DOI: 10.61782/fa.2023.0959

forum acusticum 2023

REAL-TIME SPEECH SEGREGATION AND LISTENING EFFORT DURING SPEECH-ON-SPEECH MASKING – EXAMINATION BASED ON A VISUAL WORLD PARADIGM

Hartmut Meister[1*] **Thomas Koelewijn**[2] **Deniz Başkent**[2]
Khaled Abdellatif[l]
[1]Department of Otorhinolaryngology,
Head and Neck Surgery, University Hospital of Cologne,
Jean Uhrmacher Institute, Germany
[2] Department of Otorhinolaryngology,
University Medical Center Groningen, the Netherlands

ABSTRACT

The ability to segregate competing auditory streams is a necessary prerequisite to focus attention on the target speech during speech-on-speech masking. The Visual World Paradigm (VWP) gives insight into language processing by visually presenting items associated with the speech signal and capturing gaze fixations. Using a novel VWP, the aim of this study was to determine the time course of speech-on-speech segregation when competing sentences are presented and to collect pupil responses as a measure of listening effort. Matrix sentences of the structure ”name-verb-number-adjective-object” were used as speech material. A target and a masker sentence were diotically presented via headphones. Different target-to-masker ratios (TMR) were applied. In parallel the VWP visually presented the number and the object word of both the target and the masker sentences. Participants were instructed to focus their gaze to the number and the object belonging to the target sentence. Additionally, speech recognition performance was determined in an offline experiment to compare the results with the gaze fixations and pupil dilations. Initial results show that the proposed VWP is suited for an objective assessment of sentence-based speech-on-speech segregation and corresponding listening effort on a fine-grained temporal level.

Keywords: *speech-on-speech masking, visual world paradigm, listening effort*

1. INTRODUCTION

Verbal communication frequently takes place in situations with several simultaneous talkers (see “cocktail party problem”, [1]). These situations are typically associated with a high listening effort as the competing talkers have to be segregated in order to stream the target and to focus attention on the information of interest [2]. Segregation is based on different characteristics such as voice and level cues, and the position of the talkers (overview in [3]). However, these cues are not necessarily fully available to listeners with hearing impairment.

Performance in speech-on-speech masking can be assessed by means of speech audiometry. Target speech is presented against a speech background and the resulting speech recognition is determined. Additionally, it would be helpful to apply an objective measure that allows assessment of speech segregation on a fine-grained temporal level. Here, we propose a novel visual world paradigm (VWP, overview in [4]) aiming at giving information about the time course of speech-on-speech segregation and to assess pupil dilatation as a measure of listening effort.

The fundamental of the VWP is the close link between gaze fixations and speech processing. Accordingly, the VWP

[*]***Corresponding author***: hartmut.meister@uni-koeln.de

relies on the simultaneous presentation of auditory and corresponding visual stimuli. The present study shows initial results of the proposed method based on the assessment of young normal-hearing listeners.

2. METHODS

The Oldenburg sentence test (OLSA, [5]) was used for assessing speech recognition in speech masking. The OLSA is a matrix test presenting sentences composed of five words (name – verb – numeral – adjective – object) and ten possible alternatives for each word position, for example "Stephen buys seven wet knives" or "Thomas gives 18 white cups". The stimuli consisted of two competing sentences of the same length differing in each word position. The target sentence was always indicated by the name "Stephen". For the VWP the number and the object of both sentences were presented on a computer monitor (see Fig. 1). The task of the listeners was to direct their gaze on the corresponding number and object of the target sentence while ignoring the masker. Gaze was assessed by an eye-tracker (SR Research eyelink 1000 plus). Simultaneously, the pupil size was determined as an indication of listening effort. Additionally, a speech recognition test was conducted with the same stimulus material.

As a cue for segregating the two competing sentences the intensity of the masker sentence was varied resulting in target-masker ratios (TMR) of 0, 2.5, 4.5, and 6.5 dB. Twelve normal-hearing listeners aged 22 to 27 years participated in the study. The study protocol was approved by the local ethics committee.

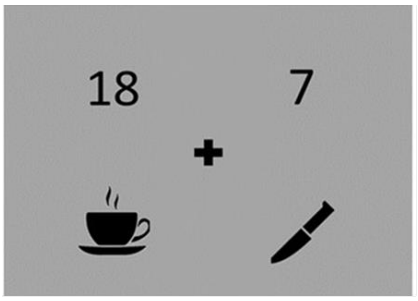

Figure 1. Example of the visual stimulus for the competing sentences "Stephen buys seven wet knives" and "Thomas gives eighteen white cups" including a fixation cross at the center of the screen.

3. RESULTS

The speech recognition test yielded mean scores of 66, 83, 94 and 98 % correct word identification for the TMRs of 0, 2.5, 4.5, and 6.5 dB, thus covering the range from intermediate to near perfect performance. Fig. 2 shows the results for gaze fixation (upper panel) and pupil size (lower panel) for the corresponding TMRs.

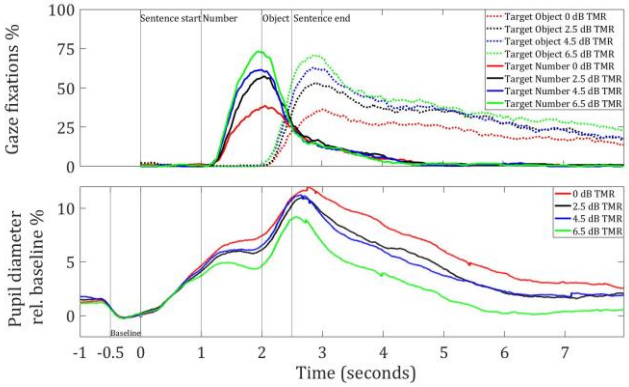

Figure 2. Time course of gaze fixations (upper panel) and pupil dilation (lower panel) for TMRs of 0, 2.5, 4.5, and 6.5 dB. Sentence starts at t=0 s, number and object are presented at about 1 and 2 s, respectively.

The upper panel of Figure 2 reveals the gaze fixations on the number and the object of the target sentence. Both are related to the target-masker-ratio with higher TMRs showing a larger proportion of fixations and a faster growth of the fixation curve. The lower panel of Figure 2 displays the change in pupil diameter relative to the baseline obtained 500 ms prior to the sentence start. The time course of the pupil dilation also reflects the presentation of the number and the object of the target sentence as well as the different TMRs. In general, lower TMRs are associated with a larger change of the pupil diameter relative to the baseline, indicating higher listening effort.

4. PRELIMINARY CONCLUSIONS

This study presents a novel Visual Word Paradigm aiming at giving information about speech-on-speech segregation on a fine-grained temporal level. Matrix sentences were used, which are well suited to assess the effects of speech masking on the one hand and to provide the corresponding visual stimuli of the VWP on the other hand. Different

TMRs were applied providing intensity cues for segregation the target sentence from the masker sentence.
The initial result obtained with the young normal-hearing listeners reveal that both, gaze fixations and pupil dilation reflect the time-course of the competing sentences by giving information on the processing of the number and the object of the target sentence. The effect of the different TMRs is clearly visible, showing a lower proportion of fixations and a higher pupil dilation for the more adverse TMRs. This is despite the fact that the most favorable TMRs yielded near perfect speech recognition.
A thorough assessment of the proposed VWP will be given with a more detailed growth curve analysis, which is subject to present work. Future examinations will include additional segregation characteristics, such as voice cues and different talker positions. Last not least, future studies using the proposed method will consider listeners with hearing loss or cochlear implants in order to give more information on the detrimental effects of auditory impairment on speech-on-speech recognition.

5. ACKNOWLEDGMENTS

This study was supported by a grant of the Deutsche Forschungsgemeinschaft to HM (Reference ME 2751/6-1).

6. REFERENCES

[1] E.C. Cherry: “Some experiments on the recognition of speech, with one and with two ears.” *Journal of the Acoustical Society of America*, 25, 975–979, 1953. doi:10.1121/1.1907229

[2] T. Koelewijn, A.A. Zekveld, J.M. Festen, S.E. Kramer: “Pupil dilation uncovers extra listening effort in the presence of a single-talker masker.” *Ear and hearing*, 33(2), 291–300, 2012. doi:10.1097/AUD.0b013e3182310019

[3] A.W. Bronkhorst: “The cocktail-party problem revisited: early processing and selection of multi-talker speech.” *Attent. Percept. Psychophys*. Jul;77(5):1465-87, 2015. doi: 10.3758/s13414-015-0882-9

[4] F. Huettig, J. Rommers, A.S. Meyer: “Using the visual world paradigm to study language processing: a review and critical evaluation.” *Acta psychologica*, 137(2), 151–171, 2011. doi:10.1016/j.actpsy.2010.11.003

[5] K. Wagener, V. Kühnel, B. Kollmeier: “Entwicklung und Evaluation eines Satztests in deutscher Sprache I: Design des Oldenburger Satztests [Development and evaluation of a German sentence test – Part I: Design of the Oldenburg sentence test]”. *Z. Audiol.* 38(1):4–15. 1999.