONS

The present study, although limited in its scope and restricted to sentence domains, shows that oro-facial movements are sensitive to different sentence types. Considering the fact that at least some of the questions showed a very similar F0 movement, it appears that oro-facial expressions may deliver important cues to the interpretation of sentences in face-to-face communication. Further research, including perceptual evidence, should (dis)confirm this hypothesis.

Our study reveals that exclamations are produced with more raised eyebrows, (and more opening of the left eye) in comparison to other sentence types. Assertions are produced in a similar manner but with slightly less pronounced gestures. Also, for these two types of sentences F0 is lowered (at the sentence-final position), which suggests a compensatory effect between acoustic cues and oro-facial expressions, (see [25]), a point that requires further investigation at the syllable level. Furthermore, the most intense eyebrow furrowing is produced in echo, guessing and incredulity. These sentence types are also produced with the lowest eyebrow positions. Other sentence types are produced with oro-facial expressions that are generally less extreme. It should also be pointed out that there was a huge inter-speaker variation, as indicated by large confidence intervals.

It is also worth pointing out that the eyebrows did not move in the same manner. The right eyebrow was more raised than the left eyebrow, a result also previously found for questions in German [25]. However, [26] found for French a greater variation in movement magnitude for the *left* eyebrow, a result that the authors tentatively attributed to different involvement of brain hemispheres in the communication process. Another explanation could refer to intercultural differences where questions are marked either by right or left eyebrow raising, a point that also requires further investigation.

Acknowledgments

Research for this paper was done in the context of the ERC Advanced Grant SPAGAD: Speech Acts in Grammar and Discourse, receiving funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 787929. We would like to thank all participants for taking part in our experiment.

7. REFERENCES

- [1] Farkas, D. F., Bruce, K.B. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27, 81-118.
- [2] Sendra, V. C., Kaland, C., Swerts, M., Prieto, P. 2013. Perceiving incredulity: The role of intonation and facial gestures. *Journal of Pragmatics* 47, 1-13.
- [3] Krifka, M. 2022. Adjacency pairs in common ground update: Assertions, questions, greetings, offers, commands. *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue*, 94-105.
- [4] Moraes, J. 1998. Intonation in Brazilian Portuguese. In: Hirst, D., Di Cristo, A. (eds.), *Intonational Systems: a survey of twenty languages*. Cambridge, MIT Press.
- [5] Moraes, J. 2008. The pitch accents in Brazilian Portuguese: analysis by synthesis. *Proc. of the 4th International Conference on Speech Prosody Campinas*, 389-397.
- [6] Miranda, L. S., Moraes, J. A., Rilliard, A. 2019. Audiovisual perception of wh-questions and wh-exclamations in Brazilian Portuguese. *Proc. of the 19th International Congress of Phonetic Melbourne*, 2941-2945.
- [7] Enfield, N.J., Stivers, T., Levinson, S.-C. 2010. Question-response sequences in conversation across ten languages: An introduction. *Journal of Pragmatics* 42, 2615-2619.
- [8] Srinivasan, R. J., Massaro, D. W. 2003. Perceiving from the face and voice: distinguishing statements from echoic questions in English. *Language and Speech* 46, 1-22.
- [9] Torreira, F., Valtersson, E. 2015. Phonetic and Visual cues to questionhood in French conversation *Phonetica* 72, 20-42.
- [10] Crespo-Sendra, V., Vanrell, M. D. M., Prieto, P. 2010. Information-seeking questions and incredulity questions: gradient or categorical contrast? *Proc. of Speech Prosody 2010 Chicago*.
- [11] Borràs-Comes, J., Kaland, C., Prieto, P., Swerts, M. 2014. Audiovisual correlates of interrogativity: A comparative analysis of Catalan and Dutch. *Journal of Nonverbal Behavior* 38, 53-66.
- [12] Hömke, P., Levinson, S. C., Holler, J. 2022. Eyebrow movements as signals of communicative problems in human face-to-face interaction. <https://doi.org/10.31234/osf.io/3jnmt>
- [13] Bartels, C. 1997. *Towards a compositional interpretation of English question and statement intonation*. Doctoral dissertation, University of Massachusetts at Amherst.
- [14] Gunlogson, Ch. 2002. Declarative questions. *Proceedings of SALT XII*. Ithaca, NY: Cornell University, 124-134.
- [15] Stivers, T. 2010. An overview of the question-response system in American English conversation. *Journal of Pragmatics* 42, 2772-2781.
- [16] Niebuhr, O., Bergherr, S. Huth, C. Lill, J. Neuschulz 2010. Intonationsfragen hinterfragt. Die Vielschichtigkeit der prosodischen Unterschiede zwischen Aussage- und Fragesätzen mit deklarativer Syntax. *Zeitschrift für Dialektologie und Linguistik* 77, 304-346.
- [17] Petrone, C., Niebuhr, O. 2014. On the intonation of German intonation questions: the role of the prenuclear region. *Language and Speech* 57, 108-146.
- [18] Baltrušaitis, T., Robinson, P., Morency, L.-P. 2016. *IEEE Winter Conference on Applications of Computer Vision (WACV)*.
- [19] 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>
- [20] Schroff, F., Kalenichenko, D., Philbin, J. 2015. FaceNet: A Unified Embedding for Face Recognition and Clustering. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 815-823. doi: 10.1109/CVPR.2015.7298682.
- [21] Boersma, P., Weenink, D. 2020. Praat: Doing phonetics by computer (version 6.1.12) [computer program], <http://www.praat.org/>
- [22] RStudio Team 2020. RStudio: Integrated Development for R (version 4.0.0) (RStudio, Boston).
- [23] Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J. 2020. Package 'lme4' (version 1), <https://cran.r-project.org/web/packages/lme4/lme4.pdf>
- [24] Lenth, R. V. 2019. Package: 'emmeans: Estimated marginal means, aka least-squares means', <https://cran.r-project.org/web/packages/emmeans/index.html>
- [25] Žygis, M., Fuchs S. 2023. Communicative constraints affect oro-facial gestures and acoustics: Whispered vs normal speech. *Journal of the Acoustical Society of America*. Paper number 07819.
- [26] Cavé, Ch., Guaitella, I., Bertrand, R., Santi, S., Harlay, F., Espesser, R. 1996. About the relationship between eyebrow movements and F0 variations. *Proceeding of Fourth International Conference on Spoken Language Processing*, 2175-2178.