Adapting to suprasegmental lexical stress errors in foreign-accented speech

Eva Reinisch[a][1], Andrea Weber[a][2]

[a]Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, The

Netherlands

running head: Adapting to foreign-accented stress marking

Journal of the Acoustical Society of America

Corresponding author

Eva Reinisch

Carnegie Mellon University

Department of Psychology

5000 Forbes Avenue

Pittsburgh, PA, 15213

USA

E-mail: evarei@andrew.cmu.edu

Phone: +1-412-268-2656

Abstract

Can native listeners rapidly adapt to suprasegmental mispronunciations in foreign-accented speech? To address this question, an exposure-test paradigm was used to test whether Dutch listeners can improve their understanding of non-canonical lexical stress in Hungarian-accented Dutch. During exposure, one group of listeners heard a Dutch story with only initially stressed words, while another group also heard twenty-eight words with canonical second-syllable stress (e.g., *EEKhorn*, "squirrel" was replaced by *koNIJN* "rabbit"; capitals indicate stress). The twenty-eight words, however, were non-canonically marked by the Hungarian speaker with high pitch and amplitude on the initial syllable, both of which are stress cues in Dutch. After exposure, listeners' eye movements were tracked to Dutch target-competitor pairs with segmental overlap but different stress patterns, while they listened to new words from the same Hungarian speaker (e.g., *HERsens, herSTEL,* "brain", "recovery"). Listeners who had previously heard non-canonically produced words distinguished target-competitor pairs better than listeners who had only been exposed to Hungarian accent with canonical forms of lexical stress. Even a short exposure thus allows listeners to tune into speaker-specific realizations of words' suprasegmental make-up, and use this information for word recognition.

186/200 words

I. INTRODUCTION

It usually takes little effort for listeners to understand what is being said in their native language. Listening to foreign-accented speech can make this task considerably harder, even in quiet listening conditions. Foreign-accented speech deviates from native speech at multiple levels. Besides syntactic and lexical errors, the most prominent difference lies in the speakers' nonnative pronunciation of words and utterances. This includes the articulation of the segments as well as the suprasegmental properties of speech rhythm and intonation patterns. The specific deviations thereby often reflect linguistic properties of the speaker's mother tongue (e.g., Swan & Smith, 2001). For example, Dutch speakers often pronounce syllable-final voiced obstruents in English as voiceless (e.g., *seed* produced as *seat*) since syllable-final devoicing is a phonological process in Dutch. As a result, English listeners find Dutch speakers hard to understand (Eisner, Weber, & Melinger, 2010). Similarly, lexical stress in Dutch is marked by duration, pitch, and spectral tilt, and Dutch speakers use the same suprasegmental cues to mark stress in English rather than vowel reduction as native English speakers do. This too causes difficulties for English listeners (Braun, Lemhöfer, & Mani, 2011). Luckily, however, understanding a nonnative speaker can become easier over time. In the present study, we empirically investigate to what extent native listeners can adapt to foreign-accented speech. The focus is on the role of suprasegmental lexical stress errors during the recognition of spoken words. Effects of lexical stress errors are assessed in an eye-tracking study with Hungarian-accented Dutch.

Importantly, there is more than anecdotal evidence showing that comprehension of foreign-accented speech can improve with listening experience (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras, Alexander, & Nygaard,

2009). For instance, Clarke and Garrett (2004) showed rapid adaptation to foreign-accented English in a cross-modal matching task. That is, lexical decision times for visually presented English words were faster when the word matched the last word of an auditorily presented sentence in comparison to mismatching words. After as little as sixteen sentences this facilitatory matching effect was equally strong for sentences spoken with or without a foreign accent.

It is generally assumed that adaptation to foreign accents is lexically guided, just as adaptation to segmental variations is in native speech (e.g., Norris, McQueen, & Cutler, 2003; see Samuel & Kraljic, 2009 for an overview). That is, native listeners' knowledge of how existing words are pronounced (and in which sentence contexts they usually occur), allows them to compensate for non-canonical pronunciations by adjusting phoneme perception for particular speakers. Usually, when listening to a foreign accent, a large number of such adaptations is required. Therefore, adaptation to a less intelligible nonnative speaker takes more time. Lexical information is harder to access due to the larger acoustic mismatch between canonical forms and the speaker's realizations (Bradlow & Bent, 2008). Nevertheless, adaptation occurs regardless of a talker's baseline intelligibility.

Note that all of the above studies on foreign-accented speech focused on global foreign accent, and most of them used offline sentence- and word-transcription tasks (but see Clarke & Garrett, 2004 who used a speeded word matching task). Only recently, a few foreign accent studies have helped to draw a more detailed picture of which segmental deviations affect online word recognition, and how native listeners can adapt to them (e.g., Eisner, et al., 2010; Sidaras, et al., 2009). For example, Eisner et al. (2010) found that after exposure to a Dutch speaker producing devoiced word-final stops in English, English listeners showed evidence of adaptation in a subsequent

cross-modal priming task. Auditory primes with voiceless final stops (e.g., *seat*), facilitated recognition of visual targets with voiced final stops (e.g., *seed*). Additionally, Sidaras et al. (2009) found correlations between the acoustic realization of vowels in Spanish-accented English and the amount of English listeners' improvement in understanding words containing these vowels.

Foreign accents, however, manifest themselves not only at the segmental level but also at the suprasegmental level in form of non-canonical realizations of intonation contours and lexical stress patterns (e.g., Boula de Mareüil & Vieru-Dimulescu, 2006; Braun, et al., 2011; Jilka, 2000). Jilka (2000) found a large contribution of prosodic factors in accentedness ratings for German-accented English sentences and for English-accented German sentences. Moreover, Boula de Mareüil and Vieru-Dimulescu (2006) used a prosody manipulation technique to investigate the contribution of intonation in a language detection task. Spanish and Italian sentences with widely overlapping segmental content were manipulated such that the intonation contour of one language was imposed on the other language (e.g., Italian sentence with Spanish prosody). Native participants of both languages listened to these sentences and judged the native language of the speaker. Results showed that listeners relied on prosody rather than the speaker's voice and segmental information to make their judgments.

Foreign-accented prosody not only increases the perceived strength of a foreign accent but also slows down the word-recognition process. For example, as Braun, Dainora, and Ernestus (2011) have shown, an unfamiliar intonation contour in otherwise native-like produced Dutch sentences interferes with lexical access. Their results from multiple experimental tasks (word-monitoring, and cross-modal identity priming with a lexical decision or semantic judgment task) suggest that adverse

effects of suprasegmental foreign accent occur at lexical and semantic processing levels. Furthermore, Braun, Lemhöfer, et al. (2011) found that word stress as produced by Dutch learners of English adversely affects word recognition for English listeners. The lack of the typical English vowel reduction in unstressed syllables prohibited priming effects for syllable fragments. For example, hearing the first syllable "ab-" facilitated recognition of the written word "absurd" when the syllable was pronounced with a reduced vowel as native English speakers do. The same syllable "ab-" spoken with a Dutch accent (i.e., without vowel reduction), however, did not facilitate recognition of written "absurd".

The study by Braun, Lemhöfer, et al. (2011) showed processing difficulties for native listeners when lexical stress was produced on the correct syllable but with incorrect acoustic cues. The main cue for native listeners in their study was segmental reduction. In contrast, the present study investigates effects of lexical stress errors by looking at *suprasegmental* stress cues realized on the wrong syllable of a word. Moreover, the present study tests whether native listeners are able to *adapt* to suprasegmental stress errors, as they are able to adapt to segmental mispronunciations.

Adaptation to suprasegmental information is not self-evident. Even though Dutch listeners have been shown to use lexical stress information to recognize words (see the literature review in section A below), most words in Dutch can be recognized by segmental information alone. Adaptation to segmental and suprasegmental cues could thus be restricted to cases where they provide lexical information, for example, in the case of suprasegmental information to lexical tones (Mitterer, Chen, & Zhou, 2011). Moreover, unlike in studies on segmental adaptation where phoneme categories can be adjusted, the previous exposure to stress errors may make listeners rely less on these cues since they learned that stress cues are not reliable.

Alternatively, exposure to stress errors may improve listeners' comprehension of foreign-accented speech, as does exposure to segmental errors. Since stress can be cued by multiple acoustic features (duration, pitch, RMS-amplitude), correctly and incorrectly stressed syllables could be marked by a different combination of these stress cues. The continued use of stress information in word recognition despite the perception of incorrect stress would then suggest the adaptation to fine phonetic detail in the realization of suprasegmental stress cues. If correctly and incorrectly stressed syllables are cued by the same acoustic features, it will likely not be possible to distinguish correctly and incorrectly stressed words by means of stress (as they share the same cues). But knowledge about the presence of errors in the speaker's accent may make it easier to *recover* from the misleading information and thus facilitate word processing for listeners who heard stress errors before. Adaptation to stress errors can thus provide detailed insight into the processing of suprasegmental information.

The use of Hungarian-accented Dutch provides a well-suited language combination for the present study. First, primary lexical stress in Dutch is not restricted to a specific syllable in a word (see e.g., van Heuven & Hagman, 1988, for frequency counts of stress locations in Dutch), so correct and incorrect stress can be compared in the same position in a word. Second, unlike in English where unstressed syllables are often reduced, lexical stress in Dutch is mainly marked suprasegmentally. Hence adaptation to suprasegmental foreign accent can be tested. Third, Dutch listeners use suprasegmental lexical stress for word recognition and are sensitive to lexical stress errors. Shifting stress towards the beginning of a word is especially detrimental for Dutch listeners (e.g., van Heuven, 1985; see section A below). Fourth, in Hungarian, stress always falls on the word-initial syllable. Since

properties of the native language are regularly transferred to the second language (e.g., Archibald, 2003, Swan & Smith, 2001), Hungarian learners of Dutch are expected to erroneously stress the first syllable in Dutch words even in cases where not appropriate (see e.g., Archibald, 1993, pp. 114ff, for findings on stress placement in Hungarian learners of English). Given these properties of the languages under investigation, it is possible to address the question of adaptation to suprasegmental stress errors with a naturally existing foreign accent.

A. Lexical stress in Dutch and Hungarian

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 lexical word stress in Dutch are suprasegmental and include duration, pitch, spectral tilt, and RMS amplitude. Primary stressed syllables are longer, louder, have a higher pitch and more energy in the high frequency bands than unstressed syllables (e.g., Nooteboom, 1972; Slujiter & van Heuven, 1995). Vowel reduction, which, for example, is the main stress cue in English, is less common in Dutch.

Importantly, Dutch listeners have been shown to use lexical stress information to recognize words (e.g., Reinisch, Jesse, & McQueen, 2010; van Donselaar, Koster, & Cutler, 2005; van Leyden & van Heuven, 1996). When guessing a word from its first syllable, Dutch listeners gave responses that matched the correct degree of stress for that syllable in 80% of the cases (van Leyden & van Heuven, 1996). Similarly, in a fragment priming task, Dutch listeners responded faster to the printed target "octopus" if they heard the segmentally and suprasegmentally matching fragment prime *OCto-* than if they heard a segmentally unrelated fragment *eufo-* (van Donselaar et al., 2005). A fragment-prime matching the target in segments but mismatching in

stress (*okTO-*), in contrast, had an inhibitory effect. That is, the target "octopus" was recognized more slowly when fragment prime and target mismatched in stress than when the target was preceded by a segmentally unrelated prime.

The time course of Dutch listeners' use of acoustic stress cues during word recognition was most clearly illustrated in a recent visual-world eye-tracking study (Reinisch et al., 2010). Using segmentally overlapping target-competitor word pairs such as *OCtopus – okTOber*, Reinisch et al. showed that Dutch listeners are able to distinguish these words by means of their stress patterns alone: when hearing the Dutch word "octopus", Dutch listeners already fixated the printed word "octopus" more than the competitor word "oktober" before they could have possibly processed any disambiguating segmental information (i.e., the /p/ in "octopus"). Dutch listeners thus use suprasegmental cues to lexical stress early during word recognition.

This sensitivity of Dutch listeners to lexical stress during spoken-word recognition also makes them susceptible to lexical stress errors. Moving suprasegmental stress cues in a word to the wrong syllable significantly slows down processing and increases incorrect identification of words (Cutler & Koster, 2000; Koster & Cutler, 1997; van Leyden & van Heuven, 1996; van Heuven, 1985). Koster and Cutler (1997), and Cutler and Koster (2000), report that erroneous stress placement is as detrimental to word recognition as are segmental mispronunciations. Van Heuven (1985), and van Leyden and van Heuven (1996), further suggest that an erroneous stress shift towards the beginning of the word is more detrimental to word recognition than is a stress shift towards the end of the word.

Hungarian, in contrast to Dutch, is a fixed-stress language where stress always falls on the word-initial syllable. Stress in Hungarian is therefore not informative for lexical distinctions. Hungarian speakers process stress automatically as illustrated by

the fact that stress violations trigger a mismatch negativity in electrophysiological recordings (Honbolygó, Csépe & Ragó, 2004). Hungarian native speakers are, however, not aware of stress differences. When asked to discriminate words that differ only in stress, Hungarian listeners show a perceptual "stress deafness", similar to speakers from other fixed-stress languages such as Finnish or French (Peperkamp, Vendelin, & Dupoux, 2010). The main acoustic stress correlate in Hungarian is considered to be intensity (Fónagy, 1958, cited from White & Mády, 2008). In addition, stressed (word-initial) syllables are landing sites for sentence-level pitch accents, making pitch a possible cue to stress (Siptár & Törkenczy, 2000, pp. 21-26). Lengthening of stressed syllables, however, even when they carry a pitch accent, is not observed in Hungarian (White & Mády, 2008), possibly because vowel length is phonemically distinctive (Mády, Bombien, & Reichel, 2008).

B. The present study

Given these differences between lexical stress in Hungarian and Dutch, and the differences in how speakers of these languages process stress in their native languages, it can be expected that Hungarian speakers of Dutch produce lexical stress errors and that Dutch listeners may encounter difficulties in processing these mispronounced words. The aim of the study was therefore to shed light on how Dutch listeners cope with incorrect word-initial stress in Hungarian-accented Dutch. In particular, we investigated whether a short exposure to the specific realization of lexical stress is sufficient to use non-canonically produced stress cues during word recognition. In this way, findings on listeners' flexibility in handling segmental variations will be extended to the adaptation to foreign-accented suprasegmental information.

In an exposure-test paradigm, native Dutch listeners were first exposed to the speech of a Hungarian learner of Dutch. One group of listeners heard a Dutch story with only initially stressed words during exposure, while another group also heard Dutch words with unstressed initial syllables. The Hungarian speaker, however, marked unstressed initial syllables with high pitch and RMS-amplitude, which are both stress cues in Dutch. The latter listener group was thus familiarized with the speaker's specific stress errors. Subsequently, listeners' recognition of a new set of words was assessed. All words were spoken by the same speaker. An eye-tracking paradigm similar to Reinisch et al. (2010) was implemented in order to track the word recognition process over time.

Eye tracking uses the fact that listeners spontaneously fixate on visually presented objects or printed words that match listeners' current hypotheses about what is being said (Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974; Huettig & McQueen, 2007; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). That is, as the acoustic signal unfolds, listeners entertain hypotheses about possible words being said and modulate these hypotheses when new acoustic information becomes available. Since eye movements can be continuously recorded, it is possible to monitor how listeners change their lexical hypotheses over time. The eye-tracking paradigm has previously been shown to be sensitive enough to reveal listeners' use of fine phonetic detail during spoken-word recognition (e.g., Reinisch, Jesse, & McQueen, 2011; Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006), including the use of suprasegmental cues to lexical stress in Dutch (Reinisch, et al., 2010).

In the present study, listeners' eye movements to Dutch target-competitor pairs were tracked. Targets and competitors overlapped segmentally on the first syllable but

differed in their canonical stress pattern (e.g., *HERsens - herSTEL*; "brains"- "recovery"). Additional target-competitor pairs were tested in which both words had either canonical stress on the initial syllable (e.g., *KAssa - CAssis* "cash register" - "cassis") or on the second syllable (e.g., *paNEEL - paNIEK* "panel" - "panic"; see Appendix A). Regardless of the canonical stress pattern, the speaker of the experiment always placed stress on the initial syllable, using duration to mark correctly stressed syllables, and pitch and amplitude to mark incorrect stress. The main questions addressed were whether Dutch listeners who had been exposed to the Hungarian speaker's specific stress errors before, (1) would be better at using this information during word recognition than listeners who had just been familiarized with the speaker's global foreign accent, and (2), would be able to use this information early during word processing, before disambiguating segmental information became available.

If exposure to stress errors facilitates word recognition, then listeners exposed to stress errors should be better at recognizing targets in cases where canonical stress distinguishes target and competitor (*HERsens - herSTEL*). By using an eye-tracking paradigm with segmentally overlapping target-competitor pairs, any adaptation effects during a time window spanning these overlapping parts would suggest an early locus of the adaptation effect during lexical processing. Later or longer lasting effects of exposure would support a more general adaptation to stress errors. In that case, targets in the same stress pairs with second syllable stress (*paNEEL - paNIEK*) may also be recognized more easily by experienced listeners. By definition, these words cannot be distinguished by means of their stress patterns, but experienced listeners may recover more easily from stress errors than listeners who were not exposed to stress errors beforehand. Finally, if exposure to additional accent features aids the recognition of

the speaker's words independent of the specific cues at stake (here: the presence of stress errors), then a facilitation would be expected, even in the recognition of stress pairs with correct initial-syllable stress (KAssa - CAssis).

II. METHOD

A. Participants

Sixty native speakers of Dutch (mean age 21), who were raised monolingually, participated for a small payment. At the time of testing they were enrolled as students at Radboud University in Nijmegen. They reported no hearing problems and had normal or corrected-to normal vision. None of the participants spoke any Hungarian or was familiar with the Hungarian accent.

B. Materials

Two versions of a short Dutch story about a bear in the forest were created for exposure (spoken duration approximately 2.3 minutes). In one version of the story, only monosyllabic words or words with initial stress occurred ("no-evidence story"). In a second version ("evidence story"), twenty-eight words of the no-evidence story were replaced with semantically fitting words that have stress on the second syllable (e.g., *EEKhorn* "squirrel" became *koNIJN* "rabbit").

For the eye-tracking study, thirty-two bisyllabic Dutch word pairs were selected as critical stress pairs. None of the words had occurred in the story. In critical stress pairs, the two words overlapped segmentally up to the first phoneme of the second syllable but contrasted in their canonical stress pattern. One word had stress on the first syllable, the other had stress on the second syllable (e.g., *HERsens - herSTEL*; see Appendix A). All words had a full vowel in their initial syllable. Another thirty-two target-competitor pairs with segmental overlap and matching stress patterns were selected. These will be referred to as same stress pairs. In sixteen of the same stress

pairs both words had stress on the first syllable (e.g., *KAssa - CAssis* "cash register" - "cassis"), and in the other sixteen pairs both words had stress on the second syllable (e.g., *paNEEL - paNIEK* - "panel" - "panic"; see Appendix A). The two members of all target-competitor pairs served alternatively as target and competitor, that is, both members were presented visually on the screen, and each one was presented auditorily as target to half of the participants.

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. Target-competitor pairs were matched on their log-transformed CELEX lexical frequency (Baayen, Piepenbrock, & Gulikers, 1995), such that for the critical stress pairs initially stressed words and non-initially stressed words did not differ from each other in mean frequency ($t(31)=-0.17$, $p=.86$). Target-competitor pairs for each participant (half initially, half not initially stressed) were also matched for frequency ($t(63)=-1.62$, $p=.11$). Distractor pairs were slightly less frequent than target-competitor pairs ($t(63)=2.05$, $p<.05$).

A female Hungarian learner of Dutch recorded both versions of the story and multiple tokens of all target-competitor pairs embedded in the sentence *Klik op het woord [TARGET]* ("Click on the word [TARGET]"). Words were recorded in pseudo-random order such that members of target-competitor pairs did not directly follow one another to prevent intentional contrastive production of correct and incorrect stress. Word stress was not particularly marked in the reading list. At the time of recording the speaker was twenty-nine years old and had been living in the Netherlands for six years (attending Dutch university courses during the sixth year). 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. Instructions about stress placement were necessary to obtain consistent recordings with initial stress on all items. After all, language learners are sometimes aware of their errors, and in a formal recording situation may try to suppress them as much as possible. This, however, is not what we wanted to achieve. By instructing the speaker to not suppress her Hungarian accent we were able to elicit a naturally existing accent (as opposed to the artificial manipulation of specific stress cues) under relatively controlled conditions.

Two phonetically trained native Dutch speakers listened to all recorded tokens and judged the perceived stress location irrespective of the correct stress pattern. Tokens used for the eye-tracking task were selected on the basis of those judgments to ensure that all words were clearly perceived as having word-initial stress.

C. Acoustic Measures

Acoustic measurements of the stimulus material were made to examine which acoustic cues our speaker used to indicate stress and whether there were differences in her realization of correctly and incorrectly stressed words. Duration, pitch, and RMS-amplitude were measured on the vowels of the first syllables of all target-competitor pairs from the critical stress pairs in the eye-tracking study (e.g., *HERsens - *HERstel*; incorrect stress is marked with an asterisk), as well as of the critical words from the story that differed between listener groups (e.g., *EEKhorn* vs. *KOnijn*). In addition, the first vowels of twenty-eight bisyllabic words that occurred in both versions of the exposure story were measured as a control (e.g., *DIEren* "animals" - these were the same words for both groups). Measuring words from the story should ensure that, during exposure and test, the acoustic differences between correctly and incorrectly stressed words were comparable and that the stress cues for words that occurred in both versions of the story did not differ from each other.

Measurements were taken only on the vocalic portion of the syllable where regular glottal pulsing occurred. All measurements were taken automatically using Praat software (version 5.1, Boersma & Weenink, 2009). Pitch points were located between 70 Hz and 400 Hz and hand-corrected where octave-jumps occurred. To ensure that our speaker had indeed marked stress suprasegmentally rather than by vowel quality differences, the first and second formants of the vowels in the critical stress pairs from the eye-tracking task (e.g., *HERsens - *HERstel*) were measured in addition to the suprasegmental measurements. Using Praat standard formant-detection settings for female speakers, the first and second formants were extracted at the midpoints of the vowels. Paired *t*-tests comparing correctly and incorrectly stressed words confirmed that words of the critical stress pairs in the eye-tracking task did not differ in vowel quality (the mean difference in F1 was 17 Hz, $t(30)= 0.911$, $p=.37$; the mean difference in F2 was 2 Hz, $t(30)=-0.053$, $p=.96$). Stress was thus implemented suprasegmentally and not by a difference in vowel quality.

Table 1 shows duration, pitch, and RMS-amplitude values for the two sets of words from the exposure story: the twenty-eight critical word pairs which differed in (correctness of) stress location between groups (e.g., *EEKhorn - *KOnijn*) and twenty-eight control words which were the same for both groups (e.g., *DIEren*). Incorrectly stressed words in the evidence story were marked by a relatively higher pitch and RMS-amplitude 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 statistical comparison of these measurements cannot be interpreted straightforwardly since these words were not matched on their segmental make-up (e.g., *EEKhorn* vs. *koNIJN*). Little difference in either, duration, pitch, or RMS-amplitude was found for

measurements of the control words that were correctly stressed on their initial syllables in both versions of the story. This ensured that no unintended acoustic differences between groups were present during exposure.

Table 2 further shows comparisons of duration, pitch, and RMS-amplitude on the words' initial syllables for targets and competitors of the critical stress pairs from the test phase (i.e., the eye-tracking task). Here direct comparisons of acoustic measures could 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 and larger RMS-amplitude. Correct and incorrect stress cues were thus consistent in the exposure and the test phase.[3]

The acoustic differences between the first vowels of *HERsens* and **HERstel* are the result of using a freely produced (i.e., not artificially manipulated) accent for the present investigation. Our speaker was instructed to produce initial-syllable stress on all words as she would naturally do in Hungarian. But since there are no comparable production data of Hungarian-accented Dutch available, it can not be determined whether the difference between correct and incorrect initial stress is representative of Hungarian accent in general or just of our speaker. Importantly, however, for the aim of exploring adaptation to suprasegmental errors, we can benefit from these acoustic differences. As hypothesized in the introduction, listeners may not only learn about the presence of stress errors during exposure but rather adapt to the specific distribution of stress cues on correctly and incorrectly stressed syllables. In this way we can test the adaptation to fine phonetic detail on a suprasegmental level.

D. Design and Procedure

Participants were seated 80 cm in front of a 40.5 cm by 30.5 cm computer screen. Eye movements were recorded with an SR Research Eyelink 1000 system at a sampling rate of 1000 Hz. Stimulus presentation was controlled by Experiment Builder software (SR Research). Auditory stimuli were presented over headphones at a comfortable listening level. During exposure, half of the participants heard the story with only initially stressed words (no-evidence group), while the other half heard the story including twenty-eight non-initially stressed words (evidence group). The latter group thus received evidence that the speaker stressed all words on the initial syllable. During exposure, participants simply listened and viewed a fixation cross. The following visual-world task was the same for everyone. On every trial, listeners first saw a fixation cross for 800 ms. Then four printed words (i.e., target, competitor, and the two distractors) were displayed in four quadrants of the screen in mono-spaced, lower case, Lucida Console font, size 35. At the same time, participants heard the sentences instructing them to click with the computer mouse on one of the words (i.e., the target). The start of the audio was timed such that the preview of the printed words was always 2000 ms relative to acoustic target onset. Listeners had thus ample time to read the words in the display and access the phonological representations including the canonical stress patterns before the target word was heard. In this way possible orthographic or other visually mediated effects could be diminished (see e.g., Huettig & McQueen, 2007; Salverda and Tanenhaus, 2010).

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. The printed words disappeared from the screen 1800 ms after acoustic target onset or at listeners' response, and the next trial

started automatically 1000 ms after the response. Every tenth trial a drift correction was carried out to adjust for head movements.

III. RESULTS

A. Analyses

Only trials on which participants clicked on the correct word were analyzed. For this reason, 79 trials (2.06 %) had to be excluded. About half of the incorrect clicks (39) were on the competitor; the remaining incorrect clicks (40) fell outside predefined areas within which clicks and fixations were counted as being on a word. This area spanned a circle of diameter 12 cm around the center of each word. Further, all trials containing the word pair *BAlie - balLET* ("counter"-"ballet") were excluded from all analyses as the authors erroneously selected this pair despite different vowel phonemes on the initial syllables. The two phonemes spelled "a" differ not only in the degree of stress but also in phonemic vowel quality and quantity. Eye-movement data were first categorized in 4 ms bins and were then analyzed in two larger time windows. The first time window (T1) spanned the time from 200 ms after target onset to 200 ms after each word pair's segmental divergence point. The time that is commonly used as estimate for the interval needed to program and launch a saccade is 200 ms (see, e.g., Hallett, 1986; Matin, Shao, & Boff, 1993). Thus 200 ms after target onset is the earliest point in time that is related to the processing of the target word. During T1, any preference for the target over the competitor word is likely to be driven by differences in the realization of correct and incorrect stress since disambiguating segmental information could not yet have been processed. T1 thus allows us to test effects of exposure on early phases of the word-recognition process. The second time window (T2) spanned from the segmental divergence point shifted by 200 ms to the point in time where the highest proportion of target fixations across

all conditions occurred (i.e., at 1084 ms). At this point in time listeners should have recognized the target. T2 thus allowed us to test longer lasting effects of processing stress information in the presence of segmentally disambiguating information. It is very common in eye-tracking studies that effects last longer than the duration of segmental overlap (see e.g. frequency effects in Dahan and Gaskell, 2007); effects need time to build up and also to disintegrate. Effects of group at T2 would suggest that listeners in the evidence group are better at recovering from the encounter of stress errors than the no-evidence group.

Statistical analyses were carried out separately for the critical stress pairs and the two same-stress conditions. Linear mixed-effects models were used as provided in the lme4 package (Bates & Sarkar, 2007) in R (version 2.10.0; The R foundation for statistical computing). The dependent variable was the difference between logistically transformed fixation proportions to targets and competitors (see e.g., Barr, 2008, and Jaeger, 2008, for discussions on the benefits of transforming proportion data). *Participant* and *item pair* were entered as random factors. This allowed the intercept of the regression models to vary by participant and item pair with the restriction that the mean of this random variation was zero (see Baayen, Davidson, & Bates, 2008). *Listener group* (no-evidence group, coded as -0.5; evidence group coded as 0.5; between participant factor) was entered as a fixed factor in all analyses. For the critical stress pairs the additional factor *stress* (correct: 0.5; incorrect: -0.5; within participant factor) and its interaction with *listener group* were entered. All factors were coded such that the overall mean was mapped onto the intercept. As a consequence of this coding, a significant positive intercept indicates listeners' preference to fixate the target over the competitor. Regression weights assigned to each of our binary factors indicate how the two levels of each factor vary around the

intercept. A positive regression weight for the factor *stress*, for instance, indicates that the preference for the target over the competitor (i.e., our dependent variable) is larger for the level of the factor mapped onto 0.5 (i.e., correct stress). The exact amount of target over competitor preference for the correct stress condition is then given by the value of the intercept plus one half of the absolute value of the regression weight for the factor *stress*. A significant effect can be inferred if a regression weight is statistically different from zero. Due to the contrast coding of factor levels, a significant effect translates to a main effect for this factor. If, in the analysis of the critical stress pairs, the interaction term was not significant, the model was re-fitted without interaction. Models were then compared with a log-likelihood ratio test. If the simpler model fit the data better, this model will be reported. P-values were based on Markov chain Monte Carlo sampling, using 10,000 runs.

1. *Critical stress pairs*

Figure 1 shows fixation proportions over time to correctly and incorrectly stressed targets and competitors from the critical stress pairs for both listener groups. Fixations on the distractors are not plotted in order to keep fixations on the target and competitor distinct and visible. Fixations on distractors, however, were comparable to target and competitor fixations between 0 and 200 ms after target onset, and steadily decreased once acoustic information of the target-competitor pairs was being processed. The vertical lines in Figure 1 show the two time windows of analysis. The temporal boundaries of the first time window (T1) were normalized for each word and plotted within the average segmental divergence point at 553 ms after target onset (including the 200 ms fixation lag).

Statistical analyses show that during T1, Dutch listeners already looked more often at the target than the competitor, indicating that they used lexical stress

information to distinguish between segmentally overlapping words. As the dependent variable is a measure of target preference and all fixed factors were contrast coded, this effect is shown by the positive regression weight of the intercept term, which is significantly different from zero ($b_{\text{intercept}}$=0.224, SE=0.09, t=2.46, $p_{\text{(MCMC)}}$<.05). That is, averaged across exposure groups, listeners used acoustic differences in correctly and incorrectly stressed words to distinguish the target from its competitor during early phases of the word recognition process (see also Reinisch et al., 2010). However, this target preference was modulated by the experimental manipulations. The main effect of *stress* shows that words with correct initial stress received more fixations than words with incorrect stress (b_{stress}=0.373, SE=0.18, t=2.05, $p_{\text{(MCMC)}}$<.05). While the main effect of *listener group* was not significant (b_{group}=0.144, SE=0.18, t=0.79, $p_{\text{(MCMC)}}$=.44), there was a marginally significant interaction between *stress* and *listener group* ($b_{\text{group*stress}}$=0.647, SE=0.36, t=1.78, $p_{\text{(MCMC)}}$=.07). The statistical model including the interaction also fitted the data marginally better than the model without the interaction ($\chi^2_{\text{(df=1)}}$ = 3.16, p = .075, log-likelihood for model with interaction = -5047.7; without interaction = -5046.1). Given the overall fixation preference for targets over competitors, we decided to follow up on this interaction even though it just failed to reach the conventional p<.05 threshold. We reasoned that results about the underlying patterns for each listener group can further inform us about the process of adaptation to stress errors. However, some caution in the interpretation of these results is warranted.

Table 3 shows the results of these additional analyses, which indicate that only listeners in the evidence group were able to use stress information to distinguish the target from the competitor before disambiguating segmental information became available (cf. significant intercept term for the evidence group; Table 3). It was also

the evidence group that carried the effect of stress with more target fixations occurring for correctly stressed words (main effect of stress; Table 3). Specifically, words with correct initial stress (*HERsens*) could be distinguished from their non-initially stressed competitors based on the stress information alone. Note that participants encountered the competitor (*herSTEL)* only in its written form, hence no match between the spoken target and the competitor with canonical second-syllable stress was expected. In contrast, targets with incorrect initial stress (**HERstel*), could not be differentiated from their initially stressed competitors (*HERsens*).[4] Neither an effect of target preference nor stress was found for listeners in the no-evidence group (see Table 3).

During T2, the time window immediately following the word pair's segmental divergence point, listeners showed overall a strong target preference ($b_{intercept}$=4.555, SE=0.13, t=35.90, $p_{(MCMC)}$<.001). Listeners from the evidence group, however, were better at recognizing targets than listeners from the no-evidence group (b_{group}= 0.557, SE=0.21, t=2.67 $p_{(MCMC)}$<.01). Moreover, correctly stressed words were recognized better than incorrectly stressed words (b_{stress}= 0.891, SE=0.16, t=5.63, $p_{(MCMC)}$<.001).

In order to test whether listeners were able to further adapt to the speaker's stress errors in the course of the test phase, *experiment half* (first half, coded -0.5, second half, coded 0.5) was added as a fixed-factor to the statistical models. Model comparisons between models in which non-significant interactions and factors were successively eliminated showed that during T1 the best fitting model is the one reported above (i.e., including *stress, listener group,* and their interaction). *Experiment half* did not interact with *listener group* or *stress* nor did it show a main effect. Table 4 reports statistics for the model comparisons.

The same model-fitting procedure was applied for listeners' fixations during T2. Here the two-way interaction between *experiment half* and *stress* was significant

and therefore stayed in the final model ($b_{\text{intercept}}$=4.555, SE=0.13, t=36.17, $p_{\text{(MCMC)}}$<.001; b_{group}=0.558, SE=0.21, t=2.70, $p_{\text{(MCMC)}}$<.01; b_{stress}=0.884, SE=0.16, t=5.59, $p_{\text{(MCMC)}}$<.001; b_{half}=0.082, SE=0.16, t=0.58, $p_{\text{(MCMC)}}$=.61; $b_{\text{stress*half}}$=-0.745, SE=0.32, t=-2.31, $p_{\text{(MCMC)}}$<.05). The effect of *stress* (correctly stressed targets are recognized better than incorrectly stressed targets) was larger in the first than in the second half of the experiment. No interaction, however, was found between *experiment half* and *listener group*. That is, during the course of the experiment, the effect of stress changed equally for both listener groups.

The results for the critical stress pairs showed that lexical stress errors hindered word recognition for the no-evidence group, but that brief exposure to the specific errors allowed listeners in the evidence-group to adapt to the non-canonical realization of lexical stress and to use this information for word recognition. While the effect was only marginally significant during T1 it turned into a robust effect during T2. Exploring possible adaptation effects in the course of the test phase showed that the effect of stress became smaller over the experiment but did so similarly for both listeners groups and only during T2.

2. Same stress pairs

Figure 2 shows fixation proportions to targets and competitors for the same-stress pairs. Statistical analyses reported in Table 5 indicate that during T1 participants could not distinguish between targets and competitors, neither in pairs with correct initial stress nor in pairs where initial stress was incorrect. Furthermore, there was no effect of *listener group* during T1 (see Table 5). During T2, however, when segmental information became available, listeners fixated the target more than the competitor. Moreover, for pairs with incorrect initial stress (i.e., the words that should have been but were not stressed on their second syllable) a significant effect of *listener group*

was found (see Table 5). Participants from the evidence group showed a stronger target preference than participants from the no-evidence group. For the stress pairs with correct initial stress no significant difference between groups was found even during T2 (Table 5). *Experiment half* did not affect the recognition of either type of same stress pairs and therefore will not be discussed further.

Even though the same stress pairs could not possibly be distinguished by means of their stress patterns they allowed us to further specify the level at which adaptation to stress errors occurs: (1) at a very specific level where stress is relevant to distinguish the target from the competitor (no effect expected), (2) at an "intermediate" level wherever stress errors occur (effect for incorrectly stressed pairs expected), or (3) at a very global level such that more information about an accent facilitates word recognition in general (effect for all word pairs expected). As in the second case above, listeners in the evidence group had adapted to lexical stress errors and used this information for the recognition of words in which these specific errors were relevant (i.e., for the same stress pairs where both words were incorrectly stressed). No advantage for the evidence group was found where stress errors were not relevant because the realization of word initial stress matched the canonical stress pattern of target and competitor.

IV. DISCUSSION AND CONCLUSIONS

The present study showed that native Dutch listeners are able to quickly tune into a Hungarian speaker's realization of lexical stress in Dutch. Dutch listeners who had heard the speaker's non-canonical marking of unstressed initial syllables before (i.e., the evidence group) were able to use this lexical stress information in a subsequent word recognition task. This is in contrast to the listeners in the no-evidence group who had only been exposed to the speaker's general pronunciation peculiarities (i.e., her

global foreign accent) but not to her lexical stress errors. Listeners from the no-evidence group suffered from stronger competition than listeners in the evidence group. This confirms earlier findings that lexical stress errors hinder word recognition in Dutch (e.g., Cutler & Koster, 2000; Koster & Cutler, 1997; van Heuven, 1985).

Differences in target recognition between the two listener groups appeared to develop over time. In the early time window (T1) where only suprasegmental lexical stress information could distinguish the target from the competitor, we found that for the critical stress pairs listeners recognized correctly stressed targets better than incorrectly stressed targets, but the group differences only started to build up. That is, the effect of listener group was visible only after following-up on a marginally significant interaction between *listener group* and *stress*, which calls for caution in the interpretation of the results. Nevertheless, this hint at an emergence of group differences during T1 was followed by a strong and lasting group effect during T2.

Importantly, during T2 group differences occurred not only for the critical word pairs but also for words from same stress pairs with incorrect stress. Listeners in the evidence group were better at recognizing the targets for the target-competitor pairs that differed in correctness of stress (*HERsens* - **HERstel*), as well as for word pairs where target and competitor mismatched in their canonical stress pattern with the acoustic input (**PAneel* – **PAniek*). Familiarization with a speaker's specific realization of lexical stress cues can thus aid word recognition even when lexical stress cannot disambiguate the target word, possibly by helping the listener to recover more quickly from the misleading stress information. Only when the target's and the competitor's canonical stress pattern matched the acoustic input (*KAssa* – *CAssis*) no advantage was found for listeners in the evidence group. Additional information about a speaker's lexical stress errors did not lead to a better comprehensibility for this

speaker in general, but only helped where this specific additional information was relevant, that is, where stress errors occurred.

So what did listeners in the evidence group learn? They were presented with natural recordings in which the Hungarian speaker happened to mark correctly and incorrectly stressed syllables with different suprasegmental cues. In particular, incorrectly stressed initial syllables were marked with higher pitch and stronger RMS-amplitude than correctly stressed syllables but they had a relatively shorter duration than correctly stressed syllables. Listeners in the evidence group could have either learned that words such as *HERstel* are being incorrectly stressed on the initial syllable by the Hungarian speaker, or they could have learned to interpret high pitch in connection with a relatively short duration as a marker for unstressed initial syllables. Listeners of both exposure groups, however, were better at distinguishing the correctly stressed targets from their competitor. This suggests that even for listeners in the evidence group, high pitch served as a cue for stress, delaying the recognition of incorrectly stressed words. A relatively higher pitch in combination with short duration thus appears not to be learned as a marker of unstressed syllables. If in the present study acoustic markers had been learned as cues to unstressed syllables, a better performance would have been expected on the recognition of words with canonical second-syllable stress. That is, the competitor with canonical stress on the initial syllable should not have been an equally good match to the acoustic input as the target with the marked unstressed initial syllable.[5] However, this is not what was found. It is therefore more likely that listeners in the evidence group had adapted to stress errors. A general adaptation to stress errors is also in line with the long lasting effect of the group differences suggesting that indeed exposure to a speaker's stress errors helps to recover from perceived pronunciation errors.

With regard to previous studies on the adaptation to segmental mispronunciations, the adaptation to lexical stress errors is likely to work in a similar fashion. Importantly, adaptation to stress errors occurs just as rapidly. Hearing less than thirty tokens of non-canonical stress was sufficient for listeners to use this information for the recognition of words, notably ones they had not heard during exposure. As for the lexical guidance of the adaptation process, ample lexical information is available for the re-interpretation of lexical stress cues. Although Dutch listeners use lexical stress information to recognize words whenever possible, in most cases segmental information is sufficient to recognize words. This could have been a reason not to adapt to stress errors. It is, however, this segmental information that allows listeners to identify and adapt to stress errors.

The present study adds a piece to the puzzle about how Dutch listeners flexibly adapt to variation in the speech signal. Unlike segmental information, stress patterns are distributed over the syllables of a word. Thus, even though the present study was restricted to the case of incorrectly stressed word-initial syllables caused by a stress-shift from the second to the initial syllable, the findings suggest that listeners can adapt to foreign accent on a non-segmental (i.e., word) level. Note, however, that for the adaptation to *suprasegmental stress* errors one restriction will apply: Listeners of the target language have to use suprasegmental cues to identify stressed syllables in their native language. No adaptation, for example, would be expected in English where vowel reduction is a stronger maker of stressed vs. unstressed syllables than suprasegmental cues (e.g., Fear, Cutler, & Butterfield, 1995). What remains to be shown, however, is whether adaptation to suprasegmental information also occurs on larger "units" than the word. Braun, Dainora, et al. (2011) demonstrated processing costs for word recognition in sentences with non-native intonation contours. To

optimally tune into non-native speech listeners would also have to adapt to more general prosodic and rhythmic characteristics of the speaker or the accent. The present study is a first step in assembling the larger picture.

In summary, the present study has shown that Dutch listeners are able to adapt to lexical stress errors in Hungarian-accented speech. In particular, they are sensitive to the distribution of suprasegmental stress cues on correctly and incorrectly stressed syllables. Although incorrect cues to lexical stress hindered word recognition, listeners who were previously exposed to specific stress errors were better at recognizing the target words than listeners who only heard the speaker's global foreign accent before. This adaptation to lexical stress errors confirms listeners' sensitivity to fine phonetic detail in spoken word recognition and shows that the reliance on phonetic detail does not depend on clear native speech but on experience with a specific speech style. Importantly, the adaptation to lexical stress errors occurs rapidly after only a short exposure and facilitates word recognition wherever stress errors occur.

Footnotes

[1] Now at: Carnegie Mellon University, Department of Psychology

[2] Also at: Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands

[3] Words from targets and competitors of the same-stress pairs could not be compared since within stress location (i.e., within pairs with correct stress and within pairs with incorrect stress) the words did not naturally fall into two comparison groups. For example, there is no inherent reason to group KAssa and AANdeel and compare it with CAssis and AANdrang rather than to group KAssa and AANdrang and compare it with CAssis and AANdeel. Between-pair comparisons (i.e., comparing words from the correctly versus incorrectly stressed words) could not be interpreted due to differences in segmental make-up (e.g., *KAssa - CAssis* vs. *paNEEL - paNIEK*).

[4] This was confirmed in separate analyses of listeners' target preference for the sets of correctly versus incorrectly stressed targets (data for the evidence group only). Mixed-effects models containing only an intercept term showed that only correctly stressed targets were recognized before segmental information could provide disambiguating information (correctly stressed targets: $b_{\text{intercept}}=0.644$, SE=0.21, $t=3.1$, $p_{(\text{MCMC})}<.005$; incorrectly stressed targets: $b_{\text{intercept}}=-0.053$, SE=0.19, $t=-0.28$, $p_{(\text{MCMC})}<.005$).

[5] Note that the effect of target preference for words with pitch-marked unstressed word-initial syllables may still not be expected to be as large as for targets with correct initial stress. Dutch listeners have been shown to recognize initially-stressed words better than non-initially stressed words (e.g., Reinisch et al., 2010; van Leyden & van Heuven, 1996). Nevertheless, if the pitch cue had been learned to mark unstressed syllables, some benefit could be expected because the acoustic cue would

not have matched the initially-stressed competitor as equally well as the non-initially stressed target.

Acknowledgments

Appendix A

Target-competitor pairs used for the eye-tracking task and their English translation.

Critical stress pairs (initially stressed word is given first)

deken/decor (blanket/scenery), satan/sate (devil/satay), salie/saluut (sage/salute), thermen/termiet (thermea/termite), atlas/atleet (atlas/athlete), natie/natuur (nation/nature), piste/pistool (piste/pistol), hersens/herstel (brain/recovery), koster/kostuum (sexton/suit), boete/boetiek (fine/boutique), kano/kaneel (canoe/cinnamon), dosis/dozijn (dose/dozen), motor/motief (moterbike/motive), mening/menu (opinion/menu), koepel/coupe (dome/coupe), ekster/eczeem (magpie/eczema), mormel/ mormoon (tyke/mormon), puber/publiek (adolescent/audience), kade/cadeau (quay/gift), pasta/pastoor (pasta/priest), krokus/kroket (crocus/croquette), kater/katoen (tomcat/cotton), coma/komijn (coma/cummin), laken/lakei (sheet/lacky), gordel/gordijn (belt/curtain), forum/forel (forum/trout), balie/ballet (counter/ballet), gieter/gitaar (watering can/guitar), robot/robijn (robot/ruby), kokos/cocon (coconut fibre/coccon), paling/paleis (eel/palace), marmer/marmot (marble/marmot)

Same stress pairs (both initial stress)

aandeel/aandrang (share/urgency), afschuw/afspraak (disgust/appointment), buffel/buffer (buffalo/buffer), ingang/ingreep (entrance/surgerey), invloed/inval (influence/invasion), inzicht/inzet (insight/effort), kassa/cassis (cash register/cassis), kennel/kennis (kennel/acquaintance), meter/metro (metre/subway), middag/middel (afternoon/waist), moeder/moedig (mother/brave), onrecht/onrust (injustice/commotion), opstand/optie (uprising/option), poedel/poeder (poodle/powder), uitzicht/uitspraak (view/pronounciation), voedsel/voetstuk (food/pedestal)

Same stress pairs (both second-syllable stress)

tabak/tablet (tobacco/tablet), paneel/paniek (panel/panic), verschil/verslag (difference/report), insekt/instinct (insect/instinct), matras/matroos (mattress/sailor), limiet/limoen (limit/lime), verdriet/verdrag (grief/convention), piloot/pilaar (pilot/pillar), gebruik/gebak (habit/pastry), geloof/geluk (faith/luck), applaus/april (applause/April), barbier/barbaar (barber/barbarian),

begin/begrip (beginning/understanding), bedrag/bedrijf (amount/company), modaal/modern (modal/modern), gebouw/gebrek (building/lack)

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419-439.

Archibald, J. (1993). *Language Learnability and L2 Phonology: The Acquisition of Metrical Parameters*. Dordrecht, The Netherlands: Kluwer Academic Publishers (pp. 110-149).

Baayen, H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effect modeling with crossed random effects for subjects and items. *Journal of Memory and Language,59*, 390-412.

Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457-474.

Bates, D. M., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using S4 classes* (version 0.999375-27) [software application]. Retrieved from http://www.r-project.org (date last viewed 06/06/2011)

Boersma, P. & Weenink, D. (2009). PRAAT: doing phonetics by computer (version 5.1) [software application] retrieved from http://www.praat.org (date last viewed 06/06/2011)

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

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

Braun, B., Dainora, A. & Ernestus, M. (2011). An unfamiliar intonation contour slows down online speech comprehension. *Language and Cognitive Processes, 26*, 350-375.

Braun, B., Lemhöfer, K., Mani, N. 2011. Perceiving unstressed vowels in foreign-accented English. *Journal of the Acoustical Society of America, 129,* 376-387.

Clarke, C.M., Garrett, M.F. 2004. Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America, 116,* 3647-3658.

Cooper, R. M. (1974). 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. *Cognitive Psychology, 6*, 84-107.

Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVI International Congress of Phonetic Sciences* (pp. 1913-1916). Dudweiler: Pirrot.

Cutler, A. & Koster, M. (2000). Stress and lexical activation in Dutch. in: B. Yuan, T. Huang, & X. Tang (Eds.). *Proceedings of the Sixth International Conference on Spoken Language Processing* (pp. 593-596). Beijing, China: China Military Friendship Publish.

Dahan, D., & Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory & Language*, 57, 483-501.

Eisner, F., Weber, A. & Melinger, A. (2010). Generalization of learning in pre-lexical adjustments to word-final devoicing (A). *Journal of the Acoustical Society of America, 128,* 2323.

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America, 97,* 1893-1904.

Fónagy, I. (1958). A hangsúlyról [On stress]. in P. Fábián (Ed.). *Nyelvtudományi Értekezések (Vol. 18).* Budapest, Hungary: Akadémiai Kiadó. (pp. 1-76).

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10-11--10-112). New York: Wiley.

Honbolygó, F., Csépe, V., & Ragó, A. (2004). Suprasegmental speech cues are automatically processed by the human brain: a mismatch negativity study. *Neuroscience Letters, 363*, 84-88.

Huettig, F., & McQueen, J.M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57,* 460–482.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59,* 434-446.

Jilka, M. (2000). Testing the contribution of prosody to the perception of foreign accent. in A. James, & J. Leather (Eds.). *Proceedings of New Sounds 2000*, University of Amsterdam. 199-207.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*, 141-178.

Mády, K., Bombien, L., & Reichel, U. D. (2008). Is Hungarian losing the vowel quantity distinction? In: R. Sock, S. Fuchs, Y. Laprie (Eds.), *Proceedings of the 8th International Seminar on Speech Production, Strasbourg* (pp. 445-448). Strasbourg, France: INRIA.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccadic overhead. *Perception & Psychophysics, 53*, 372-380.

Mitterer, H., Chen, Y., & Zhou, X. (2011). Phonological abstraction in processing lexical-tone variation: evidence from a learning paradigm. *Cognitive Science, 35*, 184-197.

Nooteboom, S. G. (1972). *Production and perception of vowel duration, a study of durational properties of vowels in Dutch.* Doctoral Dissertation, Utrecht University (pp. 1-165).

Norris, D., McQueen, J. M. & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238.

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics, 38*, 422-430.

Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye-movements immediately. *The Quarterly Journal of Experimental Psychology, 63,* 772-783.

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 978-996.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*, 51-89.

Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1108-1117.

Samuel, A. G. & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics, 71*, 1207-1218.

Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics, 68*, 1-16.

Sidaras, S.K., Alexander, J.E.D., Nygaard, L. C. 2009. Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America, 125*, 3306-3316.

Siptár, P. & Törkenczy, M. (2000). *The Phonology of Hungarian*. Oxford, UK: Oxford University Press. (pp. 21-26).

Swan, M. & Smith, B. (Eds.) (2001, 2nd edition) *Learner English: A Teacher's Guide to Interference and Other Problems*. Cambridge, UK: Cambridge University Press (pp. 1-362).

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology, 58A*, 251-273.

van Heuven, V. J. (1985). Perception of stress pattern and word recognition: Recognition of Dutch words with incorrect stress pattern (A). *Journal of the Acoustical Society of America, 78*, S21. Full text retrieved from: https://openaccess.leidenuniv.nl/retrieve/3572/167_101.pdf (date last viewed 06/06/2011).

Van Heuven, V. J., & Hagman, P. (1988). Lexical statistics and spoken word recognition in Dutch. In P. Coopmans & A. Hulk (Eds.), *Linguistics in the Netherlands 1988* (pp. 59-68). Dordrecht: Fortis.

van Leyden, K., & van Heuven, V. J. (1996). Lexical stress and spoken word recognition: Dutch vs. English. In M. Den Dikken & C. Cremers (Eds.), *Linguistics in the Netherlands 1996* (pp. 159-170). Amsterdam: John Benjamins.

White, L., Mády, K. 2008. The long and the short and the final: Phonological vowel length and prosodic timing in Hungarian. *Proceedings of Speech Prosody 2008*, Campinas, Brazil, 363-366.

Table 1: Mean duration (ms), mean pitch (Hz), and mean RMS-amplitude (Pascal) 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 (ms)	84	122	130	125
pitch (Hz)	204	170	200	202
RMS (Pascal)	0.090	0.063	0.075	0.081

Table 2: Mean duration (ms), mean pitch (Hz), and mean RMS-amplitude (Pascal) of the first vowels of targets and competitors of the critical stress pairs, and the comparison of their distributions.

	mean for correct stress	mean for incorrect stress	$t(30)$	p
duration (ms)	143	96	4.61	< .001
pitch (Hz)	176	184	-4.5	< .001
RMS (Pascal)	0.067	0.075	-2.02	.052

Table 3: Separate analyses of critical stress pairs for the two listener groups with an intercept term indicating target preference and the fixed factor *stress* (correct/incorrect).

	evidence group				no-evidence group			
	b	SE	*t*	*p*(MCMC)	*b*	SE	*t*	*p*(MCMC)
intercept	0.295	0.15	2.03	< .05	0.151	0.13	1.17	= .28
stress	0.696	0.25	2.74	< .01	0.049	0.26	0.19	= .85

Table 4. Model comparisons for analysis fixations related to the stress-contrast pairs during T1 testing experiment *half* as additional fixed factor. Degrees of freedom and log likelihood are given for each model. The degrees of freedom for the model comparison equal the difference in degrees of freedom between the models tested against each other (here always 1). χ^2 and p-values are given for the model comparisons.

model	degrees of freedom of the model	log likelihood of the model	Chi square for comparison	*p*-value of comparison
full model	11	-5045.6	--	--
minus interaction stress*half*group	10	-5045.7	0.12	=.73
minus interaction stress*half	9	-5045.7	0.18	=.67
minus interaction group*half	8	-5045.7	0.005	=.95
minus main effect of half	7	-5046.1	0.76	=.38

Table 5. Results for same stress pairs.

	initial stress pairs							
	T1				T2			
	b	SE	t	$p_{(mcmc)}$	b	SE	t	$p_{(mcmc)}$
intercept	-0.096	0.13	-0.73	.50	4.326	0.23	18.48	<.001
group	0.079	0.26	0.30	.77	0.376	0.30	1.26	.19
	non-initial stress pairs							
	T1				T2			
	b	SE	t	$p_{(mcmc)}$	b	SE	t	$p_{(mcmc)}$
intercept	0.172	0.13	1.35	.23	4.220	0.22	18.93	<.001
group	0.092	0.25	0.37	.71	0.897	0.29	3.06	<.005

Figure headings

Figure 1. Fixation proportions over time to target and competitor from acoustic target onset for the critical stress pairs. Solid lines represent fixations of 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.

Figure 2. Fixation proportions over time to target (black lines), competitor (dark grey lines), and averaged distractors (light grey lines) from acoustic target onset for the same stress pairs. Panel A shows the pairs in which targets were correctly stressed on the word-initial syllable and also the competitor matched the word-initial stress pattern. Panel B shows the pairs in which the targets were incorrectly stressed and also the competitor mismatched the pronunciation in stress. Solid lines represent the evidence group, dashed lines the no-evidence group. Vertical lines indicate boundaries for the time windows of analyses.