

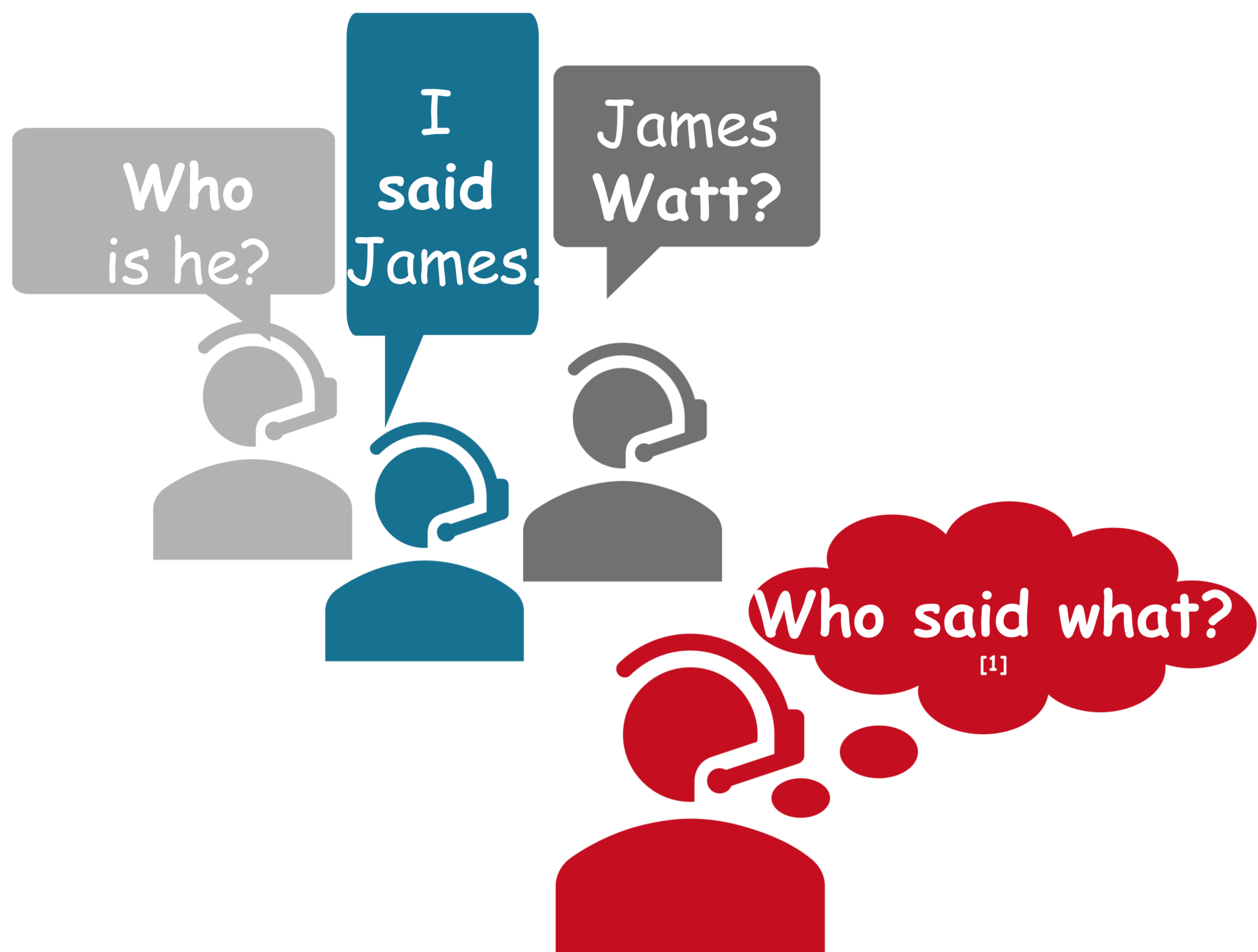
Spatial Audio and Speaker Recognition

Can Binaural Synthesis improve audioconferences?

Lars-Erik Riechert*, Laura Fernández Gallardo* und Dennis Guse*

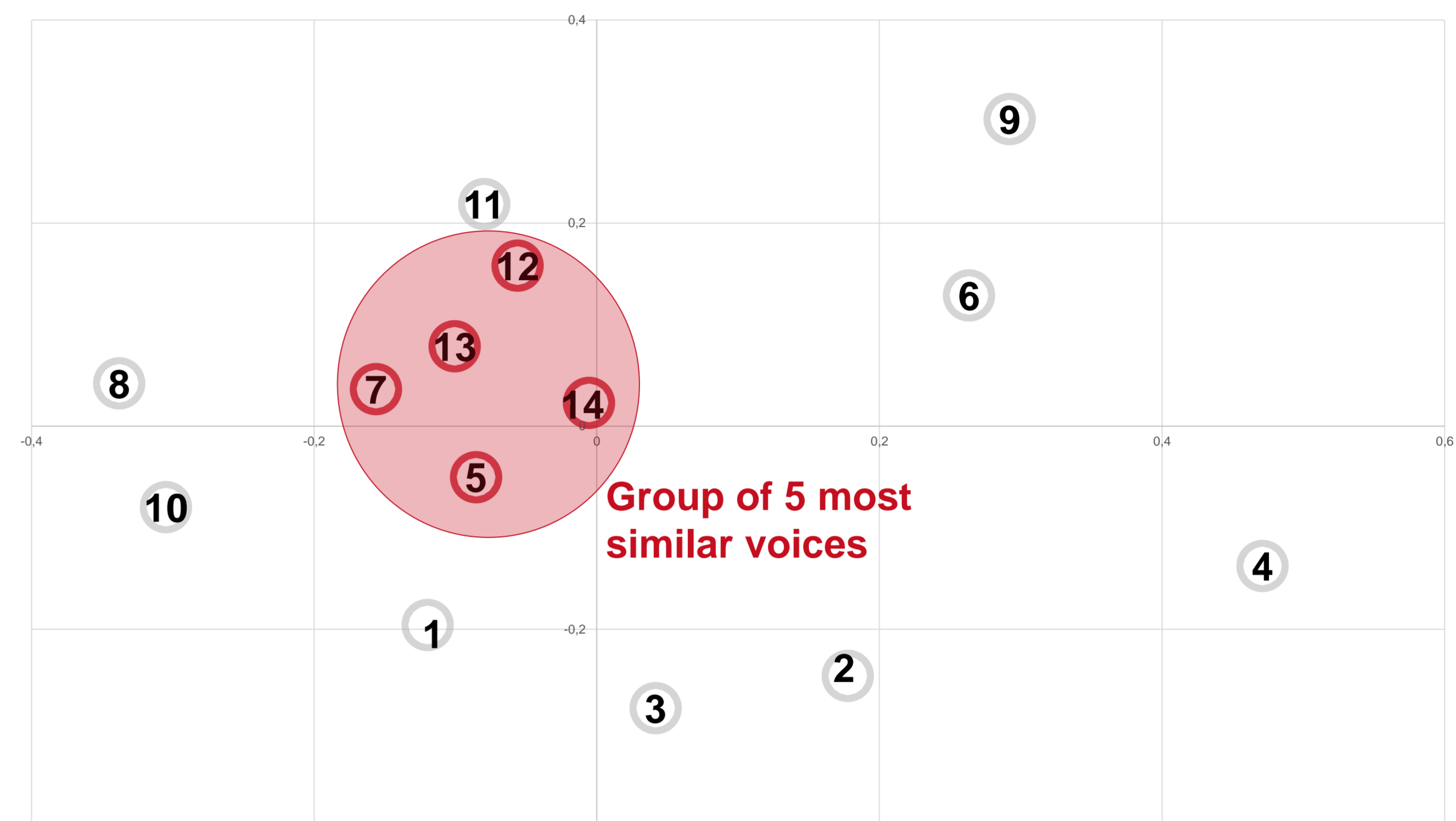
(*)TU Berlin, Quality and Usability Lab, Ernst-Reuter-Platz 7, 10587 Berlin, Germany

THE PROBLEM



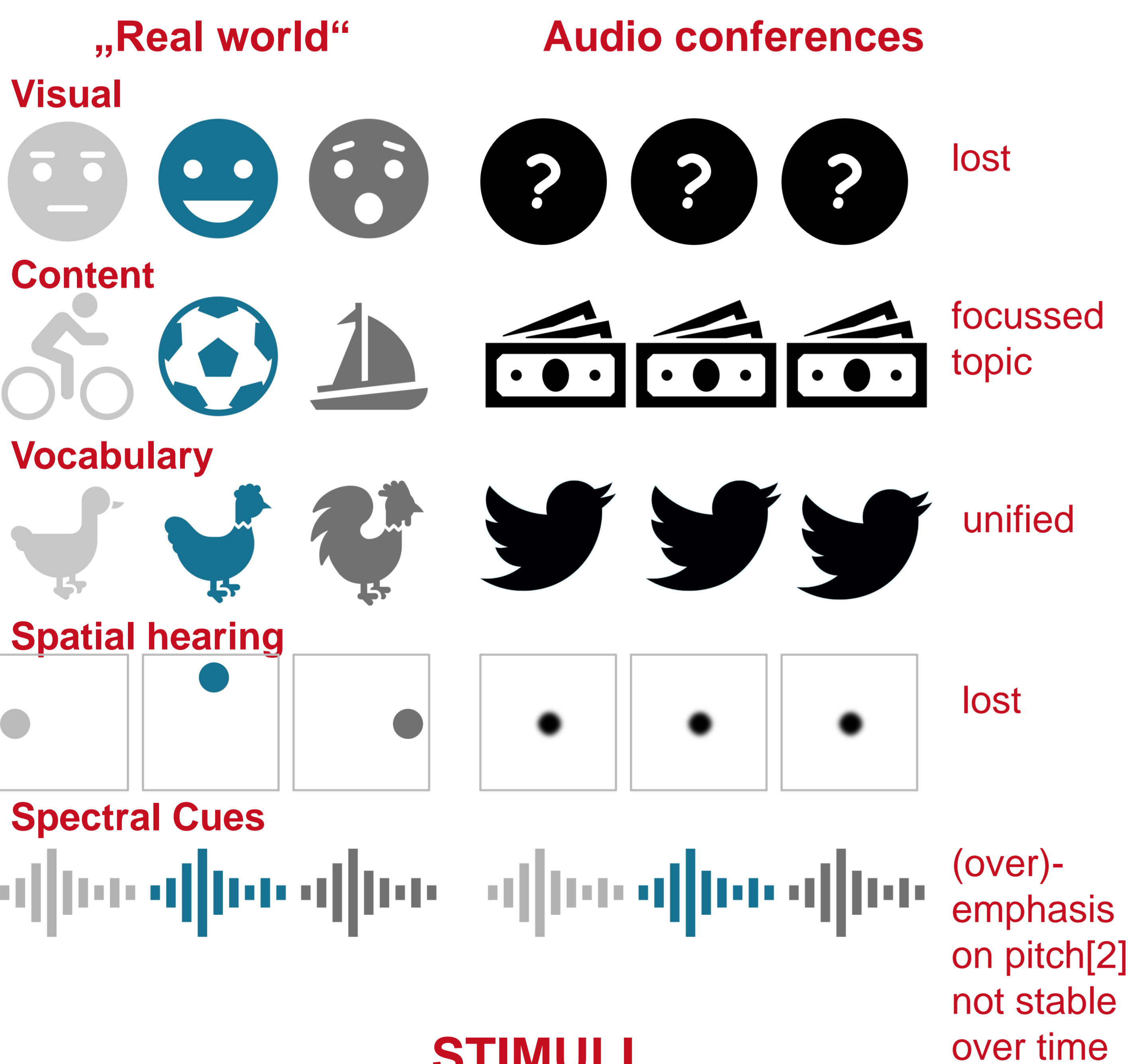
SIMILAR VOICES

An experiment with 10 participants was conducted to identify a group of 5 similar voices out of 14. To do so each participant had to perform 105 pairwise comparisons. The graphic shows the similarity values condensed to a 2D space using multidimensional scaling.



CLUES FOR SPEAKER RECOGNITION

Speaker recognition works on a multifactorial basis. Some of these factors get lost during audioconferences. Additionally a professional context leads to a unified language and speaking style. The remaining clues are often not time stable, e.g. the pitch changes due to intonation.

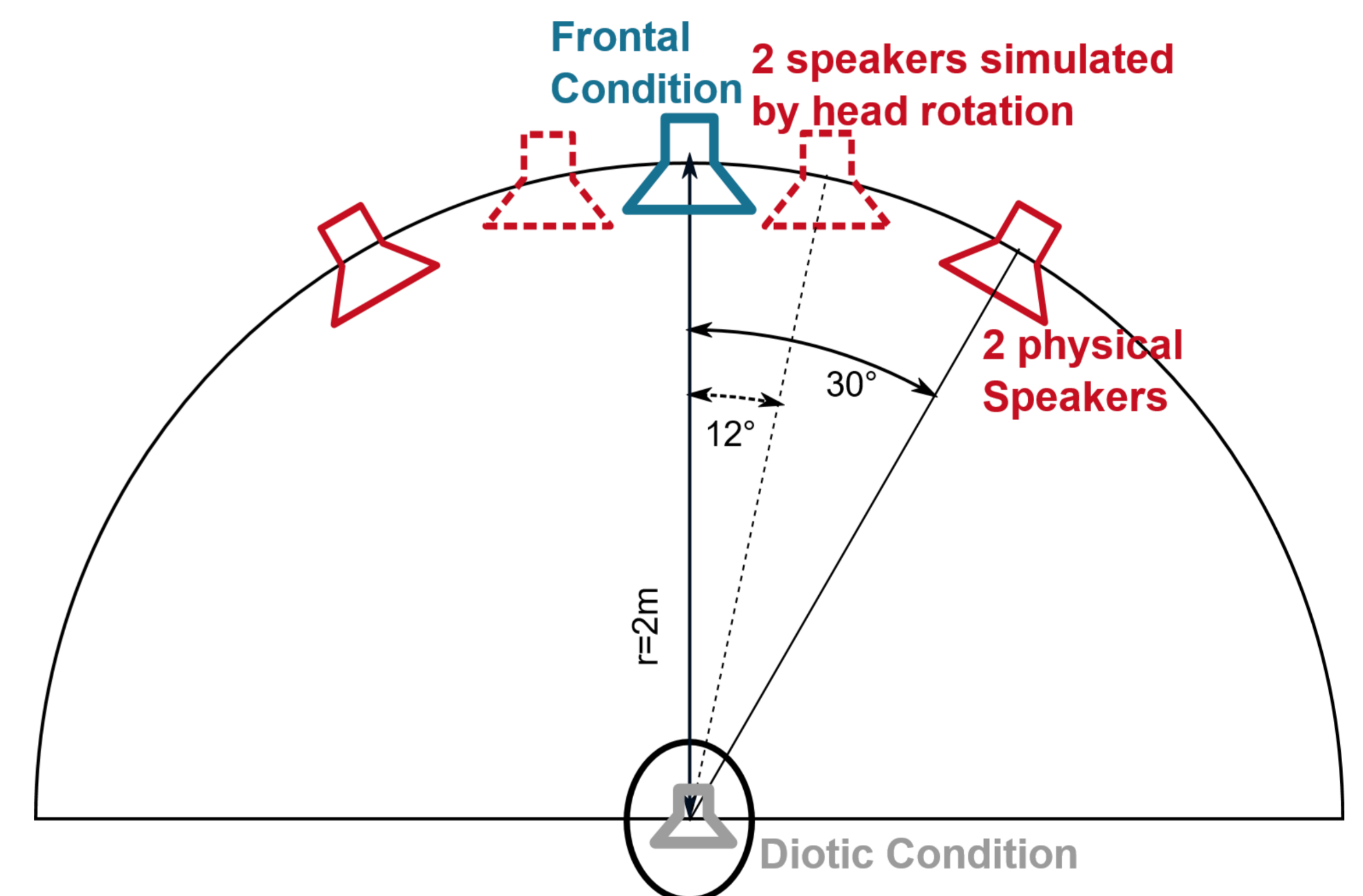


STIMULI

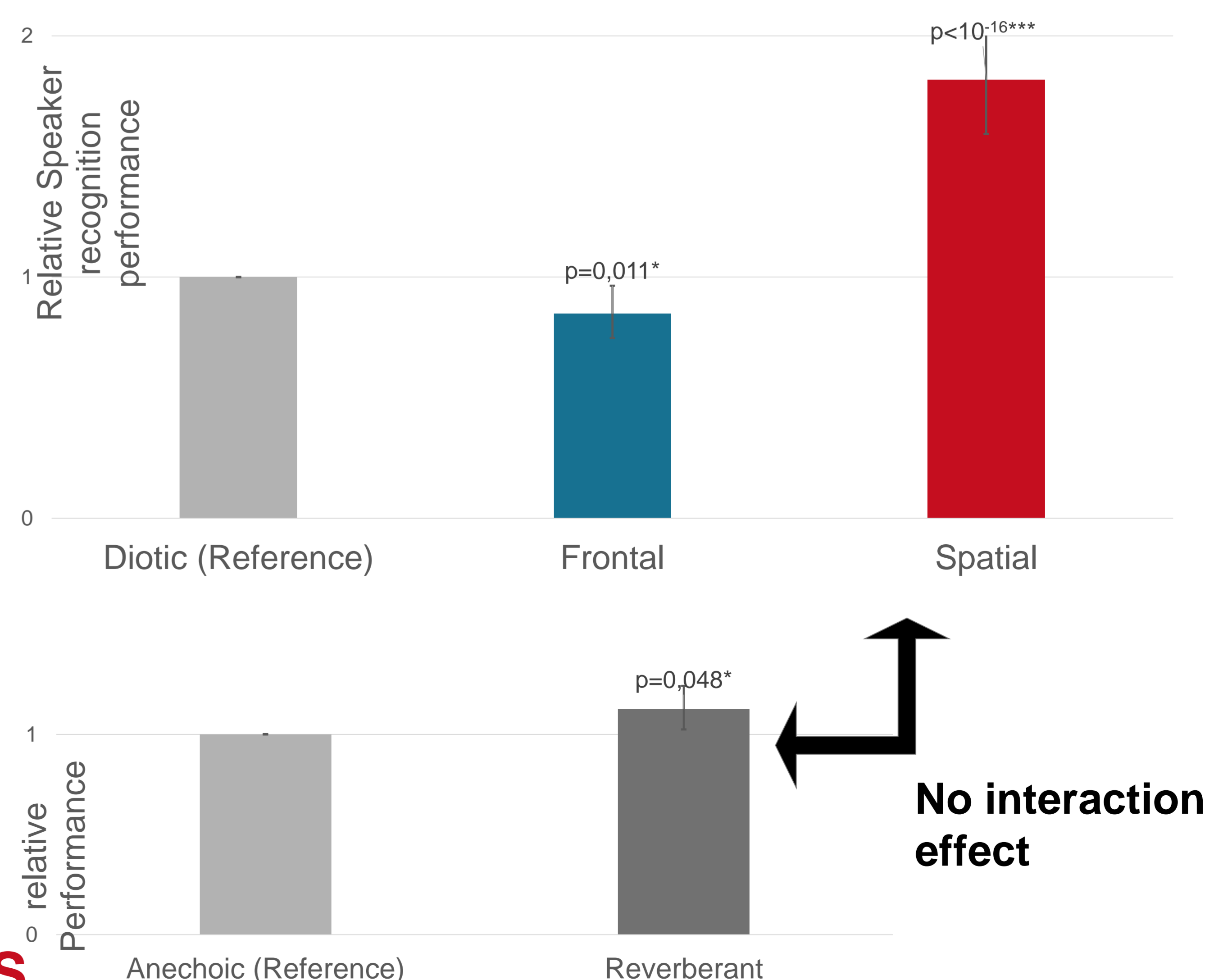
Recordings have been made with 15 female and 15 male speakers. For each talker an identical set of 25 sentences and 52 words has been recorded. From these recordings the stimuli have been prepared semiautomatically. The female voices have been used during the experiment, while the male have been used for training.

EXPERIMENTAL SETUP

Stimuli have been presented to 40 participants, either diotically or using binaural synthesis. The latter featured either 4 distinct speaker positions or a single frontal direction. As well anechoic HRIRs[3] as BRIRs with a reverberation time of 170ms [4] have been used.



RESULTS



REFERENCES

[1] - Sensible, Capt. Wot ; A&M, Santa Monica, California, United States: 1982.
 [2] - Baumann, Oliver; Belin, Pascal (2010): Perceptual scaling of voice identity: common dimensions for different vowels and speakers. In: *Psychological research* 74 (1), S. 110–120.
 [3] - Wierstorf, H., Geier, M., Raake, A., Spors, S. (2011) "A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances." 130th AES Convention, eBrief 6
 [4] - Wierstorf, Hagen. (2016). Binaural room impulse responses of a 5.0 surround setup for different listening positions [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.160761>