DOI: 10.61782/fa.2023.0607

forum acusticum 2023

THE EFFECT OF VOICE CUE DIFFERENCES ON PERFORMANCE AND EFFORT DURING SPEECH-ON-SPEECH LISTENING

Thomas Koelewijn[1,2*] **Deniz Başkent**[1,2]

[1] Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Netherlands

[2] Research School of Behavioural and Cognitive Neurosciences, University of Groningen, Netherlands

ABSTRACT

Speech perception during two talker listening conditions can be challenging and effortful, especially in hearing impaired individuals. Perceiving differences in voice cues, such as fundamental frequency (F0) and vocal-tract length (VTL), seems to help segregating competing talkers, and improve speech understanding. Pupil dilation is an objective measure for cognitive processing load while listening to speech, also referred to as listening effort. Speech-on-speech perception relies on cognitive mechanisms such as inhibition of a speech masker, and as a result can be more effortful than a non-speech masker. However, it is unknown how voice cue differences can affect effort during speech-on-speech listening.

In this study, participants listened to everyday sentences masked by competing speech consisting of random sentence segments (target to masker ratio = -6 dB) both uttered by the same talker. During the experiment, F0 and VTL voice cues of the speech masker were systematically manipulated and listening effort was measured by means of pupillometry. Results show that when F0 and/or VTL differed between target and masker speech speech-on-speech listening improved. Improvements in performance co-occurred with smaller peak pupil dilation responses during listening, indicating a decrease in listening effort. These outcomes provide a first insight on the impact of voice discriminability on listening effort.

[*] ***Corresponding author***: *t.koelewijn@rug.nl*

Keywords: *Voice cues, listening effort, speech perception, pupillometry.*

1. INTRODUCTION

Pupillometry research has shown that perception of speech in competing masker speech is more effortful than listening to speech in stationary noise [1]. This additional effort is mostly attributed to more engagement of cognitive mechanisms, such as attention [2], working memory [3], and language related processing to segregate competing talkers and reconstruct missing information [4]. These additional cognitive processes for normal hearing (NH) listeners come with the benefit of a more advantaged speech reception threshold (SRT) for speech-on-speech listening compared to speech in non-speech noise, indicating that a certain amount of listening effort can be advantageous.

Interestingly, while relatively good speech-on-speech performance is shown for target and masker speech uttered by talkers of opposite sex, speech-on-speech listening with target and masker speech uttered by talkers from the same sex have shown to result in a less advantaged SRT [5]. In recent studies (El Boghdady and colleagues [6]–[9]), target and masker speech were uttered by the same female talker, while the voice pitch (fundamental frequency; F0) and vocal tract length (VTL) voice cues of the masker speech were systematically manipulated. F0 and VTL voice cues help discriminate between talkers from the same or different sexes and were shown to be particularly important for speech-on-speech performance [8]. While target to masker ratios were fixed (-6 dB), average speech-on-speech intelligibility scores drastically improved with increasing F0 and VTL differences between target and masker speech.

These outcomes suggest that, in addition to cognitive processes, voice perception processes also play an important role in speech-on-speech processing, possibly in segregating competing talkers, and therefore might affect listening effort.

In the current study we investigate the effect of F0 and VTL voice cue difference between target and masker speech on speech-on-speech intelligibility and listening effort. For this we used a similar design as El Boghdady et al. [6] to examine speech perception in speech maskers of varying voice cues, to which we added pupillometry as an autonomous and objective measure for cognitive processing load, i.e., listening effort. We hypothesized that, in line with El Boghdady et al. [6] a difference in F0 and VTL voice cues between target and masker voice should result in an improvement in performance. In addition, we expected the pupillometry results to show a systematic decrease in listening effort when target and masker voices perceptually differ for F0 and VTL voice cues.

2. METHODS

2.1 Participants

Twelve normal hearing adults (self-reported gender 7 males, 5 females; age range 21-25 yrs., median age 23.5 yrs.), recruited at the University of Groningen and the University Medical Center Groningen, participated in the study. Participants reported normal or corrected-to-normal vision and no dyslexia, epilepsy, and/or history of developmental disorders. All participants were native Dutch speakers, provided written informed consent in accordance with the Medical Ethics Committee of the University Medical Center Groningen (METc 2018/427)], and received an hourly compensation.

2.2 Stimuli and task

During the speech-on-speech task, everyday Dutch sentences uttered by a female talker from the VU98 corpus [10] (lists 5-19) were presented as the target sentences. The speech masker consisted of concatenated 1 second samples taken from a different set of sentence lists (2-4) from the same corpus, uttered by the same female talker. Both F0 and VTL voice cues were either the same, or individually or both altered by 4 semitones, for which processing was done in STRAIGHT [6]. This resulted in a total of four conditions. Each trial started with a fixation dot appearing 2 seconds before auditory stimulus onset. Next, the masker speech was presented, and after 3 seconds, the target sentence was presented as well. Following the target sentence offset, the masker sentence continued for 0.5 seconds, and a response prompt was presented 2.5 seconds after that. Participants were instructed to respond by repeating the full sentence as correctly as possible and their response was recorded during the experiment. The target sentence was presented at a target to masker ratio of -6 dB, and the combined target and masker stimulus at an overall sound level of 65 dB SPL.

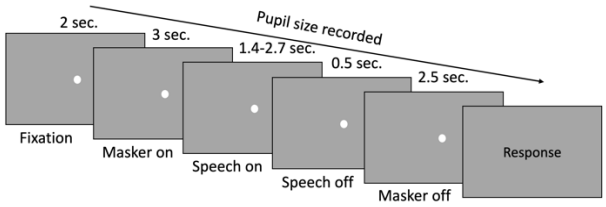

Figure 1. Illustration of a trial. The time window where pupillometry measurements were recorded is shown by an arrow, starting 2 sec before masker onset, and stopping 3 sec after masker offset.

2.3 Procedure and apparatus

After providing informed consent, participant's hearing was tested by means of pure tone audiometry, and the participants filled out a demographics questionnaire.

First, participants were provided two short practice sessions. The first practice session contained 6 practice sentences (list 1), presented without a masker. In the second practice session, 6 new practice sentences (list 1) were presented, but this time with a masker with both F0 and VTL concurrently deviating by 8 semitones. Participants had to repeat each sentence out loud to practice the procedure, like the actual experiment. However, differently than the experiment, and for an effective familiarization, after the participants response, both auditory and visual feedback was provided by presenting the sentence a second time with the sentence written out on screen.

For the experiment, 104 sentences (4 voice conditions x 26 trials) were presented in random order divided over 4 blocks with breaks in between. During each trial, the pupil dilation was recoded from the onset of the trials till the responds prompt, using a Tobii Pro Fusion eye-tracker (Tobii Pro AB, Sweden) at calibrated light conditions (see for same procedure[11]).

The entire session lasted approximately 1.5 hours and was performed in a sound-treated room. Stimuli were presented through Sennheiser HD 280 pro headphones via a MOTU Ultra Lite mk4 soundcard.

3. RESULTS

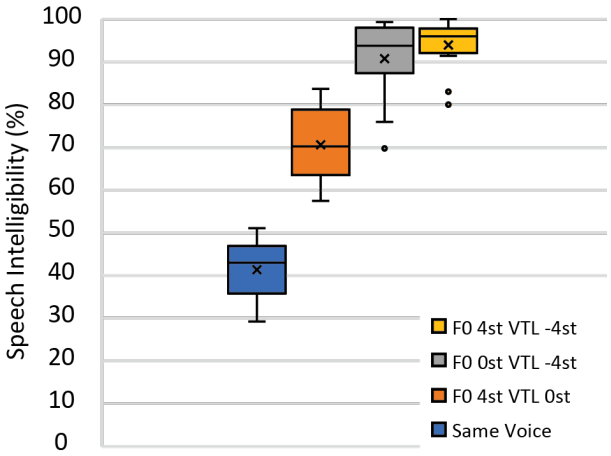

Figure 2. Speech-on-speech intelligibility scores for the four voice cue difference conditions, shown as pooled from all participants. Boxes extend from the lower to the upper quartile (interquartile range, IQ), and the midline indicates the median. The whiskers indicate the highest and lowest values no greater than 1.5 IQ, and the dots indicate the outliers, i.e., data points larger than 1.5 IQ.

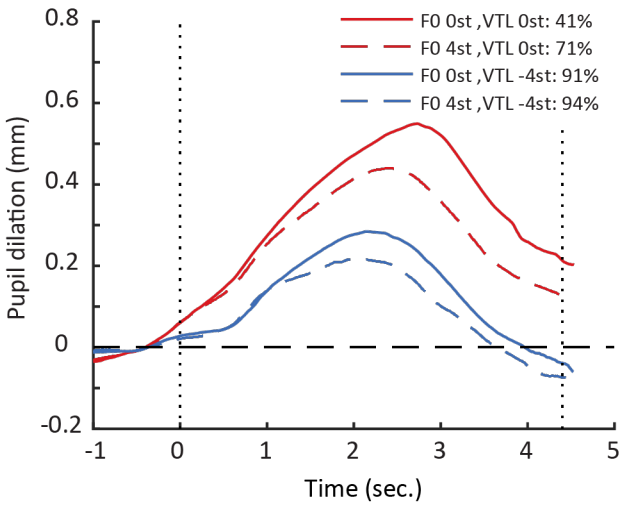

Figure 3. Pupil responses for the four voice conditions, averaged over participants. The onset of the target sentences was at 0 s. The baseline was the average pupil diameter over one second preceding the onset of the target sentence. The legend shows percentage words correct per voice condition.

3.1 Performance

The intelligibility score based on correct number of words repeated per target sentence (See Fig. 2 for the median scores per condition) were analyzed using generalized linear mixed-effect model (GLMM). The outcomes of a Type II Walt Chi-square test showed a significant main effect of voice difference ($X^2_{(3, N = 12)} = 1050.3$, $p > .001$). Post-hoc analysis showed significant contrast between conditions ($p > .001$) except for no significant effect being shown between the delta-VTL and delta-F0+VTL conditions ($p = .112$), which both reached a performance ceiling.

3.2 Pupillometry

Pupil data pre-processing was in line with Bicer et al. [11] and performed in MATLAB. The baseline-corrected pupil traces were averaged for each condition and for each participant and peak pupil dilation (PPD), peak pupil dilation latency (PPDL), and averaged baseline pupil diameter values were calculated. These outcomes were analyzed by separate 2x2 ANOVAs. PPD showed a significant main effect for F0 [$F_{(1,11)} = 22.36$, $p < .01$,], VTL [$F_{(1,11)} = 68.63$, $p < .001$], and no interaction [$F_{(1,11)} = 1.02$, $p = .335$] indicating independent effects of the individual voice cues on cognitive processing load. PPDL showed a significant main effect for F0 [$F_{(1,11)} = 8.88$, $p = .013$], VTL [$F_{(1,11)} = 28.90$, $p < .001$], and no interaction [$F_{(1,11)} < 1$] suggesting that increased cognitive load results in longer processing time. Finally, baseline showed no significant effects [$F_{(1,11)} < 1$], indicating similar levels of cognitive arousal for all voice conditions.

4. DISCUSSION

The results of this study show that when F0 and/or VTL voice cues were different between target and masker speech, in line with previous research [6]–[9], speech-on-speech listening became better. Improvements in performance co-occurred with smaller PPD and shorter PPDL responses during listening. This indicates a decrease in processing load and processing duration when target and masking speech voice cues F0 and VTL deviate, which suggests a decrease in listening effort.

These outcomes provide a first insight on the impact of voice discriminability on listening effort. The results show that voice cue processing affects the pupil dilation response. It could be that either changes in perceptual processes are reflected in pupil responses, or that the impact of improved voice discriminability on talker segregation at an attention related processes stage affects the pupil response.
Finally, these outcomes might provide an explanation for the increased listening effort hearing impaired individuals (e.g., CI-users [12]) experience during listening to speech in adverse conditions, since voice cue perception has shown to be affected in this population [13]. Future research will have to show how voice discriminability affects the PPD in hearing impaired listeners, which might result in pupillometry becoming a powerful diagnostics tool in the audiology clinic.

5. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Etienne Gaudrain for his contributions.

6. FUNDING

This work was supported by a VICI grant (918-17-603) from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw) to the last author, the Heinsius Houbolt Foundation, and a Rosalind Franklin Fellowship.

7. REFERENCES

[1] T. Koelewijn, A. A. Zekveld, J. M. Festen, and S. E. Kramer, "Pupil Dilation Uncovers Extra Listening Effort in the Presence of a Single-Talker Masker," *Ear and Hearing*, vol. 33, no. 2, pp. 291–300, 2012.

[2] T. Koelewijn, H. de Kluiver, B. G. Shinn-Cunningham, A. A. Zekveld, and S. E. Kramer, "The pupil response reveals increased listening effort when it is difficult to focus attention," *Hearing Research*, vol. 323, pp. 81–90, May 2015.

[3] T. Koelewijn, A. A. Zekveld, J. M. Festen, J. Rönnberg, and S. E. Kramer, "Processing Load Induced by Informational Masking Is Related to Linguistic Abilities," *International Journal of Otolaryngology*, vol. 2012, pp. 1–11, 2012.

[4] J. Rönnberg *et al.*, "The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances," *Front. Syst. Neurosci.*, vol. 7, 2013.

[5] J. M. Festen and R. Plomp, "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *The Journal of the Acoustical Society of America*, vol. 88, no. 4, pp. 1725–1736, Oct. 1990.

[6] N. El Boghdady, E. Gaudrain, and D. Başkent, "Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users?," *The Journal of the Acoustical Society of America*, vol. 145, no. 1, pp. 417–439, Jan. 2019.

[7] C. J. Darwin, D. S. Brungart, and B. D. Simpson, "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," *J. Acoust. Soc. Am.*, vol. 114, no. 5, p. 2913, 2003.

[8] D. Başkent and E. Gaudrain, "Musician advantage for speech-on-speech perception," *The Journal of the Acoustical Society of America*, vol. 139, no. 3, pp. EL51–EL56, Mar. 2016.

[9] L. Nagels, E. Gaudrain, D. Vickers, P. Hendriks, and D. Başkent, "School-age children benefit from voice gender cue differences for the perception of speech in competing speech," *The Journal of the Acoustical Society of America*, vol. 149, no. 5, pp. 3328–3344, May 2021.

[10] N. J. Versfeld, L. Daalder, J. M. Festen, and T. Houtgast, "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," *The Journal of the Acoustical Society of America*, vol. 107, no. 3, pp. 1671–1684, Mar. 2000.

[11] A. Biçer, T. Koelewijn, and D. Başkent, "Short Implicit Voice Training Affects Listening Effort During a Voice Cue Sensitivity Task With Vocoder-Degraded Speech," *Ear & Hearing*, vol. Publish Ahead of Print, Jan. 2023.

[12] A. E. Perreau, Y.-H. Wu, B. Tatge, D. Irwin, and D. Corts, "Listening Effort Measured in Adults with Normal Hearing and Cochlear Implants," *J Am Acad Audiol*, vol. 28, no. 08, pp. 685–697, Sep. 2017.

[13] D. Başkent, A. Luckmann, J. Ceha, E. Gaudrain, and T. N. Tamati, "The discrimination of voice cues in simulations of bimodal electro-acoustic cochlear-implant hearing," *The Journal of the Acoustical Society of America*, vol. 143, no. 4, pp. EL292–EL297, Apr. 2018.