Behavior in triadic conversations in conditions with varying positions of noise distractors

Ľuboš Hládek and Bernhard U. Seeber

Audio Information Processing, Technical University of Munich, 80333 Munich, Germany
E-Mail: {lubos.hladek;seeber}@tum.de

Introduction

Face-to-face communication can be challenging in noisy environments for everyone [1] and people often experience difficulties with speech communication in cocktail party situations. Despite identifying factors that influence spoken communication, such as the level and position of background noise sources and talkers, and the exact position of the listener's head, the precise reasons for listening problems are often unknown. Therefore, replication of the problems in laboratory is difficult which closely relates to the fact that the acoustics at the ear of people who communicate in cocktail party situations change rapidly due to the movement of participants and sound sources, and reverberation [2]–[4]. To provide a broader perspective on communication in realistic situations, the present study aims to investigate the movement behavior of people engaged in triadic conversations. We accomplish this by creating a precisely controlled audio-visual virtual reality using the real-time Simulated Open Field Environment (rtSOFE) [5].

Previous studies have analyzed behavior during real communication in acoustic scenes involving groups of two, three or more individuals discussing pre-selected generic topics such as spotting differences in picture cards, weather, movies, moral dilemmas, or general topics related to hearing problems, e.g., [6]–[8]. Alternatively, participants have engaged in conversation without any specific instructions about the topic and can choose to discuss anything they like [9]. Typically, these studies involved seated participants, with the primary experimental manipulation being the level of background noise. The studies analyzed various parameters such as head movement orientations, interpersonal distance, changes in speech production effort or parameters related to conversation analysis. However, many of these studies used diffuse sound fields, which limit participants' access to the benefits of spatial unmasking [1]. Therefore, any horizontal head movements would only have a marginal effect on speech intelligibility from an acoustic standpoint. Situations with a discrete noise source, on the other hand, can show the highest possible benefit of head orientation for speech intelligibility. Spatial unmasking benefits may also relate the phenomenon of undershooting behavior observed in these studies, where listeners offset the horizontal angle of their head with respect to the talker, rather than looking directly at the speaker, and this behavior is commonly observed in conversations of three people [6], [9].

The present study investigated whether participants' movement behavior in free, unscripted conversations is affected by changes in the spatial configuration of interfering noise and whether profiles of head orientation benefits related to spatial unmasking can be manipulated. This manuscript presents preliminary data from two participant groups, focusing on analyzing undershooting behavior, which is believed to be influenced by the spatial configuration of interfering noise source. We expect that multiple distributed noise sources would trigger a different profile of spatial unmasking than a single noise source, resulting in a different type of orientation behavior of the participants. Our previous conference contribution [10] provides a summary of the methods and preliminary analysis of interpersonal distance using the data from the current experiment.

Methods

The experiment involved two groups of three participants who were tested for face-to-face communication in acoustic scenes. Prior to the experiment, standard audiometric testing was conducted to assess the pure-tone thresholds of the participants, which were found to be below or equal to 20 dB SL. The study was approved by the university's ethical committee (65/18S), and all participants provided written informed consent.

The experiment was conducted in the real-time Simulated Open Field Environment (rtSOFE ver. 4.0) [5]. The rtSOFE is a comprehensive system for audio-visual virtual reality creation housed within a full anechoic chamber. The auralization is based on the high-performance real-time room acoustic simulation and auralization software rtSOFE, which is a freely available real-time implementation of the image source method [11]. The audio signals were delivered through loudspeakers placed on a square-shaped frame that defined the experimental area (4m x 4m) of acoustic free-field where multiple people could interact for the communication experiment. Four silent projectors were used to project a visual representation of the environment in all horizontal directions around the participants.

The rtSOFE system had 61 loudspeakers, with 36 positioned at the height of 1.4 meters at 10 degree intervals in the horizontal plane, and the rest distributed at elevations above and below this plane. The space was further equipped with a video-based motion tracking system that accurately and frequently recorded the position and orientation of tracking objects. The participants wore plastic crowns with reflective spheres attached to record their head positions and orientations during the conversational experiment. Two of the participants were equipped with a high-quality head-set microphone connected to the system via a low-latency wireless audio transmission system, while the third participant used a wired head-set microphone. All microphones were equalized for each participant using a reference measurement microphone before the start of the experiment, and the loudspeakers were calibrated for flat frequency response in the range of 100 Hz to 18 kHz. The motion tracking system

was routinely calibrated with respect to a pre-defined reference point. One participant was also equipped with a full-body motion tracking suit and an eye tracker, although these data were not analyzed for this study.

The experimental environment consisted of the audio-visual simulation of an underground station, modeled according to a real station in Munich city center [12]. The geometric acoustic model and visual model are freely available [13], the acoustic model was previously evaluated and tested in terms of preservation of speech intelligibility cues. Participants saw the visuals of the underground station and their speech was picked-up by the microphones and reverberated in real-time. The simulation included only reflections starting from first order since the direct sound was the own speech of the participants. A discrete noise source, without any visual representation, was added to the acoustic simulation from a nearby location. Real-time acoustic simulation was performed using the rtSOFE software, a room acoustic simulation and real-time low-latency convolution, controlled by the Unreal Engine rtSOFE Controller plugin that relayed information about the position and orientation of acoustic objects and receivers from UE to rtSOFE. Visual rendering was achieved using nDisplay and a powerful visual rendering computer that was connected to the four video projectors.

Upon arriving at the rtSOFE laboratory and completing all necessary calibration procedures, participants were positioned in one of three initial positions (P1, P2, or P3). They were informed that they were free to move around as desired, with the exception of P1, where the participant wore a wired microphone and wired eye-tracker and was instructed to be mindful of the cables. All participants reported using English regularly in their studies or work and believed they had sufficient proficiency to engage in small-talk conversations with colleagues or fellow students. Although some participants knew each other from previous university courses, this was not a requirement for participation. Prior to the main experiment, the participants spent significant time together in one room during a preparation period, which helped facilitate their acquaintance.

The objective of the experiment was for participants to talk for 27 minutes about any topic of their choice while speaking in English, their second language. The only condition that was systematically manipulated during the experiment was the spatial configuration of the noise source. A broad-band speech-shaped noise source without temporal modulations was added to the underground scene and presented always at 72.2 dB SPL measured at the center. The noise source was placed in one of the predefined positions of the underground scene (1, 4, 7, or 11), which effectively were either in Front, to the Left, at the Back, to the Right of the participant at position P1. Uncorrelated noise could be also coming from all four positions at the same time but with the level equalized to 72.2 dB SPL at the center, to create a situation with a somewhat diffuse sound field and with a reduced potential for spatial unmasking. The sixth condition was the quiet condition. Each condition was held constant for 90 seconds when it always changed to another condition. Each condition was repeated three times leading to 27 minutes for the whole conversation. The order of conditions was randomized for each group of participants.

For this manuscript, only data from motion tracking were analyzed. The motion tracking data were recorded using the Optitrack Motive (v 2.0.1) software and recomputed with a newer version of the same software (v 3.0.1), which helped to reduce the number of measurement artifacts. The motion tracking data were recorded in synchrony with the sound presentation system (eSync 2.0, Optitrack), the starting points and the endpoints of the recording were determined from the Motive network stream using a MATLAB (v9.9) GUI that controlled the pace of the experiment and switched conditions.

Results

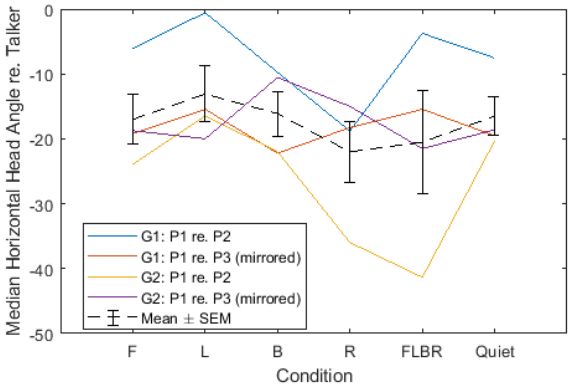

Figure 1: Analysis of orientation behavior, demonstrating an undershoot of the target azimuth. The data show median horizontal head angle of two participants at position P1 with respect to either participant at position P2 or P3 during different conditions (x-axis) in times when the respective reference participants were speaking. The figure shows individual data (color lines) from two groups (G1 and G2) and the mean across individual data (black dashed line) with SEM error bars.

Figure 1 shows median horizontal angles of two individual participants with respect to two fellow interlocutors in different experimental conditions (x-axis). The figure also shows across-subject mean. These data reflect horizontal head angles of participants at position P1 when either of the two other interlocutors (at P2 or P3) was speaking.

The data show that average undershooting was approximately constant and ranged from 15 to 20 degrees across all experimental conditions. The figure also shows substantial individual differences, but this seems to be an overall offset between participants.

Discussion

This preliminary analysis evaluated undershooting behavior of two participants at starting position P1 with respect to other two fellow interlocutors during the times when the reference interlocutors were speaking. The findings in the condition with distributed noise sources (FLBR) and Quiet condition can be compared with previous research investigating undershooting behavior in different levels of background noise, which also reported undershooting in a similar range of values [6] or slightly lower by a few degrees [9].

The results indicated a consistent undershooting pattern across all experimental conditions, including the Quiet condition, thus undershooting had little dependence on noise source location. However, change of undershooting with respect to noise source location is present is individual data. Although, this is only a preliminary analysis, one possible explanation for the constant undershooting is that the pattern reflects a general strategy that is optimized for many different situations, where background noise may originate at different positions with different profiles of spatial unmasking, while the strategy enables to preserve visual cues for speech perception [14] and be consistent with gaze aversion [15].

Our study investigated the behavior of participants with minimal movement restrictions in an unscripted experimental paradigm. While the findings showed a consistent undershooting pattern, it is important to note that the evaluation was conducted on a limited number of participants. Therefore, further data collection is necessary to draw more definitive conclusions. Nevertheless, our study serves as a foundation for future investigations of movement behavior in people with hearing aids and people who experience problems in spoken communication in cocktail party situations. Such investigations could help improve future hearing technologies and contribute to a better understanding of the core principles of behavior in acoustic scenes.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 352015383 – SFB 1330, Project C5. rtSOFE development was supported by the Bernstein Center for Computational Neuroscience, BMBF 01 GQ 1004B.

References

[1] A. W. Bronkhorst, “The cocktail-party problem revisited: early processing and selection of multi-talker speech,” Attention, Perception, Psychophys., vol. 77, no. 5, pp. 1465–1487, Jul. 2015, doi: 10.3758/s13414-015-0882-9.

[2] S. Gatehouse and M. A. Akeroyd, “The Effects of Cueing Temporal and Hearing-Impaired Listeners,” Trends Amplif., vol. 12, no. 2, pp. 145–161, 2008.

[3] W. O. Brimijoin, D. McShefferty, and M. A. Akeroyd, “Undirected head movements of listeners with asymmetrical hearing impairment during a speech-in-noise task,” Hear. Res., vol. 283, no. 1–2, pp. 162–168, 2012, doi: 10.1016/j.heares.2011.10.009.

[4] L. V. Hadley, W. O. Brimijoin, and W. M. Whitmer, “Speech, movement, and gaze behaviours during dyadic conversation in noise,” Sci. Rep., vol. 9, no. 1, p. 10451, Dec. 2019, doi: 10.1038/s41598-019-46416-0.

[5] B. U. Seeber, S. Kerber, and E. R. Hafter, “A system to simulate and reproduce audio–visual environments for spatial hearing research,” Hear. Res., vol. 260, no. 1–2, pp. 1–10, Feb. 2010, doi: 10.1016/j.heares.2009.11.004.

[6] L. V. Hadley, W. M. Whitmer, W. O. Brimijoin, and G. Naylor, “Conversation in small groups: Speaking and listening strategies depend on the complexities of the environment and group,” Psychon. Bull. Rev., vol. 28, no. 2, pp. 632–640, Apr. 2021, doi: 10.3758/s13423-020-01821-9.

[7] T. Beechey, J. M. Buchholz, and G. Keidser, “Eliciting Naturalistic Conversations: A Method for Assessing Communication Ability, Subjective Experience, and the Impacts of Noise and Hearing Impairment,” J. Speech, Lang. Hear. Res., vol. 62, no. 2, pp. 470–484, Feb. 2019, doi: 10.1044/2018_JSLHR-H-18-0107.

[8] O. Tuomainen, L. Taschenberger, S. Rosen, and V. Hazan, “Speech modifications in interactive speech: Effects of age, sex and noise type,” Philos. Trans. R. Soc. B Biol. Sci., vol. 377, no. 1841, pp. 1–23, 2022, doi: 10.1098/rstb.2020.0398.

[9] H. Lu, M. F. McKinney, T. Zhang, and A. J. Oxenham, “Investigating age, hearing loss, and background noise effects on speaker-targeted head and eye movements in three-way conversations,” J. Acoust. Soc. Am., vol. 149, no. 3, pp. 1889–1900, Mar. 2021, doi: 10.1121/10.0003707.

[10] L. Hladek and B. U. Seeber, “Effects of noise presence and noise position on interpersonal distance in a triadic conversation.,” in Proceedings Internoise 2022, 2022.

[11] B. U. Seeber and T. Wang, “real-time Simulated Open Field Environment (rtSOFE) software package.” 2021, doi: https://doi.org/10.5281/zenodo.5648304.

[12] L. Hladek, S. D. Ewert, and B. U. Seeber, “Communication Conditions in Virtual Acoustic Scenes in an Underground Station,” in 2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA), 2021, pp. 1–8, doi: 10.1109/I3DA48870.2021.9610843.

[13] L. Hladek and B. U. Seeber, “Underground station environment.” 2022, doi: 10.5281/zenodo.5532643.

[14] A. MacLeod and Q. Summerfield, “Quantifying the contribution of vision to speech perception in noise.,” Br. J. Audiol., vol. 21, no. 2, pp. 131–41, May 1987.

[15] C. Acarturk, B. Indurkya, P. Nawrocki, B. Sniezynski, M. Jarosz, and K. A. Usal, “Gaze aversion in conversational settings: An investigation based on mock job interview,” J. Eye Mov. Res., vol. 14, no. 1, May 2021, doi: 10.16910/jemr.14.1.1.