DOI: 10.61782/fa.2025.0453

FORUM ACUSTICUM EURONOISE 2025

IMPACT OF ROOM ACOUSTIC MISMATCH ON CONVERSATIONAL DYNAMICS IN VIRTUAL ONE-ON-ONE CONFERENCING: PRELIMINARY RESULTS

Pablo Abehsera-Morell[1*] **David Poirier-Quinot**[1] **Brian F.G. Katz**[1]
[1] Sorbonne Universite, CNRS, Institut Jean Le Rond d'Alembert UMR7190, F-75005 Paris, France

ABSTRACT

The acoustics of a room can have an impact on speech production, influencing factors such as speech rate and vocal effort. In augmented reality teleconferencing, virtually rendered remote interlocutors can be auralised to create the impression of sharing a room with the user. Acoustical mismatches between the real rooms of the user and the remote interlocutors can hinder interaction — for instance, a remote talker in an acoustically-treated office might sound unnatural to a user located in a highly reverberant space, such as an untreated office with hard surfaces or a position close to a bare wall. This study investigates how such a mismatch can affect conversational interactions. Participant pairs engaged in audiovisual teleconferencing while experiencing either matched or mismatched auralised room acoustics. Each participant perceived the conversation happening in a single common acoustic, though that acoustic may be different from each participant's perspective. We present preliminary results on the impact of these conditions on interaction dynamics. The interactions between interlocutors were analysed for different matched or mismatched conditions, with the findings providing insights into the impact of said conditions on the quality of the interactions.

Keywords: *room acoustics, augmented reality, teleconferencing, interaction, vocal effort.*

**Corresponding author*: *pablo.abehsera_morell@sorbonne-universite.fr.*

1. INTRODUCTION

In mixed reality environments, virtual acoustic sources are rendered using auralisation to simulate desired acoustic properties. In augmented reality (AR), virtual sound sources overlaid onto a user's real-world environment should behave acoustically as if they were physically present in that space [1]. For instance, in an AR teleconference, one would expect the voice of a remote interlocutor to be auralised to match the acoustic characteristics of the listener's environment. Since each participant is in a different location, the acoustics at each end may vary significantly, leading to distinct auditory experiences for each interlocutor.

The auditory experience created by a room's acoustics can influence how a person adjusts their speech while talking inside it. Previous work has investigated the relationship between reverberation time and speech parameters such as level [2–4], rate [2, 5] or voicing period duration [6]. These adjustments often serve to improve intelligibility, allowing one's speech to be well understood by others in the room while minimising effort for any listeners [7].

In an AR teleconference where interlocutors find themselves in acoustically different environments, it could be expected that each one might adjust their own speech to suit the acoustics of the room they physically find themselves in. Because each one will be auralised in their interlocutors' acoustics, however, these speech adjustments could actually be detrimental to intelligibility on their interlocutors' end, hampering their interaction.

In this study, we set to assess the impact of such a mismatch between room acoustics on the conversational dynamics between two interlocutors in an audiovisual teleconferencing system with virtual acoustics. An experiment was conducted where participant pairs held conversations in either the same or different acoustics on each end. Three different acoustic conditions with distinct features were used.

2. METHOD

2.1 Virtual teleconferencing system

An audiovisual teleconferencing system was set up across two acoustically isolated studio rooms, consisting of a screen, a webcam (Logitech Streamcam), reference headphones (Sennheiser HD-650), a measurement microphone (Bedrock Audio BAMT1), and a chair placed $1.5\,\mathrm{m}$ away from the screen and microphone in each room. The system was controlled via a computer in a third room, which provided real-time rendering in simulated acoustics using the `ircamverb` artificial room reverberator of the Spat5 library [8] in Cycling ′74 Max. The virtual source and receiver were set $1.5\,\mathrm{m}$ apart. Each participant heard themselves and their interlocutor auralised within a distinct virtual room. The direct sound component was muted in each participant's auralisation of themselves.

System gains were calibrated, using a dummy-head (Neumann KU-100), so that the direct sound output from the headphones resulting from a person talking in the other room at $1.5\,\mathrm{m}$ from the microphone produced the same level as dry speech at $1.5\,\mathrm{m}$ from the dummy-head.

Figure 1. Experiment setup in each room (left, right) and what each participant saw (center).

2.2 Protocol

To assess the impact of room acoustic mismatches on conversational dynamics, participant pairs were instructed to have a conversation in the teleconferencing system, split into trials, with the virtual acoustics varying across different trials. In each trial, participants would find themselves in one of three acoustic conditions, outlined in Table 1. The motivation was to have conditions with increasing reverberance and decreasing clarity, making the first best suited for conversation and the last worst. Each participant

Parameter	Cond. A	Cond. B	Cond. C
Volume (m^3)	50	5000	200
$T30_{mid}$ (s)	0.30	4.65	4.91
$C50_{mid}$ (dB)	21.1	9.20	4.23

Table 1. Room parameters for each acoustic condition. *mid* refers to the average value across the $500\,\mathrm{Hz}$ and $1\,\mathrm{kHz}$ bands, across both left and right channels.

perceived the conversation happening in a single common acoustic during each trial, though that acoustic could be different from each participant's perspective.

The experiment consisted of 11 trials, with the first 2 lasting one minute and always having the same acoustics in the same order (**A** first, then **B**), in order to accustom participants to the task. The remaining 9 trials, containing all the possible combinations for acoustic condition on each end, lasted two and a half minutes each and their order was randomised for each participant pair.

At the end of each trial, the participants were asked to answer a short questionnaire, consisting of two statements: (Q1) *You found it easy to follow the conversation*, (Q2) *Your interlocutor had no trouble following the conversation*. For both statements, a 7-point Likert scale was displayed, with 1 meaning completely disagree and 7 meaning completely agree.

In order to keep the conversation flowing, a suggested topic of discussion was displayed at the top of each screen, and participants were instructed to speak only about the currently displayed topic. A list of 70 topics was curated, ranging from trivial ("*Should pineapple go on pizza?*") to more serious ("*Should art be separated from the artist?*"), and participants were allowed to cycle through the list at will (be it during a trial or in between trials) using either a handheld clicker or a keyboard. They were allowed to skip topics whenever they wanted and were free to not discuss a given topic if they did not want to.

10 participant pairs took part in the experiment, comprised of 3 female/male, 1 female/female, and 7 male/male. The mean participant age was 26.6 ± 2.7 years.

3. RESULTS

Analyses of variances (ANOVAs) were conducted to assess the impact of (**1**) *participants' own acoustic condition*, (**2**) *the interlocutor's acoustic condition* and (**3**) *mismtach between the two* on participant ratings and top-level objective parameters. The first-order interaction terms between factors were also analysed. Statistical significance was determined for p-values below a 0.05 threshold. The notation $p < \varepsilon$ is adopted to indicate p-values below 10^{-3}. Post-hoc pairwise comparisons for significant factors were made with Tukey-Kramer adjusted p-values, or with Wilcoxon rank-sum p-values for unbalanced comparisons. Participant ratings were z-score normalised, in order to prevent participants who used the extremes of the scale more from overpowering those who used them less, and viceversa. The shorter first two trials were excluded from analysis.

3.1 Subjective ratings

The participants' own acoustic condition had a significant impact on participants' ratings for question Q1 ($F = 170.6$, $p < \varepsilon$). As seen in Figure 2, participants found it easier to follow the conversation in acoustic **A** than in acoustic **B** (0.93 ± 0.12 vs. 0.02 ± 0.15), and in acoustic **B** than in acoustic **C** (0.02 ± 0.15 vs. -0.95 ± 0.15). The same statistically significant trend was observed for ratings for question Q2, which were generally slightly higher than their Q1 counterparts.

As illustrated in Figure 2, the acoustic condition of the interlocutor had no impact on ratings for either question. Likewise, no significant interaction effect was observed between participants' acoustics on either question.

3.2 Objective metrics

The own acoustic condition was observed to have a significant impact on the level of participants' speech ($F = 13.7$, $p < \varepsilon$). Participants' RMS speech level (dB), calculated per trial, was higher than the average level across all trials in acoustic **A** than in acoustics **B** and **C**, where the speech levels were similar and lower than the average per-participant level across-trials (0.39 ± 0.22 dB vs. -0.21 ± 0.16 dB vs. -0.25 ± 0.19 dB). The acoustic condition of the interlocutor had no impact on participants' speech level. These results are illustrated in Figure 3.

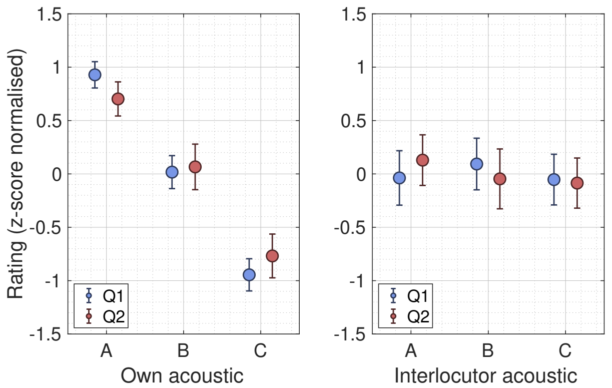

Figure 2. Participants' ratings on Q1 and Q2 as a function of (left) their own and (right) their interlocutor's acoustic condition. Mean and 95% confidence intervals (CI) displayed. Due to the normalisation, the ratings are expressed in terms of standard deviations from each participant's mean rating.

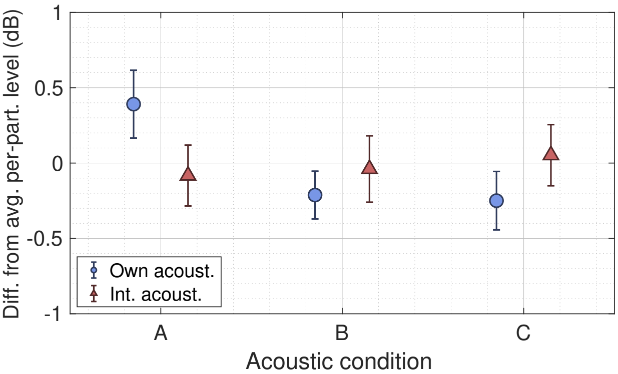

Figure 3. Difference from per-participant average RMS voice level across all trials, per own acoustic condition. Mean and 95% CI displayed.

4. DISCUSSION

The ratings of own ease in following the conversation and the observed change in voice level as a function of one's acoustic condition suggest that the teleconferencing system's auralisations created an ecologically valid environment, where, as seen in real life, room acoustics can impact voice production and perception and, therefore, conversational dynamics. The observed impact of one's own acoustics on ratings were as expected.

The ratings of the interlocutor's ability to follow the conversation could suggest that the acoustic condition of the interlocutor did not affect how the interlocutor was perceived, meaning that a mismatch of acoustics may not impact conversational dynamics. Participants often said, however, that they found it significantly harder to rate their partner, and, sometimes, that they felt uncomfortable giving them a low rating. It could be argued that the protocol may need to be changed in order to get meaningful responses to such a question.

Additional objective parameters should be analysed, to get a more comprehensive understanding of how the conversational dynamics evolved across trials.

5. CONCLUSION

In this paper, preliminary results of a study on the impact of room acoustic mismatches on conversational dynamics in virtual conferencing was presented. A virtual teleconferencing system employing a real-time auralisation engine was used, where participant pairs tasked with having a conversation could either have identical or mismatched acoustics.

A preliminary analysis of subjective ratings and high-level objective parameters suggests that the system was ecologically valid, as trends observed in real-life room acoustics, such as an increase in voice level in less reverberant conditions, were also observed in the system. Overall, the results indicate that one's own acoustics impact how the conversation is perceived, but no significant impact of the interlocutor's acoustics was observed. Further analyses must be conducted in order to fully understand how both acoustics impacted the conversational dynamics.

6. FUTURE WORK

Future work will focus on continuing the analysis of other objective parameters, such as speech rate, state probabilities, voicing period duration, or speaker alternation rate, to name a few. These analyses should allow a better-informed judgement on how the acoustic conditions impacted the conversations had by participants, but also on how ecologically valid the system was. Moreover, further subjective data was collected in the form of open comment questions and post-experiment informal interviews, and this data will also be analysed.

It could also be of interest to run a next iteration of the experiment, with an improved system based on the results of the analysis, and where more factors could be tested, such as interactions between more than two people, or testing different acoustic conditions that are less extreme and more similar to real-life scenarios that could present themselves in AR teleconferencing.

7. ACKNOWLEDGEMENTS

We would like to thank the participants who took part in the listening experiment. This work was carried out in the context of the SONICOM project (`www.sonicom.eu`) that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101017743.

8. REFERENCES

[1] A. Neidhardt, C. Schneiderwind, and F. Klein, "Perceptual Matching of Room Acoustics for Auditory Augmented Reality in Small Rooms - Literature Review and Theoretical Framework," *Trends in Hearing*, vol. 26, Jan. 2022.

[2] J. W. Black, "The Effect of Room Characteristics upon Vocal Intensity and Rate," *J. of the Acoust. Soc. of America*, vol. 22, pp. 174–176, Mar. 1950.

[3] D. Pelegrín-García, B. Smits, J. Brunskog, and C.-H. Jeong, "Vocal effort with changing talker-to-listener distance in different acoustic environments," *J. of the Acoust. Soc. of America*, vol. 129, pp. 1981–1990, Apr. 2011.

[4] A. Astolfi, A. Castellana, G. E. Puglisi, U. Fugiglando, and A. Carullo, "Speech level parameters in very low and excessive reverberation measured with a contact-sensor-based device and a headworn microphone," *J. of the Acoust. Soc. of America*, vol. 145, pp. 2540–2551, Apr. 2019.

[5] P. Bottalico, "Speech Adjustments for Room Acoustics and Their Effects on Vocal Effort," *J. of Voice*, vol. 31, pp. 392.e1–392.e12, May 2017.

[6] A. Astolfi, A. Carullo, L. Pavese, and G. E. Puglisi, "Duration of voicing and silence periods of continuous speech in different acoustic environments," *J. of the Acoust. Soc. of America*, vol. 137, pp. 565–579, Feb. 2015.

[7] J. Rennies, H. Schepker, I. Holube, and B. Kollmeier, "Listening effort and speech intelligibility in listening situations affected by noise and reverberation," *J. of the Acoust. Soc. of America*, vol. 136, pp. 2642–2653, 11 2014.

[8] T. Carpentier, "A new implementation of Spat in Max," in *15th Sound and Music Computing Conference (SMC2018)*, (Limassol, Cyprus), pp. 184–191, July 2018.