EUROREGIO
BNAM2022
Joint Acoustics Conference
9th-11th of May, Aalborg, Denmark

Perception of Virtual Reality Based Audiovisual Paradigm for People with Hearing Impairment

Kang Sun[1,2,*], Niels H. Pontoppidan[1], Dorothea Wendt[1,3], Lars Bramsløw[1]

[1]Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark.

[2]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark.

[3]Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark.

*knsu@eriksholm.com

Abstract
Integrating new and emerging technologies, such as virtual reality (VR), with established test methods to improve the ecological validity is gradually used by practitioners in some fields (e.g., soundscape research) but remains limited use in hearing science. In this paper, an audiovisual setup was introduced to create an augmented speech-in-noise test for people with hearing impairment (HI). The environment containing four competing talkers was recorded with 360-degree video and ambisonics audio. In the scene recreating the environment, the video was displayed in the VR headset, while the audio was presented to the participants via a circular loudspeaker array. Furthermore, a (silent and fixed) physical avatar (manikin) was included in the video recording as a placeholder for the audio stream of the target speech, which was added into sound presented over the loudspeaker array. This setup was used to test the effect of different hearing aid (HA) settings under the same condition for each participant. In total, 27 HI participants were tested. VR has been known to cause motion discomfort, which is referred as VR sickness or cybersickness nowadays. The simulator sickness questionnaire (SSQ, [11]) was used to quantify the sickness measurement. The data showed a low degree of Nausea and Disorientation, but scattered Oculomotor responses. Moreover, a general questionnaire assessing scene recreation, test method and outcome expectation was administrated. In general, the audiovisual system received high appraisal in realism; the augmented speech-in-noise test method was well accepted; participants highly agreed that the difference between programs could be distinguished. However, the sense of physical immersion decreased due to the weak binding between the avatar and the target speech. Furthermore, when comparing the three components (Nausea, Disorientation, Oculomotor) in SSQ and items in the general questionnaire, the Oculomotor was found significantly correlated to the perceived binding of target speech and the avatar. Specifically, participants who were less convinced that the speech came from the avatar, also rated the Oculomotor higher.

Keywords: virtual reality, hearing impairment, perception, audiovisual

1 Introduction

Conventionally, research fields involve human perception of sounds, such as soundscape, adopt laboratory tests in audio-only presentation. However, driven by the notion that many established methodologies in laboratory lack sufficient realism to produce adequately findings in real life, adopting a holistic manner in delivering multisensory information in laboratory studies to improve their ecological validity gets increasingly popular [1]. For example, researchers in soundscape address the role of visual component via the audiovisual interaction (e.g., [2]) in recent years. Until now, audio-only presentation has dominated in laboratory tests in hearing science. Though being a relatively new concept, ecological validity, in hearing science, refers to the degree to which research findings reflect real-life hearing-related function, activity, or participation [3].

New and emerging technologies – virtual reality (VR) – was integrated with established test method and used in soundscape studies (e.g., [4,5]). The technology is often well-perceived as it increases the immersion of human participation. In hearing science, effort has been taken recently to use VR as a part of rehabilitation tool for young persons with hearing difficulty [6,7], with a particular focus on stimulated environments [8]. However, the reproduction of real-life recorded environments in lab were not particularly interested due to the increased difficulty of augmentation to meet the study designs. VR has been known to cause motion discomfort, which is referred as VR sickness or cybersickness nowadays. One of the major concerns of applying VR on people with hearing impairment (HI) lies on that older (65+ yrs) adults reported more sickness than younger adults in simulated environments [9]. Furthermore, hearing loss was found associated with poorer spatial awareness [10], which might bring further discomfort in VR environments. Nevertheless, limited attention was paid on investigating the perception of HI on the VR technology, particularly in recorded real-life environments reproduction.

In this paper, we present a VR based audiovisual paradigm with augmented real-life recording, in which HI could perform a speech-in-noise test in a close-to-real-life ambience. To obtain a sketch of the paradigm, participants were recruited and instructed to evaluate a set of hearing aid (HA) programs. We used a well-known stimulator sickness questionnaire (SSQ, [11]) to evaluate the sickness symptoms of HI experiencing our setup. Furthermore, a comprehensive questionnaire was used to assess the reproduction quality, test method and outcome expectation.

2 Method

2.1 Audiovisual scene recording

The initial audiovisual scene recording was administrated in a canteen. Four actors (two male and two female) were recruited to sit in prescribed positions to engage a paired conversation (i.e., the two actors on the same side to the avatar had a conversation, Figure 1). The actors were provided a list of topics for non-invasive daily conversation such as hobby, vacation, food. The actors were also instructed to maintain the consistency of the conversation, that is to avoid rapid volume change or silence from both pairs. Furthermore, an avatar, represented by a B&K 4128 head and torso simulator (HATS, Brüel & Kjær, Denmark), was placed in the recording as a placeholder for the target speech in the lab reproduction. The HATS manikin only served as an avatar and had no acoustic purpose. All actors and the avatar were positioned 1.4m from the recording position, as shown in Figure 1.

At the recording position, a GoPro Fusion 360 camera was used for 360-degree video (5.2k, 4992 x 2496 resolution, 30 fps) and synchronized 4-channel first-order (B-format) ambisonics recording. The camera was set at the eye level of the avatar, and both the camera and the avatar were aligned to the centre of the table. Moreover, an additional microphone linking to a B&K 2250 sound level meter (Brüel & Kjær, Denmark) was hanged on the ceiling right above the recording position. The total duration of the audiovisual recording was 60 min. The sound level meter performed 5-min segment recordings continually and collected 12 recordings.

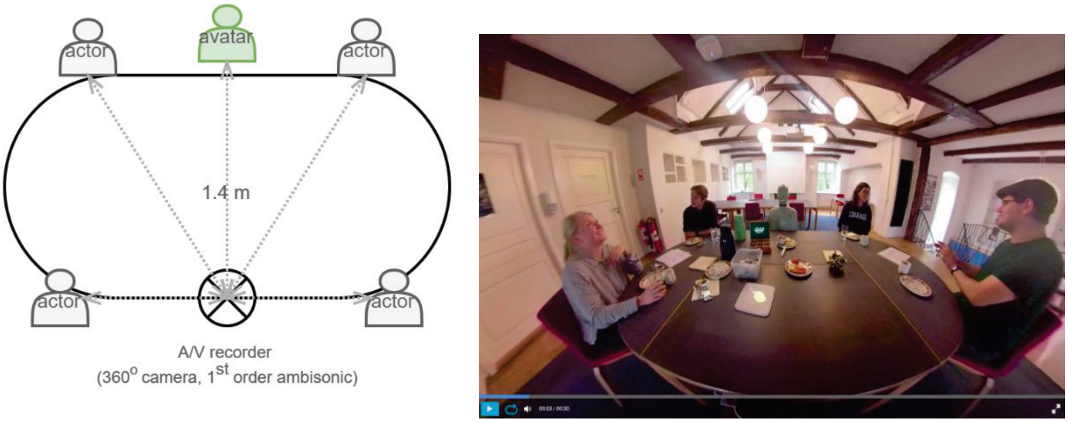

Figure 1 – Sketch of the audiovisual scene recording. (Left: top view of the relative positions of the actors, the avatar and the A/V recorder, all actors and the avatar were 1.4m from the A/V recorder, sketch is not true to scale; Right: the recorded view from the point of the A/V recorder, i.e., the participant view in the test.)

2.2 Audiovisual scene reproduction and augmentation

The reproduction of the audiovisual scene was administrated in an anechoic room. This study used Oculus Go, a stand-alone device, as the head-mounted display (HMD) for visual recreation. Though Oculus Go provided integrated headphones and even possibility to customize headphones, this option was omitted considering the test participants in this study all wore a pair of HAs. Instead, a loudspeaker array containing 24 loudspeakers forming a circle of 1.4m radius was used for sound reproduction. All loudspeakers were horizontally aligned and evenly distributed at an interval of 15 degrees, as shown in Figure 2.

Recorded footage were processed with GoPro Fusion studio. As a raw rendering, the audiovisual document was saved in a format of H.264 video codec, 5.2k resolution, and 360 audio (ambix) from Fusion Studio. This rendering was then split into a 4-channel B-formatted audio in ambix convention and a muted video document (no audio). A decoding procedure using Matlab AmbiToolbox transformed 4-channel audio into 24-channel audio while maintaining spatial characteristics. The decoding exported audio matching the loudspeaker positions as per se, and reconstructed the sound field at the centre of the array. The 24-channel audio was played via Reaper (v6.13, suggested by [12]) in this study. As the audio was presented via the loudspeakers and the video via the HMD, a time alignment procedure was implemented. First, a single channel linear time sync audio (LTC) was merged with the muted video. The merged document was saved in the HMD and played via the Skybox VR player. Then while the LTC-merged video was played, the coded signal LTC indicated a LTC reader software (Reaper) the timestamp of the audio (via an audio cable, indicated in Figure 2-Right), and thus synchronized the decoded 24-channel audio. The playback was calibrated at the array centre (i.e., participant's seat), using the sound level meter recordings as reference. To maintain the equivalence of sound level across the whole experiment, the overall average sound level was adjusted (the adjustment was limited) and fixed at 65 dB ($L_{eq,5min}$ continually throughout the 60 min recording).

Since the essential form of this study was a speech-in-noise test, the audiovisual scene augmentation used the standard speech material archive Danish HINT (hearing in noise test, [13]) sentences as target speech and the recorded scene as the mask (i.e., noise). A separate HINT audio was merged to the front loudspeaker relative to the participant's seat (the loudspeaker marked as T in Figure 2-Left, overlapped with the avatar). During the test, it was presented as that the actors were having conversations as the mask, while the avatar was delivery HINT sentences, from the participant's view. The HINT audio was also calibrated at the array centre, and a graphic user interface (developed in Matlab) was provided to the experimenter to manually adjust the speech level, and hence signal-noise ratio (SNR), during the test.

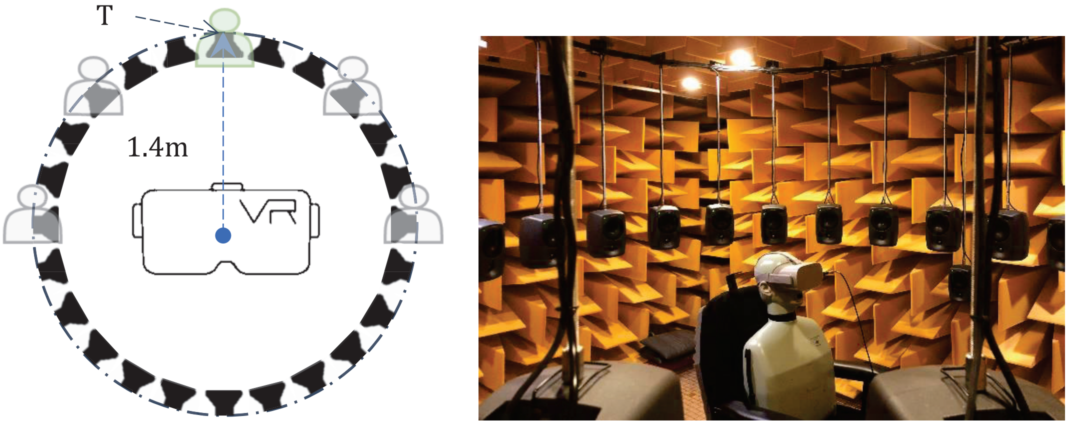

Figure 2 – Sketch of the recreation system. (Left: top view of the loudspeaker array of 24 loudspeakers in a circle of 1.4m radius, the VR position indicates the seat for the participants, the shaded portrait indicates the relative positions of the recorded avatar and actors reflected in the loudspeaker array; Right: an illustration of a test participant during the test, an audio cable was connected to the VR headset, which sent timestamp to Reaper for synchronization.)

2.3 Questionnaire

The simulator sickness questionnaire (SSQ) from [11] was used. SSQ contained 16 questions and derived into three categories: Nausea, Oculomotor, and Disorientation. The score scale, the calculation means, and the coefficients were true to [11] (details attached in Appendix). Moreover, a general questionnaire containing four aspects was designed and administered. As shown in Table 1, Q1-Q3 described the audiovisual scene recreation from perspectives including realism (Q3), immersion (Q2) and the binding between the speech and the avatar (Q1). Q2 and Q3 were adopted from [4,14]. Q4 and Q5 concerned the test method focusing on the perception of the augmentation (Q4) and the test objective (Q5). Q4 was adopted from [15]. Q6-Q8 expressed the outcome expectation and the repeatability of the test. Lastly, Q9 and Q10 asked the representativity of the reproduced scenario (Q9) and the general attitude in coping with it (Q10). All questions were answered in a 5-point likert-scale (from Strongly Disagree to Strongly Agree), started by asking "To which degree you agree with the following statement".

Table 1: Overview of the general questionnaire.

Category	No.	Questionnaire
Scene recreation	1	I sense the speech is from the avatar.
	2	I feel physically immersed in the presented scene.
	3	I feel the presented scene is real.
Test method	4	I can accept to talk to a chatbot like the avatar.
	5	I sense the difference of different programs in hearing during the experiment.
Outcome expectation	6	I believe this format of experiment will bring me better hearing aid fitting.
	7	I can describe my hearing experience to an audiologist better in this test format.
	8	I wish to experience more of this type of test scenarios.
Others	9	I encounter the type of listening environment in this experiment a lot in life.
	10	I try to avoid the presented situation in this experiment in my life.

2.4 Participants and Procedure

The protocol of the study was approved by the Capital Region of Danish Committee System on Health Research Ethics (Hovedstaden, reg. nr.: H-20068237). Twenty-seven HI (10 female, 17 male; mean_{age}=72, $\text{median}_{\text{age}}$=74, SD_{age}=8.6, age range: 51-87 yrs.) with mild to moderate hearing loss were recruited as test participants. All participants were provided with a consent form before the experiment and a signature was obtained from each participant.

The participants were instructed to evaluate a set of HA programs under the same condition. As stated before, the mask (noise) was set to 65 dB. First, in the training phase, an adaptive procedure was applied to estimate the signal-to-noise ratio (SNR), at which participants reached 50% of speech intelligibility (SNR50) with the same reference program of the HAs. Afterwards, in the test phase, each participant was tested at fixed SNR, i.e. their individual SNR50, and thus a fixed sound level of the target speech (HINT). Across all participants, the average SNR was -0.19 dB (SD=2.33, range: -5~+5). During the test, the participants sat comfortably in the chair in the centre of the loudspeaker array. The chair was height adjustable so that the loudspeaker array and the ears of the participants were set at the same level. As the sanity check of the spatial alignment of the VR view and the loudspeakers, the participants were asked to point out the avatar in the VR view at the beginning of wearing the HMD, before the loudspeakers were activated (i.e., in silence). The physical pointing position was required to be at the front loudspeaker (T). After all the test, the questionnaires were presented to the participants via a tablet.

3 Results and Analysis

3.1 Perception of the audiovisual paradigm

In [11], three principal factors were extracted from 16 symptoms in the SSQ, namely Nausea (nausea, stomach awareness, increased salivation, burping), Oculomotor (eyestrain, difficulty focusing, blurred vision, headache), and Disorientation (dizziness, vertigo). Detailed list is attached in the Appendix. Each of the three factors was used as the basis for an SSQ subscale, and together they formed an SSQ profile for an assessed item. The proportional result of the SSQ profile of the audiovisual scene was shown in Figure 3. The mean proportional score in all factors was lowest in Nausea (mean=6.70%, SD=6.12%), followed by Disorientation (mean=6.88%, SD=9.55%) and highest in Oculomotor (mean=15.87%, SD=12.30%). In similar studies using HMD and adopting the SSQ, [16] involved younger normal hearing (NH) participants to perform a visual task, which suggested a comparable Nausea (mean proportional score 18%) response but a much higher Oculomotor (39%) and Disorientation (42%) response than the current study. [17] also involved younger NH participants to experience the VR scene with a visual discrimination task, which reported no absolute score in any SSQ factors but a deviation under different test conditions. However, six out of twenty participants withdrew the test, which indicated a high degree of discomfort.

In this study, the older HI shown limited symptoms of Nausea and Disorientation after the audiovisual paradigm, which was comparable to (and even better than) the younger NH peers in other studies. Earlier research [9,18] suggested more sickness expected in older participants in complex environments. However, a correlation analysis on age and the SSQ items (including three factors and the total score) in the current study shown no significant correlations (all p>.05). The current study used a static recording of an indoor, familiar, and real-life environment [19], which could contribute to a low degree of cybersickness.

Nevertheless, a sphere space was presented to the participants and 26 (out of 27) participants experienced a pair of VR glasses for the first time. Adaptation was expected, which might explain the slight increase in Oculomotor response (compared to Nausea and Disorientation). Furthermore, a hard task was involved in the test, which might increase the Oculomotor response, despite it being an audio task. Unlike visual tasks in which participants paid high attention to the visual presentations in VR (which might evoke considerable head rotation and body turn), an audio task did not require the participants to observe the visual in detail, which might explain the relatively lower SSQ response in this study than in [16] or potentially in [17]. In addition, the high standard deviation (SD) in Oculomotor (and Disorientation) response indicated a (relatively) large individual difference in this study. Nevertheless, the low SSQ profile in this study suggested that the audiovisual paradigm was not invasive and cybersickness friendly to HI. In fact, after the test, no participant reported any discomfort at any given time.

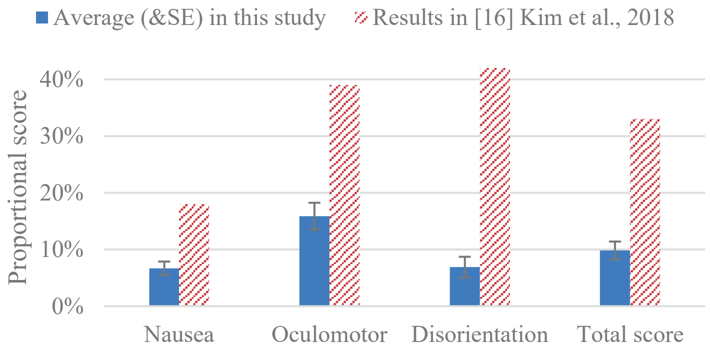

Figure 3 – SSQ profile (average score and standard error) of the audiovisual system in the test (All scores were converted to proportions to the full scale of each SSQ factors, and therefore the scale in this figure was unified. The corresponding interpretation of the degree of the symptoms: 0% – not at all, 33.33% – slightly, 66.67% – moderately, 100% – very. The slashed bars were results adopted from [16]).

Furthermore, Figure 4 illustrated the response of the general questionnaire. The participants perceived the reproduction as real (Q3, 81.47% agree or strongly agree). However, this appreciation decreased in immersion (Q2) and the binding between the avatar and the target speech (Q1). Particularly, 37.04% of participants were negative in answering Q1, while 33.33% of participants answered neutral in Q2. The results of immersion and realism were comparable to [14], which also used real-life recording reproduction but with no augmentation. Even though in this study an augmentation was attempted, the avatar still lacked features related to the audio and visual binding, such as lip movement, facial expression, and body language.

Moreover, the objective of the test was clear to the participants (Q5, 88.88% agree or strongly agree), suggesting such a paradigm was suitable for a speech-in-noise test format. A reasonably high degree of accepting the intention of augmentation (Q4, 66.66% agree or strongly agree) was achieved. Combining results in Q1, the participants understood and accepted the augmentation attempt, however, required a better tailored implementation. In addition, the repeatability of the test was high (Q8, 85.19% agree or strongly agree). The participants expected better hearing aid fitting after such type of test (Q6, 66.67% agree or strongly agree). However, describing their hearing difficulties remained a problem (Q7, 66.67% neither, disagree or strongly disagree), despite the improvement of ecological validity in this study. Lastly, though nearly 60% participants agreed that they met challenging scenarios as such in their lives (Q9). However, the attitude in coping with the challenge was almost evenly distributed, that is 44.44% of the participants wanted to face this type of scenarios while 40.74% decided to avoid them.

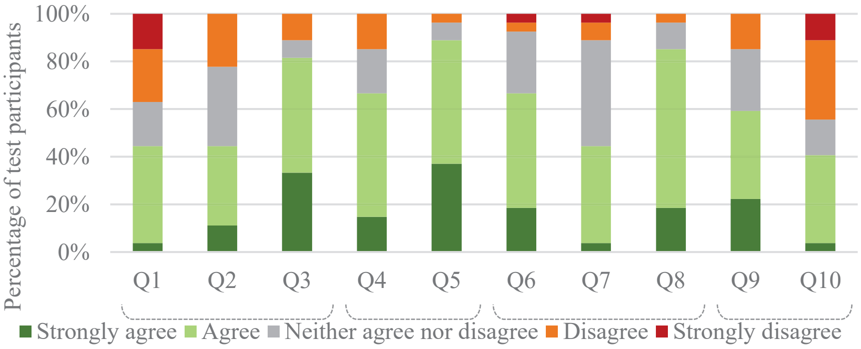

Figure 4 – The answer distribution of the questions in the general questionnaire. (Q1-Q10 refer to the question numbers in Table 1; parentheses group questions into four categories according to Table 1.)

The response of the general questionnaire was transferred into a scale comparable to the SSQ, in which “Strongly Disagree” was set to 0 and “Strongly Agree” was set to 4. Then a Person’s correlation test was applied among the SSQ factors and questionnaire items (Table 2).

On one hand, all factors in the SSQ were significantly correlated with each other, which suggested that participants who rated higher score in one factor, also rated higher in the other two factors and therefore in Total score. That is, despite the three SSQ factors being distinct [11], participants had stronger cybersickness felt discomfort from all dimensions. A significant individual difference in VR sickness measurement could be expected. On the other hand, among the items in the general questionnaire, Q3 was significantly correlated to Q6, Q8, and Q9. It indicated that the more real the participants found the reproduction (Q3), the better outcome from the test (Q6) they expected, the better willingness to conduct more of such tests (Q8) they shown, and the more of similar environments in own lives as what shown in the reproduction (Q9) they experienced. Furthermore, Q4 was highly correlated to Q6 and Q7. Namely that the participants who had a good acceptance of the virtualized scene, also expected more positive outcomes from such test. Moreover, the three questions (Q6-Q8) in outcome expectation were all significantly correlated, which suggested that the refreshing technology and yet narrative context in this study could stimulate higher outcome expectation, better difficulty description and higher willingness of repeatability, at least for some participants. Lastly, Q6 was significantly correlated to Q10, which indicated that the participants who expected better HA fitting after such test also were more willingly to take the challenge of similar difficult communication environments in their lives.

Between the SSQ factors and the general questionnaire, Q1 was found significantly correlated to Oculomotor and Total score in the SSQ. That is, the participants who were less convinced of the binding between the avatar and the target speech, also rated the Oculomotor and the Total score higher. One potential explanation might be that the participants who were less convinced by the A/V augmentation, would visually focus less on the avatar but more on the ambience in VR, and thus triggered a higher Oculomotor response. However, this test lacked recordings of the head movement of the participants, this interpretation thus cannot be examined.

Table 2: Person's correlations between SSQ items and questionnaire items.

		SSQ				Scene recreation			Test method		Outcome expectation			Others	
	Var.	N^1	O^2	D^3	TS^4	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
	N^1	—													
	O^2	0.696 ***	—												
	D^3	0.462 *	0.736 ***	—											
	TS^4	0.772 ***	0.954 ***	0.867 ***	—										
Scene recreation	Q1	-0.187	-0.523 **	-0.261	-0.407 *	—									
	Q2	0.041	0.015	-0.176	-0.05	-0.157	—								
	Q3	0.019	0.073	0.171	0.107	-0.102	0.114	—							
Test method	Q4	-0.075	-0.286	-0.164	-0.224	0.304	-0.044	0.371	—						
	Q5	0.139	0.156	0.033	0.125	-0.291	0.213	0.26	0.167	—					
Outcome expectation	Q6	-0.067	-0.212	0.042	-0.105	0.23	-0.028	0.445 *	0.472 *	0.247	—				
	Q7	-0.059	-0.316	-0.294	-0.285	0.323	3.130 e -20	0.131	0.402 *	0.123	0.702 ***	—			
	Q8	-0.216	-0.237	-0.028	-0.182	0.19	0.118	0.422 *	0.615	0.151	0.719 ***	0.476 *	—		
Others	Q9	-0.333	-0.278	-0.056	-0.242	0.054	0.28	0.423 *	0.334	0.154	0.19	0.185	0.17	—	
	Q10	0.235	0.253	0.185	0.255	0.053	-0.035	0.004	-0.145	0.03	-0.38*	-0.36	-0.294	-0.033	—

N^1: Nausea, O^2: Oculomotor, D^3: Disorientation, TS^4: Total score;
***: p<.001, **: p<.01, *: p<.05;
Q1-Q10 refer to the question numbers in Table 1.

3.2 Limitation

In this study, the major concern from the participants' feedback of this audiovisual paradigm was the binding between the fixed avatar and the target speech, i.e., the implementation of the augmentation. The visual placeholder (avatar) lacking convincing features such as lip movement and facial expression might contribute to the weak binding perception. Furthermore, the room acoustic feature (e.g., reverberation time) was not applied for the target speech material, which could introduce a decrease of realism and immersion for the participants. Last but not the least, the purpose of this paradigm was to integrate VR technology with established test method to improve the ecological validity of the test. In this test, speech intelligibility was assessed via word score of the target speech, which was far from daily conversations. Meanwhile the ambience and the noise mask were a real-life recording. This controversial combination might prohibit the participants revivificating their hearing conditions in real life.

4 Conclusion

This paper reported an audiovisual paradigm, in which an augmented real-life recording was presented to perform a speech-in-noise test. The cybersickness perception of this paradigm was measured by the SSQ and a general questionnaire. The result indicated that it was a non-invasive and cybersickness friendly setup for HI, who experienced high immersion and shared good appreciation in the study. Furthermore, it provided an important enhancement for realistic and ecological HA assessment in laboratory.

Acknowledgements

This research received funding from Innovation Foundation Denmark and William Demant Foundation. These organizations are gratefully acknowledged. The authors would like to thank all the HI participating in this study, and the four actors in the recording. The authors would also like to thank Sergio Aguirre for helping the experiment setup.

References

[1] Aletta, F., Kang, J. and Axelsson, Ö., 2016. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. Landscape and Urban Planning, 149, pp.65-74. doi: 10.1016/j.landurbplan.2016.02.001.

[2] Preis, A., Kociński, J., Hafke-Dys, H. and Wrzosek, M., 2015. Audio-visual interactions in environment assessment. Science of the Total Environment, 523, pp.191-200. doi: 10.1016/j.scitotenv.2015.03.128.

[3] Keidser, G., Naylor, G., Brungart, D.S., Caduff, A., Campos, J., Carlile, S., Carpenter, M.G., Grimm, G., Hohmann, V., Holube, I. and Launer, S., 2020. The quest for ecological validity in hearing science: What it is, why it matters, and how to advance it. Ear and hearing, 41(Suppl 1), p.5S. doi: 10.1097/AUD.0000000000000944.

[4] Sanchez, G.M.E., Van Renterghem, T., Sun, K., De Coensel, B. and Botteldooren, D., 2017. Using Virtual Reality for assessing the role of noise in the audio-visual design of an urban public space. Landscape and Urban Planning, 167, pp.98-107. doi: 10.1016/j.landurbplan.2017.05.018.

[5] Sun, K., De Coensel, B., Filipan, K., Aletta, F., Van Renterghem, T., De Pessemier, T., Joseph, W. and Botteldooren, D., 2019. Classification of soundscapes of urban public open spaces. Landscape and urban planning, 189, pp.139-155. doi: 10.1016/j.landurbplan.2019.04.016.

[6] Zirzow, N.K., 2015. Signing avatars: Using virtual reality to support students with hearing loss. Rural Special Education Quarterly, 34(3), pp.33-36. doi: 10.1177/875687051503400307.

[7] Lau, S.T., Pichora-Fuller, M.K., Li, K.Z., Singh, G. and Campos, J.L., 2016. Effects of hearing loss on dual-task performance in an audiovisual virtual reality simulation of listening while walking. Journal of the American Academy of Audiology, 27(07), pp.567-587. doi: 10.3766/jaaa.15115.

[8] van de Par, S., Ewert, S.D., Hladek, L., Kirsch, C., Schütze, J., Llorca-Bofí, J., Grimm, G., Hendrikse, M.M., Kollmeier, B. and Seeber, B.U., 2021. Auditory-visual scenes for hearing research. arXiv preprint arXiv:2 doi: 10.48550/arXiv.2111.01237. 111.01237.

[9] Keshavarz, B., Ramkhalawansingh, R., Haycock, B., Shahab, S. and Campos, J.L., 2018. Comparing simulator sickness in younger and older adults during simulated driving under different multisensory conditions. Transportation research part F: traffic psychology and behaviour, 54, pp.47-62. doi: 10.1016/j.trf.2018.01.007.

[10] Jiam, N.T.L., Li, C. and Agrawal, Y., 2016. Hearing loss and falls: A systematic review and meta-analysis. The Laryngoscope, 126(11), pp.2587-2596. doi: 10.1002/lary.25927.

[11] Kennedy, R.S., Lane, N.E., Berbaum, K.S. and Lilienthal, M.G. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. The international journal of aviation psychology, 3(3), pp.203-220. doi: 10.1207/s15327108ijap0303_3.

[12] Spatial workstation. [Online]. Available: https://facebook360.fb.com/spatial-workstation/.

[13] Nielsen, J.B. and Dau, T., 2011. The Danish hearing in noise test. International journal of audiology, 50(3), pp.202-208. doi: 10.3109/14992027.2010.524254.

[14] Sun, K., Botteldooren, D. and De Coensel, B., 2018, December. Realism and immersion in the reproduction of audio-visual recordings for urban soundscape evaluation. In INTER-NOISE and NOISE-CON Congress and Conference Proceedings (Vol. 258, No. 4, pp. 3432-3441). Institute of Noise Control Engineering.

[15] Zamora, J., 2017, October. I'm sorry, Dave, I'm afraid I can't do that: Chatbot perception and expectations. In Proceedings of the 5th international conference on human agent interaction (pp. 253-260). doi: 10.1145/3125739.3125766.

[16] Kim, H.K., Park, J., Choi, Y. and Choe, M. 2018. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. Applied ergonomics, 69, pp.66-73. doi: 10.1016/j.apergo.2017.12.016.

[17] Carnegie, K. and Rhee, T., 2015. Reducing visual discomfort with HMDs using dynamic depth of field. IEEE computer graphics and applications, 35(5), pp.34-41. doi: 10.1109/MCG.2015.98.

[18] Jang, D.P., Kim, I.Y., Nam, S.W., Wiederhold, B.K., Wiederhold, M.D. and Kim, S.I., 2002. Analysis of physiological response to two virtual environments: driving and flying simulation. CyberPsychology & Behavior, 5(1), pp.11-18. doi: 10.1089/109493102753685845.

[19] Chang, E., Kim, H.T. and Yoo, B., 2020. Virtual reality sickness: a review of causes and measurements. International Journal of Human–Computer Interaction, 36(17), pp.1658-1682. doi: 10.1080/10447318.2020.1778351.

Appendix

Table A summarized all items included in the SSQ test. All symptoms were followed by a 4-point likert-scale from 0 (not at all) to 3 (very) to indicate the degree of each symptom. The three major components (Nausea, Oculomotor and Disorientation) included a different set of symptoms in the list. To calculate the three components, the sum was first calculated by adding symptom scores and then multiplied by a coefficient.

Table A: Computation of SSQ scores (adopted from [11]).

SSQ symptom[a]	**Nausea**	**Oculomotor**	**Disorientation**
1. General discomfort	|	|	
2. Fatigue		|	
3. Headache		|	
4. Eyestrain		|	
5. Difficulty focusing		|	|
6. increased salivation	|		
7. Sweating	|		
8. Nausea	|		|
9. Difficulty concentrating	|	|	
10. Fullness of head			|
11. Blurred vision		|	|
12. Dizzy (eye open)			|
13. Dizzy (eye closed)			|
14. Vertigo			|
15. Stomach awareness	|		
16. Burping	|		
Total[b]	[1]	[2]	[3]

Score
Nausea = [1] * 9.54
Oculomotor = [2] * 7.58
Disorientation = [3] * 13.92
Total score = ([1]+[2]+[3]) * 3.74			

[a]scored (label): 0 (not at all), 1 (slightly), 2 (moderately), 3 (very);
[b]sum obtained by adding symptom scores.

Figure A- Left reported the boxplot of the SSQ profile (in original score) of the system in the test. The average score in all factors was lowest in Nausea (mean=13.43, SD=12.25), followed by Disorientation (mean=20.11, SD=27.93) and highest in Oculomotor (mean=25.27, SD=19.57). However, due to the difference between the coefficient, all factors in SSQ had a different scale. Figure A-Right further explained the correspondingly scale in each factor with markers at “not at all”, “slightly”, “moderately” and “very”. Furthermore, the average scores for each factor were also plotted (marked as “average” in Figure A-Right).

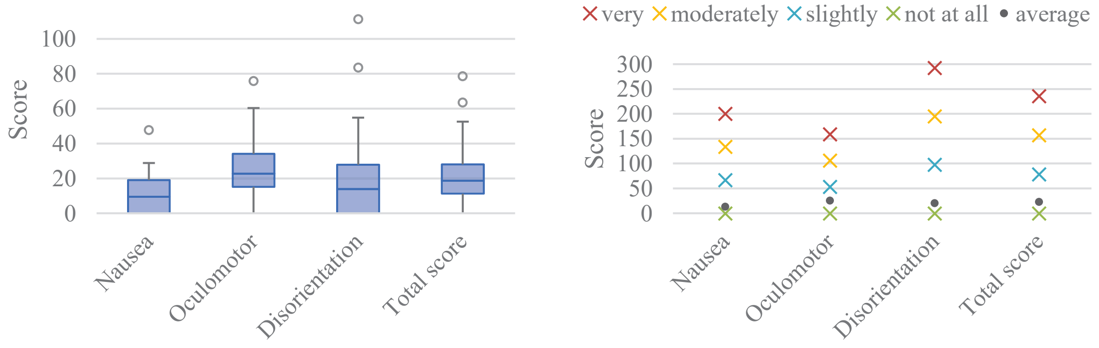

Figure A – Left: SSQ profile of the audiovisual system in the test (original score); Right: overview of SSQ subscale and corresponding verbal ticks.