

Influence of (non-) intelligible background speech on memory and listening effort in conversational situations

Cosima A. Ermert¹, Jonathan Ehret², Chinthusa Mohanathasan³, Andrea Bönsch²
Torsten W. Kuhlen², Sabine J. Schlittmeier³, and Janina Fels¹

¹*Institute for Hearing Technology and Acoustics, RWTH Aachen University*

²*Visual Computing Institute, RWTH Aachen University*

³*Teaching & Research Area Work and Engineering Psychology, RWTH Aachen University*

Email: cosima.ermert@akustik.rwth-aachen.de

Introduction

Verbal communication depends on a listener's ability to accurately comprehend and recall information conveyed in a conversation. Background noise, such as speech, can significantly impair speech processing. While previous research has explored the effects of native (intelligible) and foreign (unintelligible) background speech on memory, the tasks used in these studies were rather simple, e.g., serial recall [1].

Recent advancements in cognitive research have led to the development of the heard-text recall (HTR) paradigm [2], which can be used in a dual task design to assess both listening effort and memory performance. In contrast to traditional tasks such as serial recall, this paradigm uses running speech to simulate a conversation between two talkers. It allows for talker visualization in virtual reality (VR) [3], effectively conveying co-verbal visual cues like lip movements, turn-taking cues, and gaze behavior. While this paradigm has been investigated under pink noise [4], the impact of more realistic noise, such as speech, remains unexplored.

In this study, we administered the HTR as a dual task in VR under three noise conditions: *silence*, *intelligible speech*, and *unintelligible pseudo-speech*.

Methods

Participants

$N=24$ native German participants (9 female, age: 22-33, $M = 25.21$, $SD = 2.55$) were recruited for the listening experiment. They had to pass an audiometry (below 25 dBHL) and a Snellen Test (20/30) [6] to take part in the study. Informed written consent was obtained from all participants. They received a 10-euro voucher for a local bookstore.

Paradigm

The dual task HTR paradigm consists of a primary and a secondary task. In the primary tasks, participants listen to family stories consisting of 10 sentences. Afterwards, they are asked 9 content-related questions per text. For more details, please refer to Schlittmeier et al. [2]. The secondary task is vibrotactile (cf. [4]). Participants hold one HTC Vive controller in each hand. The controllers vibrate in four distinct patterns: short-short, long-long, short-long, and long-short. If the second vibration is a repetition of the first vibration (i.e., short-short, long-long),

participants have to click the left controller. Otherwise, they have to click the right controller. Both tasks are administered together in dual tasking and, also, in isolation to obtain a baseline performance.

Implementation

The experiment was conducted in a living room VR environment, which was presented via an HTC Vive Pro Eye HMD and created in Unreal Engine 5.3. In the VR environment, participants saw two embodied conversational agents (ECAs) [8], one male and one female MetaHuman, narrating the HTR stories as a conversation. The ECAs were animated with gestures, lip movement, and gazing. The study was implemented based on the study by Ehret et al. [3] using the StudyFramework [9].

The auditory scene was created in Virtual Acoustics v2022a [7] and reproduced via Sennheiser HD650 headphones. The scene was auralized with a generic head-related transfer function (HRTF), and headphones were equalized per participant after Masiero and Fels [10]. The target talker animations and stimuli were taken from the AuViST database [5]. In the *silence* condition, only the target talkers were audible at a distance of 1.35 m and a horizontal offset of $\pm 45^\circ$ from the frontal direction. In the *intelligible speech* condition, additional distractor sound sources were placed at $\pm 90^\circ$ from the frontal direction, i.e., to the left and right. The distractor stimuli were Oldenburger Sentence Test (OLSA) [11] sentences spoken by a male and female talker. In the *unintelligible pseudo-speech* condition, distractor sources were also placed at $\pm 90^\circ$. Instead of using foreign speech stimuli, the OLSA sentences were cut into syllables and rearranged in a random matter. Crossfading of 10 ms was applied to smooth transitions. This way, both the target and background speech had the same frequency spectrum. The stimuli were calibrated to a signal-to-noise ratio (SNR) of +3 dB(A) using an HMS III artificial head.

Procedure

After passing the screening, participants were introduced to the virtual environment. First, the vibrotactile task could be trained. Afterwards, one HTR text was presented without the dual task as training. Subsequently, participants practiced the dual tasking with one more HTR text. Then, the main experiment started. The experiment consisted of three blocks, one per noise condition. The order of blocks was balanced. In each block, 4 HTR

texts had to be completed in dual tasking (together with the vibrotactile task) and 1 HTR text in single-tasking. No texts were repeated. Additionally, 40 trials of the vibrotactile task were administered in single-tasking. The order of these tasks was balanced. The experiment took around 90 minutes.

Results

The percentage of correctly answered questions in the HTR and correctly categorized vibration patterns in the vibrotactile task were examined with the fixed factors *noise condition* (*silence, intelligible speech, unintelligible pseudo-speech*) and *number of tasks* (single, dual tasking) using generalized linear mixed models with a binomial distribution family and a logit link function. The preliminary results show a significant main effect of the number of tasks and the noise condition, indicating an influence of the content of the background noise. A peer-reviewed publication with more details and results is ongoing.

Literatur

- [1] Yadav, M, Georgi, M., Leist, L., Klatte, M., Schlittmeier, S.J., and Fels, J.: Cognitive Performance in Open-Plan Office Acoustic Simulations: Effects of Room Acoustics and Semantics but Not Spatial Separation of Sound Sources. *Applied Acoustics* 211 (2023), 109559
- [2] Schlittmeier, S.J., Mohanathanasan, C., Schiller, I.S., and Liebl, A.: Measuring text comprehension and memory: A comprehensive database for Heard Text Recall (HTR) and Read Text Recall (RTR) paradigms, with optional note-taking and graphical displays. RWTH Publications (2023)
- [3] Ehret, J., Bönsch, A., Nossol, P., Ermert, C.A., Mohanathanasan, C., Schlittmeier, S.J., Fels, J., and Kuhlen, T.W.: Who's next? Integrating Non-Verbal Turn-Taking Cues for Embodied Conversational Agents. *IWA '23* (2023) 27, 1-8
- [4] Mohanathanasan, C., Ermert, C.A., Fels, J., Kuhlen, T.W., and Schlittmeier, S.J.: Exploring short-term memory and listening effort in two-talker conversations: The influence of soft and moderate background noise. *PLoS ONE* 20(2) (2025), e0318821
- [5] Ermert, C.A., Mohanathanasan, C., Ehret, J., Schlittmeier, S.J., Kuhlen, T.W., and Fels, J.: AuViST - An Audio-Visual Speech and Text Database for the Heard-Text-Recall Paradigm. RWTH Publications (2023)
- [6] Snellen, H.: Probebuchstaben zur Bestimmung der Sehschärfe. H. Peters, 1873.
- [7] Institute for Hearing Technology and Acoustics, RWTH Aachen University, Stienen, J., Aspöck, L., and Vorländer, M.: Virtual Acoustics - A real-time auralization framework for scientific research. *Zenodo* (2020), 10.5281/zenodo.13744474
- [8] Cassell, J.: Embodied Conversational Agents: Representation and Intelligence in User Interfaces. *AI Magazine*, 22 (4) (2021), 67â€“67.
- [9] Ehret, J., Bönsch, A., Fels, J., Schlittmeier, S.J., and Kuhlen, T.W.: StudyFramework: Comfortably Setting up and Conducting Factorial-Design Studies Using the Unreal Engine, 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (2024)
- [10] Masiero, B., and Janina F.: Perceptually Robust Headphone Equalization for Binaural Reproduction. Audio Engineering Society (2011)
- [11] Kuehnel, V., Kollmeier, B., and Wagener, K.: Entwicklung Und Evaluation Eines Satztests Für Die Deutsche Sprache I: Design Des Oldenburger Satztests. *Zeitschrift Für Audiologie* 38 (1999), 4-15.