DOI: 10.61782/fa.2023.0320

forum acusticum 2023

DOES A TALKER'S VOICE QUALITY AFFECT UNIVERSITY STUDENTS' LISTENING EFFORT IN A VIRTUAL SEMINAR ROOM?

Isabel S. Schiller[1*] **Andrea Bönsch**[2] **Jonathan Ehret**[2]
Carolin Breuer[3] **Lukas Aspöck**[3]
[1] Work and Engineering Psychology, RWTH Aachen University, Germany
[2] Visual Computing Institute, RWTH Aachen University, Germany
[3] Institute for Hearing Technology and Acoustics, RWTH Aachen, Germany

ABSTRACT

A university professor's voice quality can either facilitate or impede effective listening in students. In this study, we investigated the effect of hoarseness on university students' listening effort in seminar rooms using audio-visual virtual reality (VR). During the experiment, participants were immersed in a virtual seminar room with typical background sounds and performed a dual-task paradigm involving listening to and answering questions about short stories, narrated by a female virtual professor, while responding to tactile vibration patterns. In a within-subject design, the professor's voice quality was varied between normal and hoarse. Listening effort was assessed based on performance and response time measures in the dual-task paradigm and participants' subjective evaluation. It was hypothesized that listening to a hoarse voice leads to higher listening effort. While the analysis is still ongoing, our preliminary results show that listening to the hoarse voice significantly increased perceived listening effort. In contrast, the effect of voice quality was not significant in the dual-task paradigm. These findings indicate that, even if students' performance remains unchanged, listening to hoarse university professors may still require more effort.

Keywords: *listening effort, voice disorder, virtual reality, virtual acoustics*

**Corresponding author: isabel.schiller@psych.rwth-aachen.de*

1. INTRODUCTION

The exertion of mental energy to overcome obstacles while performing a listening task is known as listening effort [1]. For university students following a lecture, such obstacles may, among other factors, relate to background sounds and the professor's voice quality. Research has shown that university professors are more prone to developing voice disorders, with a prevalence of up to 41 % compared to 6 % in the general population [2, 3]. If students are exposed to a professor's hoarse (dysphonic) voice during lectures, this may increase their listening effort. The effect of a talker's voice quality on listening effort in school-aged children has been thoroughly investigated (for a recent meta-analysis, see [4]); however, consequences for young adult listeners are still underdetermined.

Listening effort is often assessed with dual-task paradigms (DTP), which inherits the notion that cognitive capacity is limited and can be intentionally allocated between tasks [5]. In DTP, participants perform two unrelated tasks in parallel, with the primary task focusing on listening, such as spoken language comprehension, and the secondary task typically being in a different modality (e.g., performing simple motor actions in response to tactile cues). Reduced performance and/or longer response times in the secondary task are interpreted as an indication of increased listening effort in the given listening condition, assuming that more cognitive resources were needed to complete the primary task, leaving fewer resources for the secondary task.

To our knowledge, only Imhof et al. [6] have hitherto studied the impact of a talker's abnormal voice quality on university students' listening effort. In a laboratory listening experiment, the authors used a DTP as well as a subjective rating of listening effort to explore the effect of normal versus creaky voice. In comparison to a hoarse voice, which is rather rough and raspy, creaky voice (vocal fry) is characterized by a low-pitched, popping or crackling sound. Regarding the DTP, it was found that listener performance was significantly poorer on the secondary task when listening to the creaky voice, suggesting a higher listening effort. Subjective listening effort in that condition was also increased. A limitation of Imhof et al.'s [6] study is that that it only considered the auditory component of language comprehension. However, in most listening situations at universities and generally in education, spoken language comprehension is an audio-visual process.

The present study therefore combines approaches from psychology, acoustics, and virtual reality (VR) to investigate the impact of hoarseness on university students' listening effort in an immersive, close-to-real-life seminar room setting. Our hypothesis is that listening to a hoarse voice will result in decreased performance and/or longer response times in the secondary task of a DTP, indicating higher listening effort, and will also be perceived as more effortful to listen to.

2. METHOD

We conducted an audio-visual VR experiment, in a soundproof booth at the Institute of Psychology at RWTH Aachen University. The experiment lasted 60-75 minutes, from which approximately 45 minutes were spent immersed. The study was approved by the ethics committee of the Faculty of Arts and Humanities (ref. 2022_016_FB7_RWTH Aachen).

2.1 Participants

Fifty-seven participants took part in the study. However, data analysis is still in progress. This paper presents the preliminary results, based on a set of 40 university students (25 female, 15 male), aged 18-37 years ($M = 25$, $SD = 5$). Inclusion criteria were (1) native speaker of German or comparable level, and (2) normal hearing ($\leq$ 20 dB HL) according to pure-tone audiometry between 500 and 4000 Hz (ear3.0 audiometer, Auritec), and (3) normal or corrected-to-normal vision (self-report).

2.2 Task

We used a within-subject design to investigate the influence of a hoarse voice compared to a normal voice on listening effort. Both voices were presented by the same female character, an embodied conversational agent, representing a university professor. Participants completed a DTP, seated in a virtual seminar room populated with six static wooden mannequins representing fellow students (Figure 1). The primary (listening) task was Heard Text Recall [7], which involved listening to short texts (~60 s each) describing family constellations. After each text, participants answered nine content questions. The secondary (vibrotactile) task involved reacting to vibration patterns presented via two hand-held controllers. Voice quality (normal vs. hoarse) was randomly assigned during the experiment. Listening effort for each voice quality was quantified based on performance and response time measures in the secondary task. Only response times from correct trials were analyzed. In addition, we asked the participants to rate their perceived level of listening effort on a scale from 0 (*not at all effortful*) to 5 (*extremely effortful*) for each voice quality.

Figure 1. Participant immersed in the virtual seminar room with the virtual professor up front.

2.3 Listening conditions

The stimulus material was recorded in a hemi-anechoic chamber at the Institute for Hearing Technology and Acoustics using a female voice expert (a speech language pathologist and voice researcher) who read all texts in her regular voice and while imitating a voice disorder. Six speech language pathologists evaluated the degree of hoarseness in both voice qualities. The normal voice was rated as not hoarse at all, while the imitated hoarse voice was rated as moderately to severely hoarse. All samples were loudness-adjusted according to EBU R 128 [8].

To achieve a realistic background sound scenario, a binaural recording of the real-life counterpart to the virtual seminar room was played back at a level of 52 dBA. Using the room acoustic simulation software RAVEN and the auralization framework Virtual Acoustics [9], the virtual professor's target speech was rendered binaurally and adjusted to a level of 65 dBA, resulting in an SNR of 13 dB.

2.4 Procedure

Upon written informed consent and eligibility check, participants were permitted to take part in the study. They were seated at a table in a soundproof booth (exactly matching position-wise with the virtual table in the seminar room), equipped with headphones (Sennheiser HD 650) and a head-mounted display (HTC Vive Pro Eye) with two respective HTC controllers. After an eye calibration, participants completed a practice block of the secondary (vibrotactile) task and then a vibrotactile baseline block. Next, we presented the practice block for the primary (listening) task, containing one practice text with questions. This was followed by the baseline block of the listening task, which contained four texts, two in each voice quality. Two experimental blocks followed in random order. In these blocks, participants performed the dual-task paradigm with the primary and secondary tasks running in parallel. Each block consisted of six texts with questions, one presented in a normal voice quality, and the other in a hoarse voice quality. After each block, participants rated their perceived listening effort for the previous block.

3. RESULTS

An analysis of the preliminary data revealed that listening to a hoarse voice resulted in a significant increase in perceived listening effort, but this effect was not reflected in the behavioral measures. Figure 2 shows the subjective results as a function of voice quality. A paired Wilcoxon signed rank test with continuity correction showed that listening effort was significantly greater under the hoarse voice condition than the normal voice condition (V = 36.5, p = .001, d = 0.57).

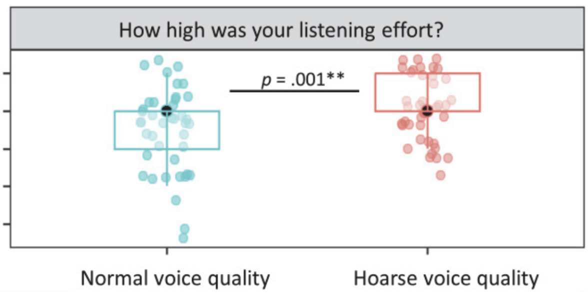

Figure 2. Perceived listening effort as a function of voice quality. The boxplots show the individual data points (colored) and the medians (black) for the normal voice (turquoise) and the hoarse voice (red).

Descriptive results from the DTP are presented in Table 1. We calculated a repeated-measures ANOVAs to assess the effect of *condition* (single-task baseline, dual-task normal voice, and dual-task hoarse voice) on secondary task performance. Although the effect of *condition* was significant (F(2, 78) = 31.42, p < .001), post-hoc comparisons using Tukey's test showed that while baseline performance was significantly better than both dual-task conditions (p-values < .001), there was no significant difference between the dual-task conditions (p = 0.54). The effect of *condition* on response time in the secondary task was modelled with a generalized linear mixed-effects model (GLMM) instead of an ANOVA. A GLMM was used because response time is usually positively skewed and these models do not require prior data transformation to yield a normal distribution [10]. Again, the overall effect of condition was significant ($\chi^2(2)$ = 77.03, p < .001), but pairwise comparisons showed that, despite response times were longer in the two dual-task conditions (p-values < .001), this effect was independent of voice quality (p = .41). In other words, whether the listening task was presented in a normal or hoarse voice had no significant influence on performance and response time measures in the secondary (vibrotactile) task.

Table 1. Descriptive results for the DTP.

Primary (listening) task		
Condition	Performance (in % correct) mean (*SD*)	Response times (in ms) mean (*SD*)
Single-task normal voice	56.48 (18.48)	NA
Single-task hoarse voice	51.98 (16.93)	NA
Dual-task normal voice	48.38 (18.73)	NA
Dual-task hoarse voice	48.08 (16.55)	NA
Secondary (vibrotactile) task		
Single-task baseline	89.43 (10.73)	592 (138)
Dual-task normal voice	69.12 (15.73)	717 (121)
Dual-task hoarse voice	72.08 (13.40)	742 (118)

4. DISCUSSION AND CONCLUSION

This study investigated the influence of a talker's hoarseness on university students' listening effort in a seminar room, using audio-visual immersive VR. Our preliminary results show that perceived listening effort was higher when students listened to a hoarse voice. This confirms the result of a previous auditory-only study, which assessed the effect of a creaky voice on university students [6]. Against our hypothesis, we found no effect of hoarseness on behavioral outcomes, suggesting that subjective experience and performance may not always align in such tasks.

The increase in perceived listening effort under the hoarse voice condition can be explained by the cognitive effort required to compensate for the degraded speech signal. When the talker's voice is hoarse, listeners need to allocate more cognitive resources to perceive and interpret the speech signal [1]. In our experiment, increased cognitive load might have led to higher perceived listening effort. However, is seems that the disturbing effect of hoarse voice was not strong enough to impede performance or prolong processing times. In the future, it will be interesting to study the effect of voice quality in relation to different SNR levels on listening effort, considering that speech intelligibility was rather high in the present study.

Overall, our findings suggest that cognitive aspects of voice perception are complex and multifaceted. They highlight the importance of considering both subjective and objective measures when investigating the influence of abnormal voice qualities, as subjective experience may not always align with objective performance. The findings also suggest that the effect of hoarseness on cognitive processing may be small. In the context of university lectures, students may still be able to understand the lecture content even if the professor's voice quality is suboptimal, but they may experience increased listening effort in doing so. Future research could investigate how a professor's poor voice quality affects students' motivation and learning outcomes, especially in situations of high background noise and prolonged exposure.

ACKNOWLEDGMENTS

Funded under the Excellence Strategy of the Federal Government and the Länder (EXS-SF-OPSF712). Isabel Schiller's contribution to this study was also funded by a grant from the HEAD-Genuit-Foundation (P-16/10-W) awarded to Prof. Sabine Schlittmeier. We thank the student assistants M. Amelung, T. Jungbauer, M. Käppel, M. Kögel, and M. Tenberg for their support.

REFERENCES

[1] M. Pichora-Fuller *et al.*: "Hearing impairment and cognitive energy the framework for understanding effortful listening (FUEL)," *Ear Hear.*, vol. 37, no. 1, pp. 5S–27S, 2016.

[2] S. Azari, A. Aghaz, M. Maarefvand, L. Ghelichi, F. Pashazadeh, and Y. A. Shavaki: "The prevalence of voice disorders and the related factors in university professors: a systematic review and meta-analysis," *J. Voice*, in press, 2022.

[3] N. Roy, R. Merrill, S. Thibeault, R. Parsa, S. Gray, and E. Smith: "Prevalence of voice disorders in teachers and the general population," *J. Speech Lang. Hear. Res.*, vol. 47, no. 2, pp. 281–293, 2004.

[4] I. Schiller, A. Remacle, N. Durieux, and D. Morsomme: "Effects of noise and a speaker's impaired voice quality on spoken language processing in school-aged children: a systematic review and meta-analysis," *J. Speech Lang. Hear. Res.*, vol. 65, no. 1, pp. 169–199, 2022.

[5] F. Paas and P. Ayres: "Cognitive Load Theory: A broader view on the role of memory in learning and education," *Educ. Psychol. Rev.*, vol. 26, pp. 19–195, 2014.

[6] M. Imhof, T. Välikoski, A. Laukkanen, and K. Orlob, "Cognition and interpersonal communication: The effect of voice quality on information processing and person perception," *Stud. Commun. Sci.*, vol. 14, no. 1, pp. 37–44, 2014.

[7] S. J. Schlittmeier, C. Mohanathasan, I. S. Schiller, and A. Liebl, "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," *Tech. rep.*, 2023. Available:. https://publications.rwth-aachen.de/record/958449/files/report.pdf

[8] European Broadcasting Union: "R128-2020: Loudness normalisation and permitted maximum level of audio signals," *Tech. rep.*, 2020. Available: https://tech.ebu.ch/publications/r128

[9] Institute for Hearing Technology and Acoustics, RWTH Aachen University: "Virtual Acoustics – A real-time auralization framework for scientific research," [Software] 2023. Available: http://www.virtualacoustics.org

[10] S. Lo and S. Andrews: "To transform or not to transform: using generalized linear mixed models to analyse reaction time data," *Front. Psychol.*, vol. 6, 2015.