Effects of Energetic Masking and Speech Modulation on Spoken Word Recognition

Dorina Strori [1], *Johannes Zaar*[2], *Odette Scharenborg*[3], *Sven Mattys* [1]

[1] University of York
[2] Technical University of Denmark
[3] Radboud University Nijmegen

dorina.strori@york.ac.uk

Index Terms: spoken word recognition, lexical memory, specificity effects, energetic masking, speech modulation

1. Introduction

Previous research indicates that listeners encode both linguistic and talker-related specifications of the speech signal in memory (Goldinger, 1998). Recent evidence suggests that non-linguistic sounds co-occurring with spoken words are also incorporated in our lexical memory (Pufahl & Samuel, 2014; Cooper et al, 2015). We argue that this "sound-specificity effect" might not be due so much to a word-sound association per se as to the different acoustic glimpses of the words that the associated sounds (maskers) create. We anticipated that contrasted acoustical glimpses would lead to a sound-specificity effect. Independently, we also tested the hypothesis that the extent to which a masker is integral (i.e., not easily segregable) to the word plays a role in the emergence of the effect. We predicted that there would be an effect of the sound change on word recognition memory only when the maskers were rendered integral to the words, not when they were not.

2. Method

In a series of six recognition-memory experiments, we assessed listeners' recognition memory performance for spoken words. The first experiment aimed at replicating the voice-specificity effect, hence the stimuli consisted of bi-syllabic, high frequency words, spoken by a male and a female talker. The next experiments investigated the sound-specificity effect in different contexts, thus the stimuli consisted of word-sound pairs (two sounds, the same set of words). We tested the different acoustical glimpses scenario in the next three experiments by varying the level of energetic masking from exposure to test, to create contrasted vs. non-contrasted glimpses. Two car horn sounds with an on/off pattern were used. In the last two experiments, we rendered the maskers integral to the words through pitch and intensity modulation. A violin and a cat sound were used as maskers. All the experiments in the series consisted of an exposure phase, a short delay and a memory test phase. During exposure, participants listened to the stimuli played one at a time binaurally, and performed an "animate/inanimate" semantic judgement task on the words only, to encode them in memory. After a 5-minute break they completed the test phase, where they had to decide whether the word was heard during exposure (*old*), or not (*new*). Accuracy and response latencies for each trial were measured.

3. Results

Our first experiment replicates the voice-specificity effect, such that the overall recognition accuracy for same-voice word repetitions was higher than that for different-voice repetitions. In the next three experiments, we observed a sound specificity-effect only when there was a high energetic masking contrast between the stimuli in the two phases. Namely, the sound change from exposure to test led to a drop in the overall recognition accuracy for previously heard words only when the glimpses resulting from the masking of two different sounds on the same word were highly contrasted. There was no effect on the response latencies. In the fifth experiment, we found an effect of the sound change on both recognition accuracy and response latencies when the sounds were made more integral to the words. As anticipated, removing the integrality factor from the stimuli in the last experiment cancelled the effect.

4. Discussion & Conclusions

Our results extend previous findings of the sound-specificity effect by identifying plausible contexts in which this novel emerges. Importantly, they suggest that listeners seem to encode irrelevant non-speech information in memory, but only in certain contexts. Hence, the assimilation of background sounds into lexical memory cannot be reduced to a simple context effect. We have shown that the mere co-occurrence between the words and sounds is not enough for the emergence of a reliable sound-specific effect on recognition memory. The auditory episodes consisting of words and sounds seem stronger when they consist of contrasted glimpses of the target words. However, contrary to Pufahl and Samuel (2014)'s claim, there is no evidence that the sound *itself* is included in the episode. Furthermore, episodes seem stronger when the sound is integral to, hence difficult to dissociate from the word.

5. Acknowledgements

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no FP7-PEOPLE-2011-290000.

6. References

[1] S. D. Goldinger "Echoes of echoes: An episodic theory of lexical access," *Psychological Review*, vol. 105, pp. 251–279, 1998.

[2] A. Pufahl and A.G. Samuel, "How lexical is the lexicon? Evidence for integrated auditory memory representations," *Cognitive Psychology*, vol. 70, pp. 1–30, 2014.

[3] A. Cooper, S. Brower and A.R. Bradlow "Interdependent processing and encoding of speech and concurrent background noise," *Attention, Percepton & Psychophysics*, vol. 77, no. 4, pp. 1342–1357, 2015.