Perception of Sentence Accent in Non-native Speech in Noise

Odette Scharenborg[1], Elea Kolkman[1], and Brechtje Post[2]

[1]Centre for Language Studies, Radboud University Nijmegen, The Netherlands
[2]Department of Theoretical and Applied Linguistics, University of Cambridge, UK

o.scharenborg@let.ru.nl, elea.kolkman@student.ru.nl, bmbp2@cam.ac.uk

Index Terms: prominence detection, phoneme detection, sentence accent, native listening, non-native listening, noise

1. Introduction

In optimal listening conditions, native listeners are able to exploit prosodic cues in the speech signal signalling upcoming sentence accent to actively focus their attention to those parts of the sentence where accent will fall [1]. Highly-proficient non-native listeners are able to detect sentence prominence [2][3] and to use similar acoustic, prosodic cues as native listeners for prominence detection [1][3]. Nevertheless, they display a reduced efficiency in using prosodic information signalling sentence accent for speech processing [1]. This research investigates whether native English listeners and non-native Dutch listeners of English exploit prosodic information signalling sentence accent to aid speech perception in the presence of background noise, while pulling apart the role of preceding prosodic cues and accent on prominence detection. The English-Dutch language pair allows us to investigate the influence of prosodic information on non-native spoken-word recognition without vital mismatches at the phonological level and with a reasonably small mismatch at the sound level as Dutch and English prosodic structures for sentence accent and prosodic processing are highly similar [1].

2. Experimental set-up

Forty-eight native Dutch listeners and 47 native English listeners participated in a phoneme monitoring experiment (following [1]). Target phonemes (/p, t, k/; maximally one per sentence) appeared in word-initial position (only of nouns) in 48 English experimental sentences, and were presented in four different levels of background noise (quiet and speech-shaped-noise at SNR +5, 0, -5). An additional 48 filler distractor sentences were created, half contained a target phoneme while the other half did not. Prior to each sentence the target phoneme was shown on the screen. Participants were asked to listen within a sentence for the presence of the target sound that was specified, and press the space bar as fast as possible upon hearing the target phoneme.

Crucially, sentence accent was manipulated, using cross-splicing of the target-bearing word, in two prosodic contexts. All sentences contained prosodic context preceding the target-bearing word signalling sentence accent on the upcoming target-bearing word; however in the 'half-accented' condition, the target-bearing word was deaccented, while it was accented in the 'accented' condition. The half-accented and accented conditions thus had identical prosodic information preceding the target-bearing words. Differences between the two conditions can thus only be attributed to absence or presence of sentence accent on the target-bearing word.

3. Results and discussion

Statistical analyses on the reaction times (RT; on the correctly detected phonemes) and the target phoneme detections on the experimental sentences were carried out using generalized linear mixed effect models [4]. The RT analyses showed that English listeners suffered significantly less from increasingly more difficult listening conditions than Dutch non-native listeners (interaction of Noise with Language Group; β=0.03584, *SE*=0.01198, p<.01). Moreover, visual inspection of the RT results (see top panels of Figure 1) shows that the 'breaking point' for the Dutch listeners is at SNR +5, while this is only at SNR -5 for the English listeners. The phoneme detection analyses showed no significant difference between the two listener groups. More targeted analyses showed that at SNR 0 (β=0.07658, *SE*=0.03253, p<.05) and -5 (β=0.12508, *SE*=0.03935, p<.01) Dutch non-native listeners detected significantly fewer target phonemes in the half-accented compared to the accented condition (see bottom panels in Figure 1) while this difference was only observed at SNR -5 for the English listeners (β=0.08954, *SE*=0.03639, p<.05).

Both native and non-native listeners use prosodic information for phoneme detection; however where native listeners seem able to partially overcome the noise-induced problems at the acoustic level by using prosodic information signalling upcoming sentence accent, non-native listeners cannot do so to the same extent, even though the key cues are very similar in their own native language. These experiments suggest that the performance difference between native and non-native listeners in the presence of background noise is, at least partially, caused by a reduced exploitation of contextual information during non-native speech processing.

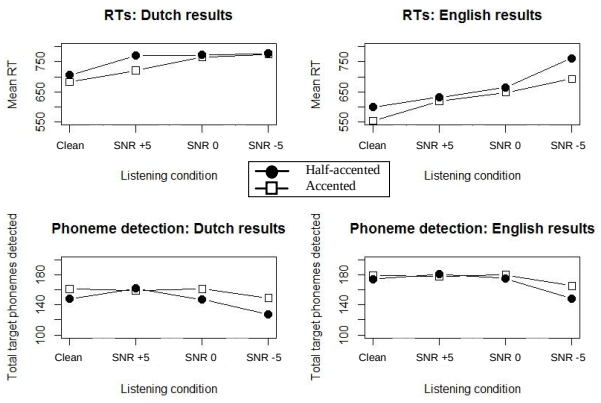

Figure 1. Mean reaction times (top panels) and total number of detected target phonemes (bottom panels) for the Dutch non-native (left) and English native (right) listeners.

This research is supported by a Vidi-grant from NWO (grant number: 276-89-003) awarded to OS.

4. References

[1] Akker, E., Cutler, A. "Prosodic cues to semantic structure in native and non-native listening.", *Bilingualism: Lang. & Cogn.* 6:81-96, 2003.

[2] Rosenberg, A., Hirschberg, J., Manis, K. "Perception of English prominence by native Mandarin Chinese speakers", *Fifth International Conference on Speech Prosody*, 2010.

[3] Wagner, P. "Great expectations – Introspective vs. perceptual prominence ratings and their acoustic correlates", *Proceedings of Interspeech*, Lisbon, Portuagal, 2381-2384, 2005.

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