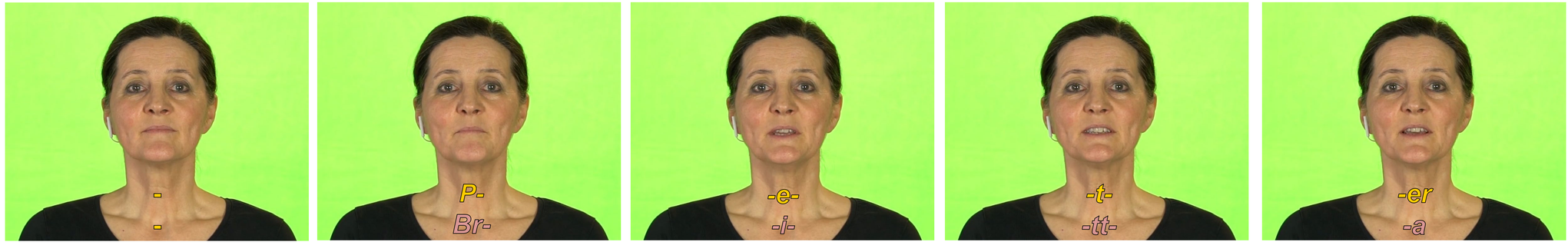
Word error and confusion patterns in an audiovisual German matrix sentence test (OLSA)

Gerard LLORACH and Volker HOHMANN
Hörzentrum Oldenburg; Cluster of Excellence Hearing4all; Medizinische Physik, Universität Oldenburg, Germany



Intoruction

- Matrix-sentence tests are an established and validated tool to evaluate speech reception thresholds and intelligibility (Kollmeier et al. 2015).
- They use non-predictable sentences, which are combinations of 5 words (name-verb-number-adjective-object, e.g. Peter kauft drei teure Dosen).
- In (audio-only) matrix sentence tests, speech intelligibility is relatively homogeneous among words. This is done to improve the precision of the test when measuring speech reception thresholds (Kollmeier 1990).

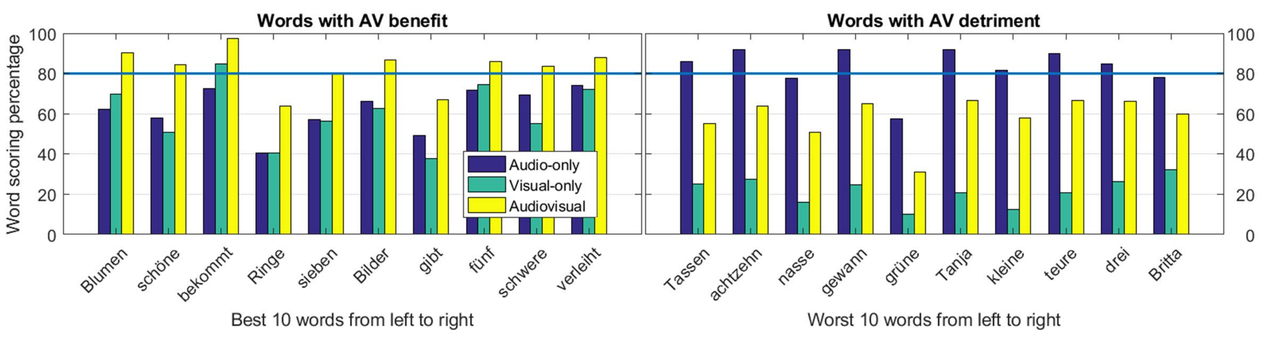
Motivation

- Speech is multimodal: lip movements and facial expressions improve speech perception (MacLeod and Summerfield 1987).
- Adding visuals to matrix sentence tests is a step towards ecological validity.

Research Questions

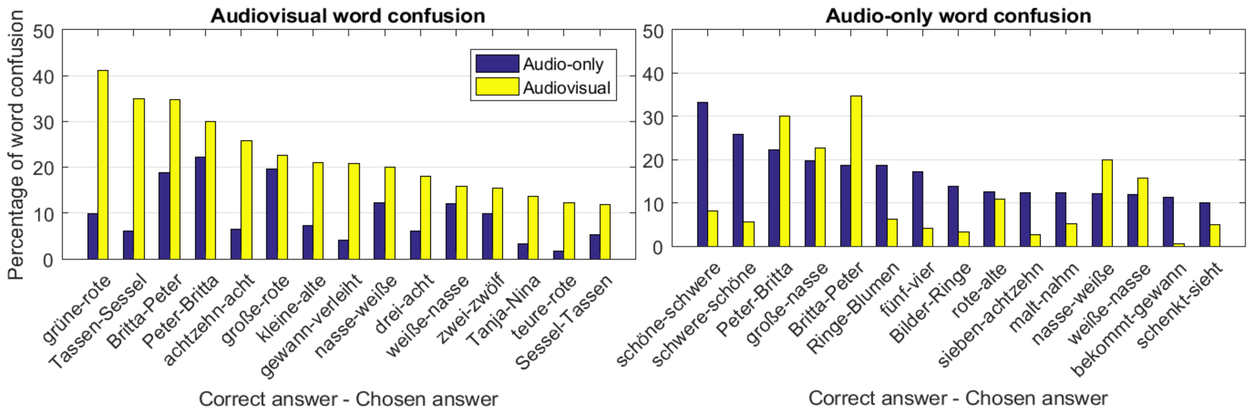
- What happens with the homogeneity of the word's intelligibility when you are able to see the face of the speaker and lip-read?
- Which words will be understood better and vice versa when visual cues are present?

Results – Audiovisual Benefit and Detriment



- The intelligibility of the words changes when adding visual cues, either making some words more intelligible in respect to the others or vice versa.
- Lip-reading (Visual-only) is related to the audiovisual benefit and detriment i.e. higher green bars in AV benefit than in AV detriment.

Results – Confusion Patterns

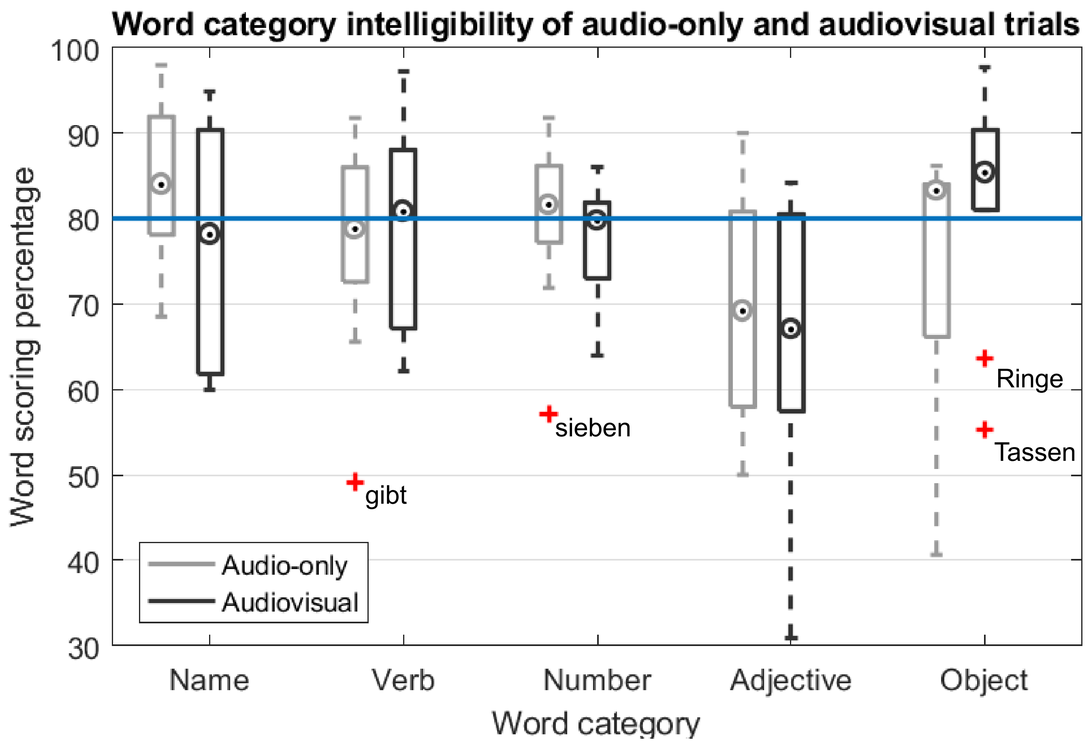


- Some pairs of words are acoustically similar (schöne-schwere), visually similar (grüne-rote) or audiovisually similar (Peter-Britta).

Method

- OLSA matrix sentence test:
  - Lists of 20 sentences in audio-only, audiovisual and visual-only modalities with test-specific noise. Closed response type (10 possible answers per word).
  - Adaptive procedure to achieve 80% Speech Reception Threshold (SRT) i.e. 80% of each sentence should be understood.
- 28 NH participants (14 female, mean age 24.9 years - range 20-29)
- 2-6 audiovisual, 2 audio-only and 2 visual-only lists per subject.

Results – Homogeneity



- Similar results for audio-only and audiovisual conditions. Adjectives are less intelligible in general.
- Only the answers of the last 10 sentences of each list were considered, as they are closer to the targeted SRT. There are around 61 and 136 appearances per word for audio-only and for audiovisual conditions respectively among all trials and subjects.

Conclusion

- The homogeneity between group of words (name-verb-number-adjective-object) in audiovisual trials is somewhat similar to audio-only trials.
- The relative intelligibility score of each word can change when adding visual cues.
- Confusion patterns can change when adding visual cues, demonstrating that acoustic and visual speech provide different information (phonemes vs visemes).
- Visual-only recognition may reach more than 50% SRT, thus testing 50% SRT in audiovisual is not recommended.

Acknowledgements

This work has received funding from the EU's H2020 research and innovation programme under the MSCA GA 675324 (ENRICH) and the Deutsche Forschungsgemeinschaft (DFG, Cluster of Excellence EXC 1077/1 "Hearing4all" and SFB1330 Project B1 and C4).

References

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M.A., Uslar, V., Brand, T. and Wagener, K.C., 2015. The multilingual matrix test: Principles, applications, and comparison across languages: A review. International Journal of Audiology, 54(sup2), pp.3-16.

Kollmeier B. 1990. Messmethodik, Modellierung, und Verbesserung der Verst ä ndlichkeit von Sprache (in German). (Methodology, modeling, and improvement of speech intelligibility measurements). Habilitation thesis. Göttingen: University of Göttingen.

MacLeod, A. and Summerfield, Q., 1987. Quantifying the contribution of vision to speech perception in noise. British journal of audiology, 21(2), pp.131-141.