Neighbourhood density influences word recognition in native and non-native speech recognition in noise

Florian Hintz & Odette Scharenborg

Centre for Language Studies, Radboud University Nijmegen, The Netherlands

f.hintz@let.ru.nl o.scharenborg@let.ru.nl

Index Terms: neighbourhood density, word frequency, noise, native listening, non-native listening

1. Introduction

Recognising spoken words in noisy environments can be challenging [1], and this is especially the case when listening in a non-native language [2]. Previous research has shown that native and non-native listeners exploit both word-initial and word-final information when recognising spoken words [3,4,5,6]. Against this background, [6,7] examined the importance of both word-initial and word-final information in native and non-native word recognition in noise. English and Dutch listeners were presented with English words either in optimal listening conditions or partially masked with speech-shaped noise at different SNR levels (i.e., -12, -6, 0 dB). Crucially, the position of the noise was manipulated, with noise either occurring on a word's onset or offset. The results showed that as listening conditions deteriorated, both native and non-native listeners recognised fewer words when word-initial information was masked compared to when word-final information was masked [6,7]. The present study investigates in more detail how masking word-initial and word-final information contributes to word misperceptions by native and non-native listeners. Specifically, we investigate how item variables such as word frequency and neighbourhood density, which have been suggested to play an important role in recognising spoken words [8], influence noise-induced misperceptions in native and non-native word recognition in noise. To that end, we re-analysed the noise trials from Scharenborg et al. [7] for which participants provided existing but incorrect words as response (i.e., misperceptions). We quantified their misperceptions by calculating the deviance between expected and typed in words and linked these difference scores to experimental (position of noise and SNR) and item variables (word frequency and neighbourhood density).

2. Analysis

For details of the experimental design and procedure see [6,7]. The present analysis focused on misperceptions by 51 English (512 data points in total) and 61 Dutch (1190 data points in total) participants. The phonetic transcriptions of the participants' responses and the target words were transformed into the "DISC" transcription format such that each phoneme corresponded to a unique symbol [9]. The Levenshtein distance [10] between both strings was then calculated and normalised for the target words' length. The target words' lexical frequency and neighbourhood density were looked up online using the SUBTLEX-UK [11] and CLEARPOND [12] databases. The individual contributions of experimental and item variables and their interactions were estimated using linear mixed effect regression models [13] in R.

3. Results and discussion

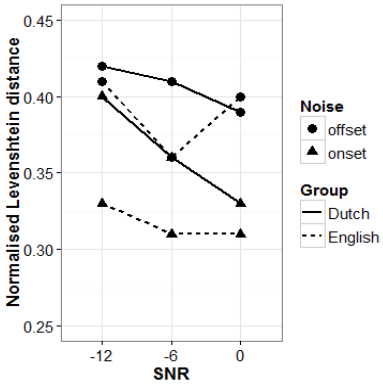

Figure 1: Normalised Levenshtein distance for the English and Dutch listener groups by position of noise and SNR.

For the English data, a significant interaction between noise position and SNR was found (see the dashed lines in Figure 1): Surprisingly, the difference in deviance between onset and offset masked words was relatively smaller at SNR -6 than at SNRs 0 and -12 (although offset noise resulted in greater overall deviance than onset noise). For the Dutch listeners, a main effect of SNR was observed: Responses deviated more strongly from the target as listening conditions deteriorated. Moreover, offset masking generally led to greater deviance between the participants' responses and the target words. Importantly, for both listener groups (though to a weaker degree in the Dutch group), the main effects and interaction were moderated by the target words' neighbourhood density (English: t = 3.31; Dutch: t = 1.77, marginally significant) in that words with more lexical neighbours led to responses with a larger deviance from the target than words with fewer lexical neighbours. Word frequency did not seem to affect the misperceptions in either listener group.

Our data extend Luce and Pisoni's proposal that neighbourhood density affects word recognition [8] to the realm of (non-native) listening in noise. The lack of an effect of word frequency on the misperceptions was surprising given, among others, its assumed role in Luce and Pisoni's model [8]. A supplemental analysis on the entire set of Scharenborg et al. where correctness (correct vs. incorrect) was used as dependent variable showed that word frequency did in fact influence the correct recognition of the presented target while neighbourhood density did not. Possibly, the influences of word frequency and neighbourhood density on the word recognition process differ in granularity.

4. Acknowledgements

This research was sponsored by a VIDI-grant from the Netherlands Organisation for Scientific Research (NWO; grant number: 276-89-003) to O.S. The authors would like to thank Juul Coumans and Polina Drozdova for collecting the data.

5. References

[1] J. Meyer, L. Dentel, and F. Meunier, "Speech Recognition in Natural Background Noise," *PLoS ONE*, vol. 8, 2013.

[2] A. Cutler, "Why is it so hard to understand a second language in noise?," *American Association of Teachers of Slavic and East European Languages (Newsletter)*, p. 16, 2005.

[3] P. D. Allopenna, J. S. Magnuson, and M. K. Tanenhaus, "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models," *Journal of Memory and Language*, vol. 38, pp. 419-439, May 1998.

[4] A. Weber and A. Cutler, "Lexical competition in non-native spoken-word recognition," *Journal of Memory and Language*, vol. 50, pp. 1-25, 2004.

[5] M. L. G. Lecumberri, M. Cooke, and A. Cutler, "Non-native speech perception in adverse conditions: A review," *Speech Communication*, vol. 52, pp. 864-886, 2010.

[6] J. Coumans, R. van Hout, and O. E. Scharenborg, "Non-native word recognition in noise: The role of word-initial and word-final information," *Proceedings of INTERSPEECH*, pp. 519-523, 2014.

[7] O. E. Scharenborg, J. Coumans, and R. van Hout, "The Use of Word-initial and Word-final Information in Native and Non-native Spoken-word Recognition," in preparation.

[8] P. A. Luce and D. B. Pisoni, "Recognizing Spoken Words: The Neighborhood Activation Model," *Ear and hearing*, vol. 19, pp. 1-36, 1998.

[9] R. H. Baayen, R. Piepenbrock, and L. Gulikers, "The CELEX lexical database on CD-ROM," ed: Philadelphia, PA: University of Pennsylvania [Distributor], 1993.

[10] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals (translated)," *Doklady Akademii Nauk SSSR*, vol. 163, pp. 845–848, 1965.

[11] W. J. B. van Heuven, P. Mandera, E. Keuleers, and M. Brysbaert, "SUBTLEX-UK: A new and improved word frequency database for British English," *The Quarterly Journal of Experimental Psychology*, vol. 67, pp. 1176-1190, 2014.

[12] V. Marian, J. Bartolotti, S. Chabal, and A. Shook, "CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities," *PLoS ONE*, vol. 7, 2012.

[13] R. H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language*, vol. 59, pp. 390-412, 2008.