Greater sensitivity to prosodic goodness in non-native than in native listeners (L)

Anne Cutler[a)]
Max Planck Institute for Psycholinguistics, Nijmegen 6500 AH, The Netherlands and MARCS Auditory Laboratories, University of Western Sydney, NSW 1797, Australia

(Received 23 December 2008; revised 16 March 2009; accepted 16 March 2009)

English listeners largely disregard suprasegmental cues to stress in recognizing words. Evidence for this includes the demonstration of Fear *et al.* [J. Acoust. Soc. Am. **97**, 1893–1904 (1995)] that cross-splicings are tolerated between stressed and unstressed full vowels (e.g., *au-* of *autumn, automata*). Dutch listeners, however, do exploit suprasegmental stress cues in recognizing native-language words. In this study, Dutch listeners were presented with English materials from the study of Fear *et al.* Acceptability ratings by these listeners revealed sensitivity to suprasegmental mismatch, in particular, in replacements of unstressed full vowels by higher-stressed vowels, thus evincing greater sensitivity to prosodic goodness than had been shown by the original native listener group. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3117434]

PACS number(s): 43.71.Hw, 43.71.Es [RSN] Pages: 3522–3525

I. INTRODUCTION

Stressed and unstressed syllables in English differ both suprasegmentally and segmentally. Word recognition experiments in English have shown, however, that listeners attend primarily to the segmental structure, and to a great extent disregard suprasegmental cues to stress. Mis-stressing disrupts word identification only or principally when vowel quality is altered (Bond and Small, 1983; Cutler and Clifton, 1984; Slowiaczek, 1990), and the processing of minimal stress pairs where vowel quality does not differ, e.g., *insight/incite*, has in many differing experiments been shown to depend on contextual criteria rather than on assigned stress (Cutler, 1986; Slowiaczek, 1991; Small *et al.*, 1988; see Cutler, 2005, for a full review of the word recognition evidence).

A most telling finding was provided by Fear *et al.* (1995), who studied the production and perception of syllables varying in vowel quality (full, reduced) and stress. Vowel quality is the segmental reflection of English stress variation. Vowels bearing (primary or secondary) stress are always full, and reduced vowels are always unstressed. The reverse implications, though, do not hold; some full vowels do not bear stress, or conversely: some unstressed vowels are not reduced. English thus effectively has three syllable types: stressed with a full vowel (common), unstressed with a reduced vowel (common), and unstressed but with a full vowel (rare). Despite their rarity, the latter cases are crucial for understanding use of stress variation in speech perception.

The study of Fear *et al.* (1995) centered on sets of four vowel-initial words; each set comprised one word in which the initial vowel bore primary stress, one with an initial vowel bearing secondary stress (and primary stress on the third or fourth syllable), one with an unstressed initial vowel (and primary stress on the second syllable), and one with a reduced initial vowel. The four stimulus types can be referred to as P, S, U, and R, respectively, and an example set is *autumn, automation, automata, atomic*. Thus the initial vowel of *automata*, a U case, i.e., unstressed but with full vowel quality, is crucial; the central question of Fear *et al.* (1995) amounted to whether listeners treated U as more like R (because it was unstressed), as more like P or S (because it had a full vowel), or as a true separate case. Table I gives the sentence contexts for this set; it can be seen that phonetic and prosodic context was kept as similar as possible. The full set of materials is listed in Fear *et al.*, 1995.

In their study, 12 speakers produced all word sets in sentence context at two speech rates. Acoustic analyses showed that the four vowel types differed, at both rates. P and S differed in duration (P was longer) and these two stressed vowels as a class were distinct from U, and U from R, on duration, F0, intensity, and spectral quality. The acoustic foundation was thus in place for U vowels to function as cases different from either stressed (P/S) or R vowels.

To examine listeners' perception of (in particular) U vowels, the vowels were exhaustively cross-spliced within each word set, giving 16 stimuli per set. In shorthand terms in which each stimulus is referred to as a vowel (P, S, U, or R) followed by a body (from a word normally having P, S, U, or R vowel), the set for *autumn* would be PP, SP, UP, and RP, i.e., the P vowel (of *autumn*) or the S vowel (of *automation*), or the U vowel (of *automata*), or the R vowel (of *atomic*), each followed by the body of *autumn* (normally having an initial P vowel). The acceptability of these 16 item types was rated by listeners, who heard them either in a neutral environment, offering no contextual support regarding the identity of the word, or in a meaningful context, in which it was clear what each word should be (see Table I); the meaningful sentences were spoken at two speech rates: normal or fast.

Fear *et al.* (1995) distinguished four possible hypothesised outcomes of this perception test. The outcomes were principally distinguished by different patterns of statistical associations for the six stimulus types involving U vowels cross-spliced with another vowel. Two hypotheses assumed that U vowels would be treated as unlike either stressed or

[a)]Electronic mail: anne.cutler@mpi.nl

 0001-4966/2009/125(6)/3522/4/$25.00

TABLE I. Example stimulus set.

Initial syllable stress	Example word	Sentence context
P (primary)	*Autumn*	Summer is the time for berries, but autumn is the time for apples.
S (secondary)	*Automation*	The factory once employed 80, but automation reduced this by half.
U (unstressed)	*Automata*	The workers were treated as if they weren't humans, but automata to be programmed.
R (reduced)	*Atomic*	Armies used to be a country's main defense, but atomic weapons changed all that.

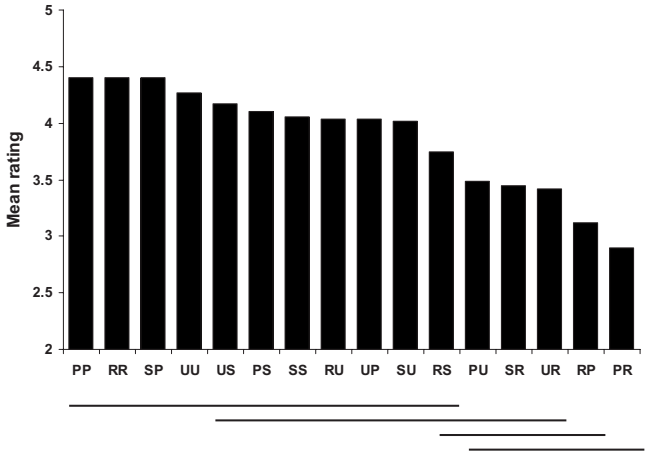

FIG. 1. Mean acceptability ratings (across participants and word sets) given by Dutch listeners to the 16 stimulus types, coded in vowel-body order (so, SP is a secondary-stressed vowel on the word body of the primary-stressed set member). Ratings were on a scale of 1–5, with maximum acceptability at 5. Ratings of word types linked by underlining do not differ statistically.

reduced vowels, and postulated a grouping linking US, SU, UP, PU, RU, and UR because they would all be perceived as mismatching. The other two hypotheses postulated a single category boundary, either based on vowel quality, with RU and UR items (different vowel quality expected versus realized) grouping distinctly from SU, US, PU, and UP (same quality), or on difference magnitude, with PU and UP (two steps apart) grouping distinctly from US, SU, UR, and RU (all one step apart). The results of Fear *et al.* (1995) most strongly supported a vowel-based categorical distinction. The acceptability ratings fell into distinct sets, linked by no statistical association, and overall SU, US, PU, and UP items were in one set, and RU and UR in the other. Further, correlation analyses of the listeners' ratings with acoustic properties of the stimuli showed the ratings to be more strongly related to measures of vowel quality than to any suprasegmental dimension. Fear *et al.* (1995) concluded that although the production data did not support a categorical distinction, listeners acted as if there were one anyway.

English stress is very similar to that in other Germanic languages such as Dutch and German (van der Hulst, 1999), but the vowel-based categoricality apparent in English listeners' responses is not seen in other Germanic languages for which listening evidence is plentiful. Mis-stressing in Dutch harms word identification even when vowel quality is unchanged (van Heuven, 1985), and suprasegmental information is used in the processing of Dutch minimal stress pairs (Cutler and van Donselaar, 2001). In both Dutch and German, word fragments cause inhibition of stress-mismatching words (van Donselaar *et al.*, 2005; Friedrich, 2002), whereas this inhibition is absent in English (Cooper *et al.*, 2002); thus *admi*-from *admiration* does not effectively inhibit *admiral*, whereas in Dutch, *domi*- of final-stress *dominant* ("dominant") does inhibit initially-stressed *dominee* ("pastor"). Sensitivity to suprasegmental cues to stress level is greater in German- or Dutch-speakers than in English-speakers.

When such speakers acquire English as a second language, the cues to stress level that they encounter and use in their native language will—as acoustic analyses of Fear *et al.* (1995) showed—also be available in the English they hear. Indeed, Cooper *et al.* (2002) found that Dutch listeners outperformed native English listeners in use of suprasegmental information. In the present study, Dutch listeners were exposed to the cross-spliced materials created by Fear *et al.* (1995), and their responses were compared with those of the native listeners in the original study.

II. METHOD

A. Materials

The stimuli of Fear *et al.* (1995) in the meaningful context, in which it was clear what each word should be, were selected for the present study. (Preliminary tests suggested that both the fast-speech condition and the neutral environment, offering no contextual cues to word identity, were difficult for non-native listeners.) There were 80 items: five word sets of four words each, each word occurring in four vowel versions. All versions contained the same word body in its sentence context, but differed in whether the initial vowel was original or was spliced in from another word of the same set. Thus in the example sentence "…but autumn is the time for apples," the word *autumn* could be, as described, a PP, SP, UP, or RP token. The test items were preceded, as in the original study, by a practice set of cross-spliced versions of *upper*, *upset*, *appeal* in sentence context.

B. Participants and procedure

Twenty-four Nijmegen University undergraduates, all native Dutch speakers with high proficiency in English, heard the sentences over Sennheiser headphones from disk, and followed written instructions (in Dutch) to rate the naturalness of the critical word specified for each sentence on a scale from 1 to 5, with 1 signifying that the word's prosodic form was completely wrong, 5 that it was completely right.

III. RESULTS

A. Acceptability ratings

Mean acceptability ratings were computed for each version of each word. The mean ratings for the 16 stimulus types (four words with four vowels each), averaged across the five word sets, are shown in Fig. 1 in order of rated

acceptability. As Figs. 5, 6, and 7 in Fear *et al.*, 1995 show, the original acceptability ratings plotted in this way showed a clear discontinuity, both in the meaningful context used here, as in the neutral environment, and overall. The present results are not discontinuous in this way, already suggesting that for the Dutch listeners no single over-riding cue is ascribed greatest weight in judging acceptability.

Statistical difference between these means was tested, as in the original study, by multiple comparisons; the association lines under the stimulus types in the figure link sets which are not statistically different from one another. In the original study, again, the multiple comparison results produced distinct sets; in both Fig. 7 of Fear *et al.*, 1995 for the meaningful context and Fig. 6 for the neutral context, despite differences in the precise ordering, the full set of 16 stimulus types always fell into two sets linked by no association lines. Here, this is not so; the four highest-rated stimuli are linked by no association lines to the five lowest-rated, but there is overlap across the seven stimulus types in between.

Of the four possible hypotheses set out by Fear *et al.* (1995), the two postulating a single category boundary are thus not supported by the present data. Instead, Fig. 1 shows a single association running from US to UR. Fear *et al.* (1995) predicted this if all cross-splicings involving U vowels were treated as mismatches, i.e., U vowels were treated as differing from both stressed and from reduced vowels.

B. Cross-experiment comparisons

Fear *et al.* (1995) conducted no analyses across participants, but in order to compare the response patterns of the present participant group with those of the original listeners, the raw data of Fear *et al.* (1995) was re-coded and mean ratings for each stimulus type were computed for each listener. The same values were computed for the present participants. Because the Dutch listeners' ratings were distributed over a different range than those of the native speakers, each group's ratings were converted to *z*-scores for a comparison across subjects via *t*-tests (uncorrelated means).

No effects of listener group were observed for most stimulus types. Just as the identity conditions were rated most highly by both listener groups, so the vowel quality differences were rated least highly by both. Significant differences emerged in exactly two conditions: U words with vowel replaced by a more highly stressed vowel (i.e., PU, SU). In each case ratings were significantly lower from the Dutch than from the native listeners ($t[46]=2.285$, $p=0.027$ for PU, $t[46]=2.292$, $p=0.027$ for SU).

C. Correlations of acceptability with acoustic factors

The acoustic properties of each stimulus token had been recorded by Fear *et al.* (1995), who conducted correlation analyses between their listeners' ratings for each token and measures of the difference between that token's original and substituted vowel in duration, intensity, F0, and spectral quality (expressed as the difference between F1 and F2 on a log scale). They found, as described, that the ratings were more strongly related to the vowel measure than to any of the suprasegmental measures (of which one, F0, was not related to the ratings at all, largely because the values for the speaker's tokens did not differ widely on this measure).

The same analyses were conducted for the present data. Again, the F0 measures did not predict listeners' ratings. There were significant correlations of rated acceptability with differences in duration ($r[59]=-0.483$, $p<0.001$) and intensity ($r[59]=-0.526$, $p<0.001$). In contrast to the native listener group, the present listeners did not show as systematic an effect of vowel quality difference; the correlation which had been the strongest for the native group did not reach significance here ($p=0.088$).

IV. DISCUSSION

The pattern of results found by Fear *et al.* (1995) with English listeners was not replicated here. The listener ratings of cross-spliced stimuli in Fear *et al.* (1995) were best described by a statistical model involving a vowel-based category distinction: stimuli in which vowel quality was preserved were statistically indistinguishable, and were rated significantly better than stimuli in which vowel quality was altered. Such a categorical division was not observed here; instead, there was a more gradient distribution across the ratings received by the stimulus types, with overlapping statistically motivated groupings, including one spanning the six cross-splicings involving U vowels. The responses of the original listeners correlated most strongly with a vowel quality measure, while the present responses correlated more strongly with acoustic measures of intensity and duration.

Direct comparisons between the two response sets showed that the differences principally occurred with words containing U vowels. Substitutions of a more stressed vowel (the PU and SU cases) were significantly less acceptable to the present Dutch listeners than to the original group.

Note that the converse of this finding is that in general, the ratings given by these non-native listeners were not significantly different from those of native listeners. Like listeners of Fear *et al.* (1995), the present group rated the items with an original vowel most highly, and the cross-spliced items where both segments and suprasegmentals mismatched least highly. The difference emerged precisely with the crucial case in which the mismatch was exclusively suprasegmental. By focusing on this case, using the paradigm devised by Fear *et al.*, (1995) it has been possible to discern a difference between these listener groups who are otherwise performing at an equivalently high level.

The findings suggest that Dutch listeners have a rather more refined appreciation of prosodic goodness than do English listeners. Specifically, they have a notion of what words with U (unstressed but full) vowels should sound like. There is good reason why they should have such a concept; U vowels are far more common in Dutch than they are in English. Cognate pairs abound to illustrate this: the first syllables of *cigar, parade, banana*, and the second syllables of *panda, cobra*, and *octopus* are all reduced in English, while the equivalent syllables in Dutch *sigaar, parade, banaan, panda, cobra*, and *octopus* all have full U vowels. Substitution of P or S vowels in place of the U vowel in such words would violate listeners' preconceptions. Cooper *et al.* (2002) found

that English syllables such as *mus-* from *music* versus *museum* could be more accurately attributed by Dutch than by English listeners, and the difference was greatest with the unstressed case (*mus-* from *museum*). English-speaking participants in that study actually performed significantly below chance with these cases. Dutch-speaking participants performed significantly above chance with the same items.

Another feature of the present results is also indicative of sensitivity to prosodic goodness. The present Dutch listener group produced a completely consistent response pattern with respect to direction of a mismatching cross-splice; mismatches in which the inserted vowel was less stressed than the original vowel were always rated more highly than mismatches in which a more highly stressed vowel replaced a less stressed (thus, as can be seen from Fig. 1, the rating for SP is above that for PS, UP above PU, US above SU, RP above PR, RS above SR, and RU above UR). This is in accord with what happens in natural speech—casual-speech processes lead to both suprasegmental and segmental reduction, so that citation forms are realised naturally in less stressed form. The reverse situation, i.e., higher stress than in the citation form, can also happen (in contrastive stress on morphemes, e.g., *He said IGnited not United*), but is far less likely.

This internal consistency, too, was absent from the responses found by Fear *et al.* (1995). For their English listeners, substitution of a more highly stressed vowel for a U vowel actually led to a higher rating than the reverse operation (with the same items as used here, SU was rated higher than US, and PU higher than UP). If, as Fear *et al.* (1995) proposed, the English listeners have internalized a vowel-based distinction (full versus reduced) only, then it may be that a P or S vowel better matches to their notion of what a full vowel should sound like than a U vowel does.

There are good reasons for English listeners to attend less to suprasegmental information in speech than listeners of other languages do. The payoff, in terms of reduction in the competitor population in spoken-word recognition, is significantly less in English than it is in Dutch and other languages (Cutler *et al.*, 2004; Cutler and Pasveer, 2006). English listeners can easily distinguish *parade* from *paradise* during the first syllable, because the vowels in the two words are different; Dutch listeners can distinguish the cognate words *parade* and *paradijs* as quickly as this only if they can use suprasegmental cues in the initial syllables, because even though the stress is different, the vowel quality is the same.

In comparison to native listeners, non-native listeners are in general disadvantaged. However, the present findings join those of Cooper *et al.* (2002) as evidence that listening skills encouraged by the native language might sometimes, when deployed in non-native listening, help to compensate for the disadvantages. The processing of stress is not the only case in which Dutch listeners to English appear to outdo native listeners; Broersma (2005, 2008) also observed that Dutch listeners can display more sensitivity than English listeners in distinguishing voicing contrasts at the end of English syllables. In that case, the native language, where voicing can contrast only syllable-initially, encouraged attention to cues within the consonant to distinguish the voicing feature. Dutch listeners could then use such cues also for English syllable-final contrasts, whereas for English listeners, vowel duration was an over-riding cue. In the stress case, the greater frequency of U vowels in Dutch has caused listeners to store knowledge of the acoustic differences between such vowels and vowels bearing stress. This knowledge can be deployed in recognizing Dutch words (e.g., to rule out *parade* during the *pa-* of *paradijs*). Even though the payoff is undoubtedly less in English than in Dutch, the same acoustic differences do obtain and hence the native listening skills can be transferred. In English, then, a Dutch listener could in principle call on these skills to determine, fractionally earlier than a native English listener, that *automatic* is being uttered, and not *automata*.

ACKNOWLEDGMENTS

Financial support was provided by the NWO SPINOZA project "Native and Non-native Listening." Thanks to Abeer Alwan for suggesting that this experiment would be likely to work, and to Sally Butterfield, Karly van Gorp, and Michael Tyler for unmissable assistance in making it happen.

Bond, Z., and Small, L. H. (**1983**). "Voicing, vowel, and stress mispronunciations in continuous speech," Percept. Psychophys. **34**, 470–474.
Broersma, M. (**2005**). "Perception of familiar contrasts in unfamiliar positions," J. Acoust. Soc. Am. **117**, 3890–3901.
Broersma, M. (**2008**). "Flexible cue use in nonnative phonetic categorization," J. Acoust. Soc. Am. **124**, 712–715.
Cooper, N., Cutler, A., and Wales, R. (**2002**). "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," Lang Speech **45**, 207–228.
Cutler, A. (**1986**). "Forbear is a homophone: Lexical prosody does not constrain lexical access," Lang Speech **29**, 201–220.
Cutler, A. (**2005**). "Lexical stress," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Oxford), pp. 264–289.
Cutler, A., and Clifton, C. (**1984**). "The use of prosodic information in word recognition," in *Attention and Performance X: Control of Language Processes*, edited by H. Bouma and D. G. Bouwhuis (Erlbaum, Hillsdale, NJ), pp. 183–196.
Cutler, A., and van Donselaar, W. (**2001**). "Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch," Lang Speech **44**, 171–195.
Cutler, A., Norris, D., and Sebastián-Gallés, N. (**2004**). "Phonemic repertoire and similarity within the vocabulary," in Proceedings of Eighth International Conference on Spoken Language Processing, edited by S. H. Kim and D. H. Youn (Sunjin Printing Co., Seoul), Vol. **1**, pp. 65–68.
Cutler, A., and Pasveer, D. (**2006**). "Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition," in Proceedings of the Third International Conference on Speech Prosody, edited by R. Hoffman and H. Mixdorff (TUD, Dresden), pp. 250–254.
Fear, B. D., Cutler, A., and Butterfield, S. (**1995**). "The strong/weak syllable distinction in English," J. Acoust. Soc. Am. **97**, 1893–1904.
Friedrich, C. K. (**2002**). "Prosody and spoken word recognition—Behavioral and ERP correlates," Ph.D. thesis, University of Leipzig, Leipzig.
Slowiaczek, L. M. (**1990**). "Effects of lexical stress in auditory word recognition," Lang Speech **33**, 47–68.
Slowiaczek, L. M. (**1991**). "Stress and context in auditory word recognition," J. Psycholinguist. Res. **20**, 465–481.
Small, L. H., Simon, S. D., and Goldberg, J. S. (**1988**). "Lexical stress and lexical access: Homographs versus nonhomographs," Percept. Psychophys. **44**, 272–280.
van der Hulst, H. G. (**1999**). "Word accent," in *Word Prosodic Systems in the Languages of Europe*, H. G. van der Hulst (Mouton de Gruyter, Berlin), pp. 3–116.
van Donselaar, W., Koster, M., and Cutler, A. (**2005**). "Exploring the role of lexical stress in lexical recognition," Q. J. Exp. Psychol. A **58**, 251–273.
van Heuven, V. J. (**1985**). "Perception of stress pattern and word recognition: Recognition of Dutch words with incorrect stress position," J. Acoust. Soc. Am. **78**, S21.