



## Effects of noise presence and noise position on interpersonal distance in a triadic conversation.

Lubos Hladek<sup>1</sup>

Audio Information Processing, Technical University of Munich  
Arcisstr. 21, 80333, Munich, Germany

Bernhard U. Seeber<sup>2</sup>

Audio Information Processing, Technical University of Munich  
Arcisstr. 21, 80333, Munich, Germany

### ABSTRACT

*People usually move in different ways during conversations but little is known whether movement behavior reflects difficulties that some people experience in communication in terms of listening and speaking. Here, we investigate movement behavior of standing participants in triadic conversations in noise with different spatial properties. In the initial analysis, we test how presence and position of noise affects interpersonal distance. The task of the interlocutors was to hold a free conversation in controlled acoustic conditions with realistic reverberation and visual scene of an underground station that was rendered in real-time in the Simulated Open Field Environment while the participants were monitored with motion tracking equipment and their voices were recorded. Two groups of people took part in the study. Different noise conditions involved presence of stationary speech-shaped noise at 70 dB SPL presented from one of four possible directions (Front, Right, Rear, Left), or from all directions simultaneously (F+R+R+L), or there was no noise (Quiet). The results show that interpersonal distance decreased when the noise was present, which confirmed previous results. Further, the preliminary analysis shows a relatively small effect of noise position on the interpersonal distance.*

### 1. INTRODUCTION

Having a conversation in a noisy situation is often challenging for people with hearing impairment but hearing research has been struggling in pinpointing the exact moments that replicate the difficulties, and it has been struggling in defining the measures that predict the difficulties with high accuracy [1]. One reason for this issue is the fact that the traditional lab measures (at least in the audiology domain) focus only on listening, although with well-controlled and well-defined stimuli. However, the problem of inter-personal communication is much more complex and involves exchange of spoken words in complex acoustic conditions, where people move and attend to multiple objects at the same time, where sounds and visual objects move, and where background noise is present. An approach that circumvents some of the issues of standard audiological testing is to analyze behavior during conversations with no or minimal intervention to the communication itself [2,3].

---

<sup>1</sup> lubos.hladek@tum.de

<sup>2</sup> seeber@tum.de

One way of reformulating the problem of communication in noise is to assume that the person performs a signal-to-noise ratio (SNR) optimization during conversation. However, the SNR is affected by a number of external factors like presence of noise, distance of interlocutors, spatial configuration of sound sources and the listener position, speech level (Lombard effect), or internal factors like attention, context of the discussion, familiarity, turn taking behavior, speech prosody, native language and so on. Among these, the focus of the current work is to analyze behavior in terms of interpersonal distance because we assume that this is one of the primary behavioral markers that relate to speech communication (e.g., [4–6]). When people have problems understanding each other in a noisy situation, they come closer to each other. This has been previously observed in a triadic conversation with sitting participants [7] or in a dyadic conversation of standing participants [8]. The current investigation addresses the question how presence and position of a noise source influences interpersonal distance in a triadic conversation of free-standing participants.

## **2. METHODS**

### **2.1. Participants**

Hence six young (1 female, 5 males) participants (including first co-author) took part in a non-structured triadic conversation (2 groups of three different people). People talked in English but it was not their mother tongue. The participants were international students, research assistants and research associates who use English during their studies and work. All participants had hearing thresholds below or equal to 20 dB HL at standard audiological frequencies between 250 Hz and 8 kHz tested using pure-tone audiometry. All participants provided written informed consent prior to the start of the experiment. Procedures and methodologies were approved by the Ethical committee of the Technical University of Munich (65/18S).

### **2.2. Environment**

The experiment was conducted in the real-time Simulated Open Field Environment (rtSOFE) which is a comprehensive set of tools for audio-visual virtual reality located in our anechoic chamber (10 m x 6 m x 4 m; l x w x h) [9,10]. To recreate the acoustic environment, the sounds (noise sources and reverberation of spoken speech) were presented over an array of 61 loudspeakers distributed horizontally and vertically inside the laboratory. The loudspeakers (Dynaudio BM6A mkII, Dynaudio, Skanderborg, Denmark) were driven by digital-to-analog converters (RME 32DA, Audio AG, Haimhausen, Germany) connected to a multi-channel sound card (RME HDSPe, Audio AG, Haimhausen, Germany). The visual presentation system consisted of four low-noise (32 dBA) projectors (Barco F50 WQXGA, Barco, Kortrijk, Belgium) projecting the visual of the virtual environment on four acoustically transparent screens surrounding the participants during the conversation.

The virtual environment was based on the visual and acoustic model of the underground station Theresienstraße [11,12] and it was rendered using a visual renderer (Unreal Engine 4.25, Epic Games, USA) which was controlling the positions of the virtual objects in the open-source room acoustic simulation tool rtSOFE (v1.1) [13]. The visuals included the surrounding walls around the listener position (in [11] position R1) with the running escalator (without escalator sound), the rendered sounds included the noise sound source (see experimental conditions below) and real-time simulation of realistic reverberation [12] of the voices of conversing people.

The voice input to the audio signal processing chain was obtained from one wired (C520, AKG) and two wireless head-worn microphones (VT 800, Voice Technologies, Switzerland, on DWR-R02DN and DWT-B01N digital wireless transmitters, Sony, Japan) connected via an analog-digital converter (Micstasy, RME, Germany) to the sound card. The noise signals were played using a software interface [14]. The signals were then convolved in real-time (buffer size 128) on 61 channels (using open source program Convolver developed at the institute [13]) and sent to the loudspeakers that were equalized in terms of frequency and phase response using 512 tap FIR filters. The delay from the (wireless) microphone to the loudspeaker was approximately 15 ms. The impulse responses of the real-time convolution were updated by the room simulation software (open source program rtsofe

developed at the institute [13]) whenever the participants moved during the conversation. Their position was tracked by the camera-based motion tracking device (OptiTrack Prime 17W, NaturalPoint Inc. Corvallis, Oregon, USA) running on 359 Hz although the position updates for the room simulation software were capped at 30 Hz due to limitations of the visual rendering. The acoustic rendering involved real-time updated early reflections (up to 100 ms) modeled using the image source method and static late reverberation obtained from multi-channel recordings of the environment (available at [11]). For details and acoustic and perceptual evaluation of the impulse responses see [12]. The acoustic rendering considered the directivity of human speech and the microphone acoustic characteristics determined individually for each participant prior to the experiment.

Since the focus of the research was behavior of people during conversation, the following data were recorded: speech and head positions (using motion tracking crowns) of all three participants. One of the participants also wore a full body motion tracking suit and an eye tracker (Pupil Core, Pupil Labs, Google, USA) data of which were also recorded but are not analyzed here. The experiment and calibration procedures were controlled by programs written in MATLAB (v9.9.0, Mathworks, USA) and Python (v3.9) via a GUI displayed on a hand-held tablet.

### **2.3. Procedures, stimuli and conditions**

The conversations were organized in triads. The task of the three participants was to talk on any topic they liked for about 30 minutes. Participants had no restrictions on the movement, except the participant who wore the eye tracker was asked to be aware of the cable and the tracking points on the motion-tracking suit. The participants stood approximately in the center of the loudspeaker array about 1.5 meters from each other and  $\pm 60$  degrees horizontally apart.

The conversation took place in a noisy environment, which was created by placing a virtual sound source with 70 dB SPL in the underground environment [11] at one of four possible positions at  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $-90^\circ$  azimuthal angle at 1.6 m distance relative to the default listener position (in [11] the default listener position corresponds to the position R1, the sources correspond to source positions 1, 4, 7, 10). The noise was a sample of steady state speech-shaped noise created from the long-term average spectrum of one OLSA list [15] by randomizing the phase of the Fourier spectrum. The level corresponded to the level of the direct sound measured at the center of the loudspeaker array, thus the actual level including reverberation was slightly higher. In another condition, the noise emanated from all four positions simultaneously (the sources were uncorrelated) and there was also a silent condition. The real-time processing of speech signals was always present during the conversation.

Therefore, there were six noise conditions: Front, Right, Rear, Left, F+R+R+L (combined Front, Right, Rear, Left), Quiet. During the conversation, the noise condition always changed after 90 seconds and each condition was played three times. Thus, the whole experiment consisted of 18 blocks of 90 seconds. The order of condition was randomized such that the same condition was never played twice in a row. The blocks changed after one or two seconds but there was no break to the conversation.

Before the conversation, the participants were instructed on the safety of the laboratory (participants were monitored by a CCTV and intercom from outside by an experimenter). The preparation started by putting on the motion tracking suit on one of the participants, placing the motion tracking markers, and calibrating the full body motion tracking system (data not analyzed here). The eye tracker was also installed on the participant (data not analyzed here). In the next step, all participants were equipped with the motion tracking crowns and head-worn microphones (one wired and two wireless). When all participants were inside the laboratory, the head tracking crowns were calibrated on each participant. Then each participant went through a calibration procedure of the microphone. During the calibration procedure, each participant stood one meter from a measurement microphone (MM 210, Microtech Gefell, Germany) and spoke for 20 seconds. By comparing the power spectra of the measurement and head worn microphones, a FIR filter was created that corrected for the differences (i.e., all microphones were set to the same level and frequency response). Just before the start of the experiment, the interlocutor with the eye tracker had to look at pre-defined points to calibrate the eye tracker.

## 2.4. Analysis

The data were recorded in a raw format using the motion tracking software (Motive v 2.0.1, NaturalPoint, USA), and using the Playrec interface for Matlab [14]. The correct time alignment of the different data sources was assured by time-syncing the motion tracking software with the sound card using a word-clock signal (eSync 2, NaturalPoint Inc. Corvallis, Oregon, USA), while at the same time the software was streaming the frame-swap time stamps over local UDP network, which were used to timestamp the beginnings and ends of audio recordings of each experimental noise condition. Small desynchronization was caused by the Matlab interface running the recording script but for the half-hour recording, it was less than 0.5 s (worst case). The quality of the raw data was manually checked, in some cases, when the motion tracking had low quality (i.e., was lost due to visual interference) the data were manually removed from the dataset. Analysis of the data was performed in MATLAB, statistical evaluation was run using software CLEAVE [16].

## 3. RESULTS

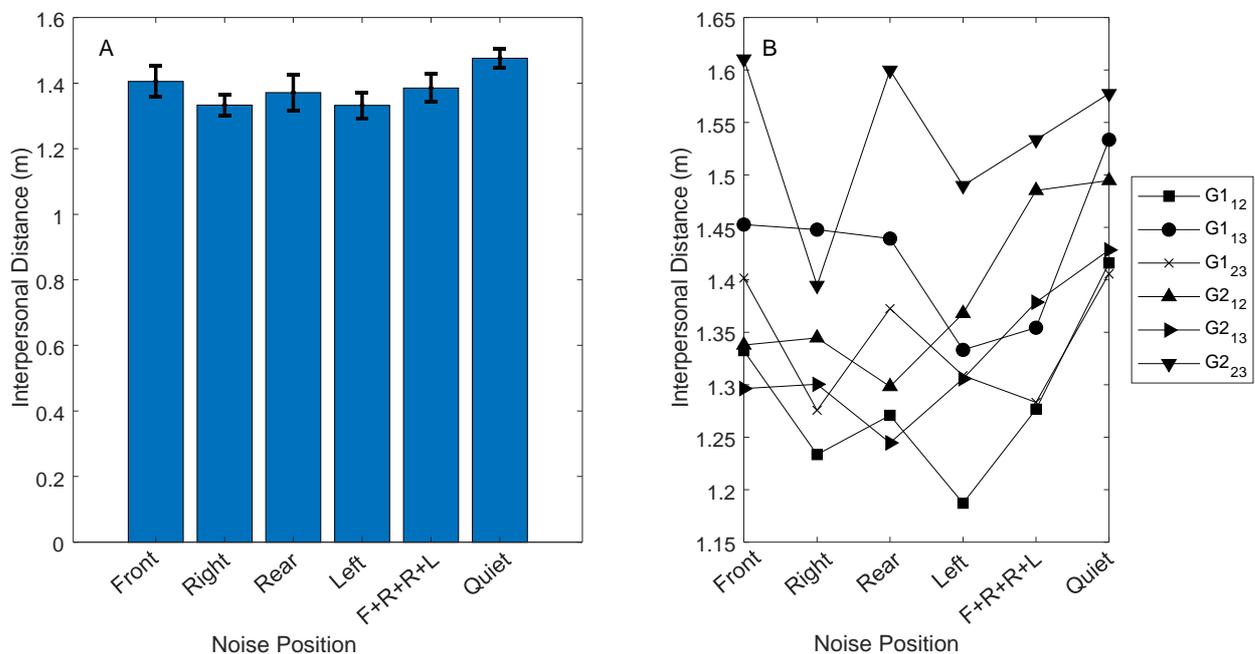


Figure 1: Interpersonal distance between pairs of interlocutors during the conversation. A – average data. Data show across-subject means, error bars show standard errors of the means for different noise conditions. B – individual data. The legend shows the group number in capital letters and index of individual pair in subscript.

Figure 1A shows inter-personal distance between each pair of participants during conversation. The data are shown for six experimental conditions given on the abscissa. Each data point in Figure 1A represents the mean of six pairs from two groups (of three people). When the data were submitted to factorial analysis ANOVA with a single factor of condition, it showed significant main effect of condition ( $F_{5,25} = 5.5$ ,  $p < 0.05$ , Geisser-Greenhouse corrected). The main effect is mainly driven by the increase of the distance in the *Quiet* condition, which was statistically significant from all other conditions (t-test,  $p < 0.0001$ , Bonferroni corrected probability: 0.0033). The post-hoc statistical analysis also showed differences between other conditions but these were smaller in magnitude. Detailed inspection of the individual data (Figure 1B) shows that the pairs G1<sub>23</sub> and G2<sub>23</sub> drive the differences between Front, Right, Rear, and Left because these two pairs show a similar zigzag pattern as the average data. People in these pairs were indeed closer to the noise sources at positions Right and Left than at Front and Rear, which might explain the across-subject differences.

## 4. DISCUSSION

The data show that the interlocutors were farther apart (on the magnitude 10 cm – 20 cm) in the Quiet condition in comparison to the conditions with one noise source or with diffuse noise. The results of this preliminary analysis (so far only of two groups each of three people) confirm previously observed trends [7,8]. The presence of noise (here at 70 dB SPL) decreased interpersonal distance of the interlocutors in a free triadic conversation by on average by ~10-20 cm which corresponds to ~1 dB change in terms of change of acoustic intensity ( $1/r^2$  law) in comparison to a condition without noise. In hearing sciences, the effects at this magnitude are often considered as small, but these results show a natural response of people to the presence of noise and it is reasonable to assume that the behavioral change served to maintain or at least support successful communication. This preliminary result supports the hypothesis that severe acoustic conditions lead to behavioral changes and suggests that the acoustic effects related to body movements may need careful and sensitive measures since they may not lead to huge effects, in terms of dB change, at least when looking at average data. Other aspects of movement behavior, like movement synchronization, switching attention, gaze movements, or specific gestures, may exhibit more details about the movement strategies that people employ when they optimize their communication with other people.

## 5. ACKNOWLEDGMENTS

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

## 6. REFERENCES

1. Keidser, G. *et al.* The Quest for Ecological Validity in Hearing Science: What It Is, Why It Matters, and How to Advance It. *Ear Hear.* **41**, 5S-19S (2020).
2. Hadley, L. V., Brimijoin, W. O. & Whitmer, W. M. Speech, movement, and gaze behaviours during dyadic conversation in noise. *Sci. Rep.* **9**, 10451 (2019).
3. Beechey, T., Buchholz, J. M. & Keidser, G. Eliciting Naturalistic Conversations: A Method for Assessing Communication Ability, Subjective Experience, and the Impacts of Noise and Hearing Impairment. *J. Speech, Lang. Hear. Res.* **62**, 470–484 (2019).
4. Latif, N., Barbosa, A. V., Vatiokiotis-Bateson, E., Castelhana, M. S. & Munhall, K. G. Movement coordination during conversation. *PLoS One* **9**, 1–10 (2014).
5. Sørensen, A. J. M., Fereczkowski, M. & MacDonald, E. N. Effects of Noise and Second Language on Conversational Dynamics in Task Dialogue. *Trends Hear.* **25**, 233121652110244 (2021).
6. Tuomainen, O., Taschenberger, L., Rosen, S. & Hazan, V. Speech modifications in interactive speech: Effects of age, sex and noise type. *Philos. Trans. R. Soc. B Biol. Sci.* **377**, (2022).
7. Hadley, L. V., Whitmer, W. M., Brimijoin, W. O. & Naylor, G. Conversation in small groups: Speaking and listening strategies depend on the complexities of the environment and group. *Psychon. Bull. Rev.* **28**, 632–640 (2021).
8. Weisser, A., Miles, K., Richardson, M. J. & Buchholz, J. M. Conversational distance adaptation in noise and its effect on signal-to-noise ratio in realistic listening environments. *J. Acoust. Soc. Am.* **149**, 2896–2907 (2021).
9. Seeber, B. U., Kerber, S. & Hafter, E. R. A system to simulate and reproduce audio–visual environments for spatial hearing research. *Hear. Res.* **260**, 1–10 (2010).
10. Seeber, B. U. & Clapp, S. W. Interactive simulation and free-field auralization of acoustic space with the rtSOFE. *J. Acoust. Soc. Am.* **141**, 3974–3974 (2017).
11. Hladek, L. & Seeber, B. U. Underground station environment. (2022). doi:10.5281/zenodo.5532643
12. Hladek, L., Ewert, S. D. & Seeber, B. U. Communication Conditions in Virtual Acoustic

- Scenes in an Underground Station. in *2021 Immersive 3D Audio from Archit. to Automot.* 1–8 (IEEE, 2021). doi:10.1109/I3DA48870.2021.9610843
13. Seeber, B. U. & Wang, T. real-time Simulated Open Field Environment (rtSOFE) software package. (2021). doi:<https://doi.org/10.5281/zenodo.5648304>
  14. Humphrey, R. Playrec. (2015). at <<https://github.com/PlayrecForMatlab>>
  15. Wagener, K. C., Brand, T. & Kollmeier, B. Entwicklung und Evaluation eines Satztests in deutscher Sprache Teil II: Optimierung des Oldenburger Satztests. *ZEITSCHRIFT FUR Audiol.* **38**, 44–56 (1999).
  16. Herron, T. C Language Exploratory Analysis of Variance with Enhancements. (2005). at <<http://www.ebire.org/hcnlab/software/cleave.html>>