SEGMENTAL AND SUPRASEGMENTAL CONTRIBUTIONS TO SPOKEN-WORD RECOGNITION IN DUTCH

Mariette Koster and Anne Cutler

Max Planck Institute for Psycholinguistics
Postbus 310, 6500 AH Nijmegen, The Netherlands
Tel +31 24 352 1911; Email anne.cutler@mpi.nl

ABSTRACT

Words can be distinguished by segmental differences or by suprasegmental differences or both. Studies from English suggest that suprasegmentals play little role in human spoken-word recognition; English stress, however, is nearly always unambiguously coded in segmental structure (vowel quality); this relationship is less close in Dutch. The present study directly compared the effects of segmental and suprasegmental mispronunciation on word recognition in Dutch. There was a strong effect of suprasegmental mispronunciation, suggesting that Dutch listeners do exploit suprasegmental information in word recognition. Previous findings indicating the effects of mis-stressing for Dutch differ with stress position were replicated only when segmental change was involved, suggesting that this is an effect of segmental rather than suprasegmental processing.

1. INTRODUCTION

Languages contain many thousands of words, but these words are constructed from a notably limited array of phonetic resources. Words can be distinguished by segmental differences: *bellow* vs. *mellow,* or *rusty* vs. *trusty;* but in many languages suprasegmental means - variations in pitch, amplitude and duration of syllables - are also used to distinguish one word from another. Thus in languages like English and Dutch, all polysyllabic words have the property of lexical stress: one syllable is marked for higher stress than the others). In English, *bellow* is stressed on the first syllable, below on the second; *trusty* on the first and *trustee* on the second. However, the words *bellow* and *below* also differ in the vowel sound in the first syllable: the unstressed first syllable of *below* contains the reduced vowel [e] - schwa, while the stressed first syllable of *bellow* contains the full vowel [e], as in *bed. Trusty* and *trustee* do not differ in vowel sounds; they are distinguished only by the stress difference. Dutch contains similar variety: *'voornaam* ("first name") and *voor'naam* ("respectable") differ only in stress, while *'regent* ("is raining") and *re'gent* ("regent") differ both in stress and in vowel quality: the second syllables show exactly the same opposition between schwa in the unstressed case and [e] in the stressed case as in the English example.

Given the large number of words in an adult listener's vocabulary, and the patently obvious speed and efficiency with which human listeners recognise spoken words, it might seem reasonable that listeners should make use of any and all information in the signal to help distinguish the actual words in the input from competing similar words in the vocabulary. Yet experimental evidence shows that English listeners make little use of suprasegmental information in spoken-word recognition: vowel quality differences are perceptually more important than stress differences [1], and minimal stress pairs with no segmental difference (such as *trusty/trustee)* are, in the earliest stages of word recognition, effectively homophonous [2], suggesting that the suprasegmental distinction plays no role at this stage.

In English, however, the correspondence between stress and vowel quality, such that stressed vowels are always full, while unstressed vowels are nearly always reduced, is pervasive. Thus almost every time two words are distinguished suprasegmentally they are also distinguished segmentally; pairs such as *trusty/trustee* are very rare indeed; and stressed versus unstressed syllables can nearly always be identified from vowel quality. The cost of processing suprasegmental information may therefore hot be warranted for English listeners by the small benefit it would offer. Dutch, although phonologically similar to English in many ways, nevertheless differs in the degree of correspondence of stress and vowel quality. In Dutch, many words contain unstressed syllables with full vowels; the unstressed syllables of *si'gaar* ("cigar") and *'cobra* ("cobra"), for instance, have full vowels, where the English counterparts have reduced vowels. Thus suprasegmentals may offer Dutch listeners sufficient information about word identity not available in segmental structure to justify whatever

cost suprasegmental processing might involve. Indeed, recent studies of the recognition of correctly-pronounced words suggest that supra-segmentals play a greater role for Dutch than for English [3].

A number of studies have examined the role of stress in word recognition by assessing the effect (if any) of misplacing word stress. In English, such studies have strengthened the conclusion that segmental structure is more important for word recognition than supra-segmental structure. Thus of all the ways one can slightly alter a word's pronunciation, alteration of vowels in stressed syllables most inhibits successful recognition [4]; mis-stressing of words has no adverse effect on word recognition in noise unless vowel quality is also changed [5]; but mis-stressing with vowel quality change renders word recognition, even without noise masking, very difficult [6] or indeed impossible [7]. Studies of mis-stressing in Dutch [8,9] have not explicitly compared the contributions of suprasegmental versus segmental structure to the effects of mis-stressing on word recognition. However these studies have revealed a pattern unknown in English: mis-stressing words which normally have final stress (e.g. saying *'sigaar)* is reported to be more harmful than mis-stressing words which normally have initial stress (e.g. saying *co'bra).*

The present investigation systematically investigated the effects on word recognition of (a) mis-stressing via only a suprasegmental change, (b) mispronunciation via only a vowel quality change, with no supra-segmental change, and (c) mis-stressing via both segmental and suprasegmental change; furthermore the effects were investigated in (d) both initially-stressed and finally-stressed words. Our experiment used a semantic judgement task which clearly required successful word recognition. Thus it provided us with a perspective on Dutch word recognition via which we could compare the contributions of segmental and suprasegmental information.

2. METHOD

2.1. Materials

84 monomorphemic bisyllabic Dutch nouns were selected, half with initial *(cobra, blunder)* and half with final stress *(fatsoen, begin);* within each set of 42, half of the words had a full vowel in the unstressed syllable *(cobra, fatsoen)* and half had a reduced vowel *(blunder, begin).* Three versions of

each word were constructed: (a) no change (i.e. correctly pronounced, (b) with one change (either in segmental or in segmental structure), and (c) with two changes, i.e. with both segmental and suprasegmental structure altered. In the no-change condition, all words were uttered in their normal form: e.g. *'cobra, fat'soen, 'blunder, be'gin.* In the two-change condition, the suprasegmental correlates of stress were assigned to the normally unstressed syllable, and the vowel in one syllable was changed: for words with vowel reduction *('blunder, be'gin),* that vowel became full *(blun'dier, 'beegin),* while for words with two full vowels *('cobra, fat'soen),* the normally stressed vowel was reduced *(co'bra, 'fatsen).* In the single-change condition, words with vowel reduction had the reduced vowel changed to full, but no suprasegmental change occurred *('blundier, bee'gin),* while words with two full vowels received suprasegmental stress on the normally unstressed syllable but no segmental change was applied *(co'bra, fatsoen).*

A further 84 filler words were chosen, similar in length, frequency and prosodic structure to the experimental words, and 16 practice items. Three versions of each of these were also constructed. All these items were recorded by a female native speaker of Dutch. All words were digitised and stored on disc. The experimental items were also measured. A prime word was chosen for each experimental, filler and practice word. For the experimental (and half the practice) target words, the primes were words related in meaning; thus for *blunder* the prime was FOUT ("mistake"), and for *begin* it was AANVANG ("start"). Relatedness between prime and target pairs was established via a pre-test. For the filler (and half the practice) words the prime was unrelated; thus the prime for the filler word *fauna* was SCHRIK ("fright").

2.2. Subjects and Procedure

Subjects were 48 native speakers of Dutch, members of the Nijmegen University community, who were paid for participating. They saw a prime word displayed on a computer screen, and then heard a spoken word; their task was to decide whether the two words were similar in meaning, and to signal their decision as rapidly as possible by pressing one of two response keys labelled YES and NO. The subjects were informed that some words would be mispronounced. The subjects were further asked to speak aloud, in its correct form, the word that they had just heard after each keypress response; when these spoken responses were not correct, the relevant

response was discarded from the data set. As all experimental words were preceded by a visual word related in meaning, the correct response for experimental words was YES, and subjects' reaction time (RT) to make this response was measured. The RTs, from a timing mark aligned with spoken-word onset to the subject's keypress, were collected by a computer running the experimental control program NESU. We also recorded the error rate (proportion of missed or erroneous responses).

3. RESULTS

Mean RTs (measured from spoken-word onset) and error rates were calculated, averaged across subjects and items for each condition; separate analyses of variance were conducted across subjects (F1) and across items (F2). We will report statistics only for effects which reached significance in both analyses. The analyses were conducted separately for the items with vowel reduction *(blunder, begin)* versus the items with full vowels *(cobra, fatsoen),* because the nature of the change in the single-change condition differed across these groups. Using the measured durations of the experimental items, a parallel analysis of RTs from word offset was also possible. One item to which hardly any subject responded correctly was discarded from all analyses *(kudde,* from the *blunder* set). The mean RTs (from word onset) and error rates for each condition are shown in Table 1.

Word Type:	No change	Single change	Two changes
'blunder	796 (5%)	853 (7%)	926 (9%)
be'gin	912 (9%)	949 (7%)	1019 (11%)
'cobra	830 (9%)	865 (10%)	1036 (24%)
fat'soen	909 (13%)	975 (10%)	999 (21%)

Table 1. Mean correct "YES" response times (in ms, measured from word onset), and mean percentage of missed or erroneous responses, for each of the four word types in the three pronunciation conditions.

The analysis of errors showed no effects of stress and no effects of either type of single change; two changes produced a significant increase in error rate for words with full vowels only. Further analyses concentrated on the differences in RT between the changed-pronunciation conditions and the baseline, no-change, condition. As the mean RTs clearly show, there were substantial mispronunciation effects: a segmental change alone, or a suprasegmental change alone, or both, all resulted in longer RTs than to the unchanged words. The main effect of the comparison between conditions was highly significant both for words with vowel reduction (Fl [2,44] = 30.84, p < .001; F2 [2,38] = 31.02, p < .001) and for words with full vowels ((Fl [2,44] = 50.11, p < .001; F2 [2,39] = 11.29,p<.001).

Table 1 also shows that RTs are in general longer to finally- than to initially-stressed words. The main effect of stress pattern was indeed significant for the words with vowel reduction (Fl [1,45] = 92.32, p < .001; F2 [1,39] = 7.92, p < .001) but for the words with full vowels it reached significance only across subjects. It is generally the case that bisyllables with initial stress have shorter duration than bisyllables with final stress; but this difference in RTs cannot be just an artefact of word duration because it also appears when the stress patterns are in fact reversed. Nevertheless we examined the effects in a further analysis of RTs from word offset. As expected, the main effect of stress pattern no longer reached significance (at all, for words with full vowels, and across items, for words with reduced vowels), suggesting that it was indeed due to differences in word duration rather than stress per se. However, this analysis also revealed that many responses in the no-change condition had been made prior to word offset, and the exclusion of these responses reduced the data set for the no-change condition to an undesirable extent. Accordingly we decided to confine further analyses to the RTs from word onset.

A segmental plus a suprasegmental mispronunciation significantly slowed RT - the difference between the no-change and the two-change condition (on average, 135 ms) was statistically significant across subjects and across items for each of the four word types separately. Collapsed across initially- versus finally-stressed words, the effects of a single change in suprasegmental (50 ms slower than correct pronunciation) or in segmental structure (47 ms slower) were virtually identical in size, and an analysis of variance comparing just the no-change and the single-change condition showed that the effect of a change did not interact with the type of change. However while a single segmental change was significant (across both subjects and items) for words of the *blunder* type, it was not significant (across

either) for words of the *begin* type; and while a single suprasegmental change was significant (across both subjects and items) for words of the *fatsoen* type, it was not significant (across either) for words of the *cobra* type.

Note that this last result for the single suprasegmental change is in contradiction to that of Van Heuven and colleagues, who reported consistently greater effects for words with initial stress (i.e. greater effects for words of the *cobra* type than for words of the *fatsoen* type).

4. DISCUSSION

These results suggest mat the effects of stress on word recognition in Dutch differ from the effects found in English; in particular, a suprasegmental change alone has as at least as strong an effect in Dutch as a segmental change. Suprasegmental correlates of stress carry more informative load in Dutch than they do in English, and in consequence they play a greater role in listeners' word recognition processes. The present results are thus consistent with the conclusions reached by Cutler, Dahan and Van Donselaar [3] in a review of evidence on the role of prosody in spoken-language comprehension: listeners will use whatever information is available and efficient. Suprasegmental information plays little role in spoken-word recognition in English, but this is because its use is not efficient in that language, due to the redundant coding of stress via segmental and suprasegmental information simultaneously. In Dutch, there is less such redundancy, and despite the considerable prosodic similarity between the two languages Dutch thus differs from English in that it warrants greater use of suprasegmental information in word recognition.

As we described in the introduction, previous studies of the effects of mis-stressing in Dutch by Van Heuven and colleagues [7,8] have suggested that initially-stressed words may show greater adverse effects of mis-stressing than finally-stressed words. We indeed replicated this asymmetry in the two-change condition and in the single segmental change condition. In the single suprasegmental change condition in the present study, however, effects were not very different for finally- versus initially-stressed words, and reached significance only for the former. This suggests that the asymmetry that Van Heuven found may reflect segmental processing rather than suprasegmental. This is exactly as would be expected if a different vowel in the first syllable of

a word is likely to activate a greater number of competing word candidates than a different stress pattern.

Finally, implications may be drawn from the present study for automatic recognition of speech: recognisers may have little to gain by attempting to exploit suprasegmentals in English, but could well benefit from using this information in the recognition of Dutch, and this once again underlines the need to tailor design of such systems to take into account language-specific phonological structure.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1]   B.D. Fear, A. Cutler and S. Butterfield, "The strong/weak syllable distinction in English", *Journal of the Acoustical Society of America,* Vol. 97, pp. 1893-1904, 1995.

[2]   A. Cutler, *"Forbear* is a homophone: lexical prosody does not constrain lexical access", *Language and Speech,* Vol. 29, pp. 201-220, 1986.

[3]   A. Cutler, D. Dahan and W. van Donselaar, "Prosody in the comprehension of spoken language: A literature review", *Language and Speech,* Vol. 40, 1997.

[4]   Z.S. Bond, "Listening to elliptic speech: pay attention to stressed vowels", *Journal of Phonetics,* Vol. 9, pp. 89-96,1981.

[5]   L.M. Slowiacek, "Effects of lexical stress in auditory word recognition", *Language and Speech,* Vol. 33, pp. 47-68, 1990.

[6]   A. Cutler and C.E. Clifton, "The use of prosodic information in word recognition", In H. Bouma & D.G. Bouwhuis (Eds.). *Attention and Performance X.* Hillsdale, N.J.: Erlbaum, 1984.

[7]   Z.S. Bond and L.H. Small, "Voicing, vowel and stress mispronunciations in continuous speech", *Perception & Psychophysics,* Vol. 34, pp. 470-474, 1983.

[8]   V.J. van Heuven, "Perception of stress pattern and word recognition: Recognition of Dutch words with incorrect stress position", *Journal of the Acoustical Society of America,* Vol.78, s21,1985.

[9] K. van Leyden and V J. van Heuven, "Lexical stress and spoken word recognition: Dutch vs. English", in C. Cremers and M. den Dikken (Eds.), *Linguistics in the Netherlands 1996.* Amsterdam: John Benjamins, 19%.