ADAPTING TO LEXICAL STRESS IN A FOREIGN ACCENT

Eva Reinisch[a] *& Andrea Weber*[a,b]

[a]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands;
[b]Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands
eva.reinisch@mpi.nl; andrea.weber@mpi.nl

ABSTRACT

An exposure-test paradigm was used to examine whether Dutch listeners can adapt their perception to non-canonical marking of lexical stress in Hungarian-accented Dutch. During exposure, one group of listeners heard only words with correct initial stress, while another group also heard examples of unstressed initial syllables that were marked by high pitch, a possible stress cue in Dutch. Subsequently, listeners' eye movements to target-competitor pairs with segmental overlap but different stress patterns were tracked while hearing Hungarian-accented Dutch. Listeners who had heard non-canonically produced words previously distinguished target-competitor pairs faster than listeners who had only been exposed to canonical forms before. This suggests that listeners can adapt quickly to speaker-specific realizations of non-canonical lexical stress.

Keywords: perceptual adaptation, foreign accent, lexical stress, eye tracking

1. INTRODUCTION

It usually takes little effort for listeners to understand what is said in their native language. Listening to foreign-accented speech can make this task considerably harder. A number of studies has shown now, however, that comprehension can improve after only little exposure to a foreign accent, with listeners even being able to generalize certain accent features to new speakers with the same language background (e.g., [3, 10]). While most previous studies focused on global foreign accent using offline sentence and word transcription tasks, a few recent studies have helped to draw a more detailed picture of what affects the perception of foreign accent [2] and how these factors influence online word recognition [5, 8].

As for the effects of short-term adaptation on online processing, the main focus has been on adaptation to segmental mispronunciations [6, 13]. But foreign accents manifest themselves also at the suprasegmental level such as intonation and lexical stress [2, 4]. This raises the question of whether listeners can not only adapt to segmental variation in foreign-accented speech but also to suprasegmental variation.

The present paper addresses this issue by asking whether Dutch listeners can adapt to lexical stress errors in Hungarian-accented Dutch. Dutch is a free-stress language in which the location of primary stress differs across words and can distinguish lexical meaning in a number of cases (e.g., *KAnon* - *kaNON*; "canon"-"cannon"; capitals indicate stress). The main correlates of stress are duration, pitch, and to a lesser extent spectral tilt (e.g., [11]). Hungarian, in contrast, is a fixed-stress language where stress always falls on the word-initial syllable and therefore is not informative for lexical distinctions. Traditionally, the main acoustic stress correlate in Hungarian was considered to be intensity but more recent studies suggest sentence-level pitch accents [12]. Given these differences it can be expected that Hungarians have difficulties in remembering stress location in Dutch and that Hungarian learners of Dutch mark stress differently than Dutch native speakers do.

Importantly, Dutch listeners have been shown to use lexical stress information at the earliest moments of word recognition in Dutch: in a visual-world eye tracking study, Dutch listeners were able to distinguish segmentally overlapping target-competitor word pairs by means of their stress patterns before disambiguating segmental information became available [9]. The present study therefore implemented the same paradigm to assess whether Dutch listeners can adapt to suprasegmental foreign accent after a short exposure phase.

During exposure, one group of listeners heard only words with correct initial stress, while another group also heard examples of unstressed initial syllables. The Hungarian speaker, however, marked unstressed initial syllables with high pitch which is also a stress cue in Dutch. Subsequently, listeners' eye movements to target-competitor pairs that overlap segmentally on the first syllable but differ in their canonical stress pattern (e.g., *HERsens* - *herSTEL*; "brains"-"recovery") were tracked. The questions were whether listeners who had been exposed to the speaker's stress errors before, (1) would be better at using this information during processing than listeners that had just been familiarized with the speaker's global foreign accent, and (2), will be able to use this information early during word processing.

2. METHOD

2.1. Participants

Sixty native speakers of Dutch who were raised monolingually participated for a small payment. They reported no hearing problems and had normal or corrected-to normal vision. None knew Hungarian nor was familiar with Hungarian accent.

2.2. Materials

A short Dutch story was created for exposure (spoken duration approximately 2.3 minutes). One version contained only monosyllabic or initially stressed words ("no-evidence story"). In a second version ("evidence story"), twenty-eight words were replaced with semantically fitting words with stress on the second syllable (e.g., *EEKhorn* "squirrel" became *koNIJN* "rabbit").

For the eye-tracking study, thirty-two Dutch bisyllabic target-competitor pairs were selected as critical pairs. In critical pairs, target and competitor overlapped segmentally up to the first phoneme of the second syllable but contrasted in their canonical stress pattern (word-initial stress vs. stress on second syllable; e.g., *HERsens* - *herSTEL*). Another thirty-two pairs with segmental overlap and matching stress pattern served as fillers (sixteen each with stress on first or second syllable). Each target-competitor pair was combined with a distractor pair that matched in stress properties but did not overlap segmentally with the target-competitor pair.

A female Hungarian learner of Dutch recorded both versions of the story and all target-competitor pairs embedded in the sentence *Klik op het woord [TARGET]* ("Click on the word [TARGET]"). At the time of recording the speaker was twenty-nine years old and had been living in the Netherlands for six years (attending university courses during the 6th year). In order to obtain consistent recordings with word-initial stress, the speaker was instructed not to suppress her natural Hungarian accent and to stress all words on the initial syllable, as she would usually do in Hungarian, irrespective of the correct Dutch stress pattern. The word pairs for eye tracking were recorded in random order to avoid intentional use of contrastive stress cues.

2.3. Acoustic measures

Duration and pitch were measured on the vowels of the first syllables of the critical words from the stories as well as on the critical target-competitor pairs from the eye-tracking study. In addition, first vowels of twenty-eight bisyllabic words that occurred in both stories were measured as control. Table 1 shows duration and pitch values for word pairs from the exposure stories.

Table 1: Mean duration (ms) and mean pitch (Hz) of the first vowels from the critical words in the stories, and the same number of matched control words.

	critical words (correct vs. incorrect initial stress)		control words (always correct initial stress)	
	evidence story	no-evidence story	evidence story	no-evidence story
duration	84 ms	122 ms	130 ms	125 ms
pitch	204 Hz	170 Hz	200 Hz	201 Hz

Incorrectly stressed words in the evidence story were marked by a relatively higher pitch than correctly stressed words in the no-evidence story. The first vowels of the correctly stressed words in the no-evidence story in turn were longer in duration than the incorrectly stressed words in the evidence story. Note that a direct comparison of these duration and pitch values cannot be interpreted straightforwardly as the words were not matched on their segmental make-up (e.g., *EEKhorn* vs. *koNIJN*). Little difference in either duration or pitch was found for measures of the control words which were correctly stressed on their initial syllables in both versions of the story. This ensured that no unintended acoustic differences were present during exposure.

Table 2: Mean duration (ms) and mean pitch (Hz) of the first vowels from the critical target-competitor pairs, and the comparison of their distributions.

	mean for correct stress	mean for incorrect stress	$t(31)$	P
duration	143 ms	96 ms	4.61	$< .001$
pitch	176 Hz	184 Hz	- 4.50	$< .001$

Table 2 further shows comparisons of duration and pitch values on the words' initial syllables for the critical target-competitor pairs. Here direct comparisons of acoustic measures can be made as the target and competitors overlapped segmentally on their initial syllables (e.g., *HERsens* - *herSTEL*). Again syllables with canonical initial stress were marked with longer duration whereas non-canonically stressed syllables were marked with higher pitch. Correct and incorrect stress cues were thus consistent in exposure and test.

2.4. Procedure

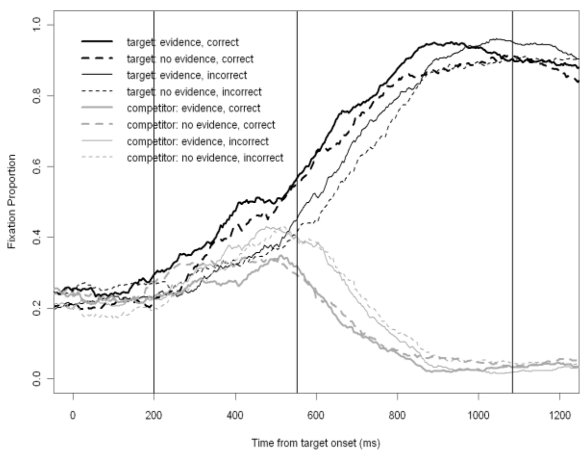
Participants were fitted with an Eyelink 1000 system (SR Research) to monitor their eye movements. During exposure, half of the participants heard the story with only initially stressed words (no-evidence group), the other half heard the story in which the twenty-eight non-initially stressed words occurred. The latter group thus received evidence that unstressed initial syllables were marked with high pitch (evidence group). During exposure, participants simply listened and viewed a fixation cross. The following visual-world task was the same for everyone. On every trial, four printed words (i.e., target, competitor and distractor pair) were displayed in the four quadrants of the screen in Lucida Console font, size 35. At the same time participants heard the sentences instructing them to click on one of the words (i.e., the target). Participants saw each target-competitor pair only once. Target-competitor pairs were counterbalanced across participants such that each word served equally often as target and competitor. For each participant half of the words had (correct) initial stress and half had incorrect stress. 1800 ms after acoustic target onset or at listeners' response the printed words disappeared from the screen. 1000 ms after the response, the next trial started automatically. Every 10th trial a drift correction was carried out to adjust for head movements.

3. RESULTS

Figure 1 shows fixation proportions to targets and competitors from the critical word pairs. Solid lines represent fixations from the evidence group, dashed lines represent the no-evidence group. The darker/thicker lines for each target and competitor indicate words with correct initial stress, the lighter/thinner lines indicate words with incorrect initial stress. The vertical lines show the two time windows of analysis. The first time window (T1) spans the time from 200 ms after target onset to a word pair's segmental divergence point shifted by 200 ms. 200 ms is the earliest point in time that is related to the processing of the target word [7]. The second time window (T2) spans from the segmental divergence point to the point in time with the highest proportion of target fixations across all conditions.

Statistical analyses were carried out using linear mixed-effects models [1]. The dependent variable was the difference between logistically transformed fixation proportions to targets and competitors. Participant and item were entered as random factors. Listener group (no-evidence group: -0.5; evidence group: 0.5), stress location (correct: -0.5; incorrect: 0.5) and their interaction were entered as fixed factors. Factors were coded such that a significant intercept indicates listeners' preference to fixate the target over the competitor. P-values were based on Markov chain Monte Carlo sampling.

Figure 1: Fixation proportions over time to target and competitor from acoustic target onset. Vertical lines indicate boundaries for the time windows of analyses.



During T1 listeners looked already more at the target than the competitor, indicating that they used lexical stress information to distinguish between segmentally overlapping words ($b_{intercept}$=0.225; $p_{(MCMC)}$<.05). The main effect of stress location shows that words with correct initial stress were recognized better than words with incorrect stress (b_{stress}=-0.389; $p_{(MCMC)}$<.05). The main effect of group was not significant during T1 (b_{group}=0.111; $p_{(MCMC)}$=.55). This lack of an effect, however, was mediated by the interaction of stress location and listener group ($b_{group*stress}$=-0.711;

$p_{(MCMC)}$<.05). Table 3 shows results for separate analyses by listener group, which indicate that only listeners in the evidence group were in fact able to use stress information to distinguish the target from the competitor before segmental information became available.

Table 3: Separate analyses of target preference for each listener group.

	evidence group		no-evidence group	
	b	$p_{(MCMC)}$	*b*	$p_{(MCMC)}$
intercept	0.279	< .05	0.169	= .21
stress location	-0.745	< .01	-0.034	= .88

During T2, the time window immediately following the word pair's segmental divergence point, listeners showed a strong target preference ($b_{intercept}$=4.564; $p_{(MCMC)}$<.001). Listeners from the evidence group were better at recognizing the target than the no-evidence group (b_{group}= 0.548; $p_{(MCMC)}$<.01). The effect of better target recognition for correctly initially stressed words stayed significant within T2 (b_{stress}=-0.907; $p_{(MCMC)}$<.001). The interaction between group and stress location was not significant and therefore was eliminated from the final model.

4. DISCUSSION

The present experiment showed that native Dutch listeners are able to tune into a speaker's specific realization of suprasegmental cues to lexical stress. Listeners who during exposure received evidence of the speaker's non-canonical marking of unstressed initial syllables were able to use this information during test to distinguish between segmentally overlapping target-competitor pairs. Listeners were even able to distinguish target and competitor by means of their stress patterns alone prior to disambiguating segmental information being available (i.e., during T1). The group that only received information about the speaker's general pronunciation peculiarities (i.e., the global foreign accent) recognized the target later and suffered from stronger competition than the evidence group even after the target and competitor could be disambiguated by segmental information (during T2).

The finding about the early use of stress information by the evidence group extends prior findings ([9]), as it demonstrates Dutch listeners' reliance on suprasegmental information even in situations where acoustic information may generally be less reliable as is the case in foreign-accented speech. This holds as long as the listeners have evidence through exposure of how the relevant suprasegmental cues are implemented.

The present findings suggest one of two things: Either listeners learned that the speaker produced correctly and incorrectly stressed initial syllables with different stress cues (i.e., duration vs. pitch) or that word-initial *unstressed* syllables are marked by high pitch. Whichever mechanism is underlying the adaption, the results show that Dutch native listeners are able to adapt to foreign-accented lexical stress. This adaptation occurs rapidly and can be applied during early phases of the word recognition process.

5. REFERENCES

[1] Baayen, H., Davidson, D.J., Bates, D.M. 2008. Mixed-effect modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390-412.

[2] Boula de Mareüil, P., Vieru-Dimulescu, B. 2006. The contribution of prosody to the perception of foreign accent. *Phonetica* 63, 247-267.

[3] Bradlow, A.R., Bent, T. 2008. Perceptual adaptation to non-native speech. *Cognition* 106, 707-729.

[4] Braun, B., Lemhöfer, K., Mani, N. 2011. Perceiving unstressed vowels in foreign-accented English. *J. Acoust. Soc. Am.* 129, 376-387.

[5] Clarke, C.M., Garrett, M.F. 2004. Rapid adaptation to foreign-accented English. *J. Acoust. Soc. Am.* 116, 3647-3658.

[6] Eisner, F., Weber, A., Melinger, A. 2010. Generalization of learning in pre-lexical adjustments to word-final devoicing (Abstract). *J. Acoust. Soc. Am.* 128, 2323.

[7] Hallett, P.E. 1986. Eye movements. In Boff, K.R., Kaufman, L., Thomas, J.P. (eds.), *Handbook of Perception and Human Performance*. New York: Wiley, 10-11, 10-112.

[8] Maye, J., Aslin, R.N., Tanenhaus, M.K. 2008. The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science* 32, 543-562.

[9] Reinisch, E., Jesse, A., McQueen, J.M. 2010. Early use of phonetic information in spoken word recognition: Lexical stress drives eye-movements immediately. *Q. J. Exp. Psychol.* 63, 772-783.

[10] Sidaras, S.K., Alexander, J.E.D., Nygaard, L.C. 2009. Perceptual learning of systematic variation in Spanish-accented speech. *J. Acoust. Soc. Am.* 125, 3306-3316.

[11] Sluijter, A.M.C., van Heuven, V.J. 1995. Effects of focus distribution, pitch accent, and lexical stress on the temporal organization of syllables in Dutch. *Phonetica* 52, 71-89.

[12] White, L., Mády, K. 2007. Heads to lose: Prosodic structure and timing in Hungarian. *Proc. Laboratory Phonology* 11, 159-160.

[13] Witteman, M.J., Weber, A., McQueen, J.M. 2010. Rapid and long-lasting adaptation to foreign-accented speech (Abstract). *J. Acoust. Soc. Am.* 128, 2486.