PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Differences in Speech-on-Speech Processing Between Musicians and Non-Musicians: The Role of Durational Cues

Elif Canseza Kaplan[1]; Deniz Başkent[2]; Anita E. Wagner[3]

[1] Research School of Behavioral and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, Netherlands

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

ABSTRACT

In the current study, we investigate the role of prosodic cues in speech-on-speech perception in musicians and non-musicians. Earlier studies have shown that musically experienced listeners may have an advantage in speech-on-speech performance in behavioral tasks (1,2). Previously, we have also shown in an eye-tracking study that musical experience has an effect on the timing of resolution of lexical competition when processing quiet vs masked speech (3). In particular, musicians were faster in lexical decision-making when a two-talker masker was added to target speech. However, the source of the difference observed between groups remained unclear. In the current study, by employing a visual world paradigm, we aim to clarify whether musicians make use of durational cues that contribute to prosodic boundaries in Dutch, in resolving lexical competition when processing quiet vs two-talker masked speech. If musical training preserves listeners' sensitivity to the acoustic correlates of prosodic boundaries when processing masked speech, we expect to observe more lexical competition and delayed lexical resolution in musicians. We will compare gaze-tracking and pupil data of both groups across conditions.

Keywords: speech-on-speech, cocktail-party effect, musical training, eye-tracking, pupillometry

1. INTRODUCTION

Musical training has been shown to have a positive effect in understanding speech in a competing talker(s) situation (1–3), also known as the cocktail-party phenomenon (4). Previously, we employed the visual world paradigm (5,6) to investigate whether the processing of speech masked by competing speakers differs between musicians and non-musicians (3). Our results indicated that musicians were faster in terms of resolving the lexical ambiguity and the speech masking engaged their attention differently. However, it remained unclear whether this positive effect of musical experience is due to musicians being better able to suppress the maskers or being more sensitive to the acoustic cues that matter in making the distinction between two words that are phonologically similar in their onset

Speech comprehension involves resolving lexical competitions between phonologically similar items and acoustic ambiguities resolves as the speech signal unfolds in time. Durational cues that contribute to prosodic boundaries in Dutch are known to play a role in resolving this lexical ambiguities (7). For instance, onset-embedded words such as, ham/hamster that share the initial syllable have inherently different durational properties in the first syllable. In the current study, we aimed to investigate whether these cues that enable the resolution of lexical competition, as they have been shown to play a role in quiet speech (7), are picked up differently by musicians and non-

[1] e.c.kaplan@rug.nl

[2] d.baskent@umcg.nl

[3] a.wagner@umcg.nl

musicians when speech is presented within a two-talker masker. We embedded monosyllabic words (ham) in polysyllabic ones (hamster), which in turn generated words containing longer syllables in spliced (ham+ster) words. Thus, there were durationally matching (hamster) or mismatching (ham+ster) target words, either presented without maskers in quiet, or within two-talker maskers. We measured participants' eye-movements and changes in pupil dilation to capture the attention and effort induced by masking. If musically experienced individuals are more attentive to these durational cues despite the background noise, they are expected to exhibit more lexical competition and a delay in the lexical resolution when the durational cues are mismatching as opposed to the matching condition.

2. METHOD

2.1 Participants

Twenty-five musically trained and twenty-six non-musician listeners with normal hearing participated in the study. The criteria for having normal hearing involved having less than 25 dB HL pure tone thresholds between 250 to 4000 Hz bilaterally. Musician criteria were based on the literature (8) to be: having more at least 10 years of training, having started music at/before the age of 7 and actively making music within the past 3 years, prior to the study.

2.2 Materials

For target sentences, we used twenty-six semantically neutral Dutch sentences, containing a polysyllabic target word (i.e., hamster) that enable the embedding of a monosyllabic word (i.e., ham). The recordings consisted of utterances of a Dutch female speaker without any regional accent (see (9) for details). The monosyllabic recordings were embedded in the polysyllabic ones to generate the target mismatching duration conditions.

The masker sentence set was taken from another corpus (10) and consisted of meaningful Dutch sentences, uttered by a female speaker without any regional accent. The same female speaker's voice was used to generate the two-talker maskers. The target speaker's utterances were embedded within two-talker masker, such that the onset of the target speaker's utterance was 500 ms after the onset of the maskers' utterance. Also, the maskers ended 500 ms after the offset of the utterance of the target speaker. If the duration of a single sentence of the masker was not sufficient to go beyond the extra 500 ms, another sentence was randomly chosen and added to the signal. All sentences were generated offline before the experiment. All sentences were presented at the same level of intensity, whether masked or quiet, of 70 dB SPL.

Additionally, black and white pictures were created to be used in the visual world paradigm. The images consisted of the pictures that referred to the target words (hamster), phonological competitors (ham) and the semantically or phonologically unrelated distractors (box).

2.3 Procedure

All participants initially underwent audiometric check. Those with normal hearing proceeded with the experiment. Following these initial tests, the experiment took place in two parts. In the first part, participants were shown the pictures utilized in the experiment and asked to name them. Experimenters made sure the pictures were referred as the same as used in the experiment. Then, the eye-tracker was calibrated.

The second part of the experiment had two blocks: in the first block, participants listened to the target speaker in quiet and the second block was consisted of the masked condition. The order of presentation of the blocks was always constant, since in the quiet block, participants were familiarized with the voice of the target speaker. Each block contained target matching and mismatching duration condition words. Before each list started, participants completed four practice trials. In each trial, participants initially saw a cross in the middle of the screen, which was followed by the simultaneous presentation of the audio and visual information. Participants were asked to pay attention to the target speaker's utterances and choose the image that they heard in the utterance, from the four images displayed on the screen. Their gaze fixations to the four images, as well as the pupil dilations were recorded.

3. RESULTS

The gaze fixations revealed that in the target matching duration condition, both groups performed similarly in terms of timing of lexical resolution. There was a slight delay for both groups in the masked condition of target matching duration condition. In the target mismatching duration condition, the timing of lexical resolution was delayed for both groups, but it was more so for the musically trained group. Both in quiet and masked speech, musically trained listeners exhibited more lexical competition and a delay in resolving the lexical ambiguity.

4. CONCLUSIONS

The target matching duration condition's results were in line with our previous findings, where we observed a delay in timing of lexical ambiguity when masking was added to the speech signal (3). In the target mismatching duration condition, musicians appeared to be affected more by the mismatching duration than non-musicians, suggesting that they may be more sensitive to the durational prosodic cues despite the speech maskers. Both groups were affected by the speech masker, indicated by the delay in lexical decision-making; however, non-musician group appeared to be less affected by the duration manipulation both in quiet and in masked speech. Results will be analyzed employing growth curve analysis, which captures how the gaze fixation curves differ across time between groups and conditions. Additionally, pupil responses are still being processed for further analysis.

ACKNOWLEDGEMENTS

We would like to thank our BCN master's student assistant Joëlle Jagersma for assisting the data collection. This project was supported by funding from the EUs H2020 research and innovation programme under the MSCA GA 67532*4 (the ENRICH network: www.enrich-etn.eu), and the second and last authors were supported by a VICI Grant (No. 016.VICI.170.111) from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw).

REFERENCES

1. Başkent D, Gaudrain E. Musician advantage for speech-on-speech perception. J Acoust Soc Am. 2016;139(3):EL51–6.
2. Swaminathan J, Mason CR, Streeter TM, Best V, Kidd G, Patel AD. Musical training, individual differences and the cocktail party problem. Sci Rep. 2015;5(11628):1–10.
3. Kaplan EC, Wagner AE, Baskent D. Are musicians at an advantage when processing speech on speech? In: Parncutt R, Sattmann S, editors. Proceedings of ICMPC15/ESCOM10. Graz, Austria: Centre for Systematic Musicology, University of Graz; 2018. p. 233–6.
4. Cherry CE. Some Experiments on the Recognition of Speech, with One and with Two Ears. J Acoust Soc Am. 1953;25(5):975–80.
5. Cooper RM. The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. Cogn Psychol. 1974;6(1):84–107.
6. Salverda AP, Tanenhaus MK. The Visual World Paradigm. In: Research methods in psycholinguistics: A practical guide. 2017. p. 89–110.
7. Salverda AP, Dahan D, McQueen JM. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. Cognition. 2003;90(1):51–89.
8. Parbery-Clark A, Skoe E, Lam C, Kraus N. Musician enhancement for speech-in-noise. Ear Hear. 2009;30(6):653–61.
9. Wagner AE, Toffanin P, Baskent D. The timing and effort of lexical access in natural and degraded speech. Front Psychol. 2016;7(MAR).
10. Versfeld NJ, Daalder L, Festen JM, Houtgast T. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. J Acoust Soc Am. 2000 Mar [cited 2018 Feb 21];107(3):1671–84.