Dutch listeners' use of suprasegmental cues to English stress

Anne Cutler [1,2], Roger Wales [3], Nicole Cooper [1] and Joris Janssen [1]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; [2] MARCS Auditory Laboratories, University of Western Sydney, Australia; [3] LaTrobe University, Bundoora, Australia
anne.cutler@mpi.nl; r.wales@latrobe.edu.au; 1okokoro@cox.net; jorisjanssen@student.ru.nl

ABSTRACT

Dutch listeners outperform native listeners in identifying syllable stress in English. This is because lexical stress is more useful in recognition of spoken words of Dutch than of English, so that Dutch listeners pay greater attention to stress in general. We examined Dutch listeners' use of the acoustic correlates of English stress. Primary- and secondary-stressed syllables differ significantly on acoustic measures, and some differences, in F0 especially, correlate with data of earlier listening experiments. The correlations found in the Dutch responses were not paralleled in data from native listeners. Thus the acoustic cues which distinguish English primary versus secondary stress are better exploited by Dutch than by native listeners.

Keywords: stress (lexical), Dutch, English.

1. INTRODUCTION

English and Dutch stress patterns are very similar [1]; however, some subtle differences have major consequences for the recognition of words. Lexical statistics show that considering lexical stress in word identification removes more competitors in Dutch than in English [2]. Listeners always attend to vowel quality to recognize words (e.g., to tell *cub* from *cab*); in English, the same vowel quality processing usually delivers the distinction between syllables which differ in stress. Thus in *SUBject* vs. *subJECT*, for instance, the vowels of the first syllables differ (the latter *sub*- has a reduced vowel). In English, listeners have nothing further to gain by also computing suprasegmental features. By extension, then, they are not explicitly considering stress. Dutch, in contrast, has less vowel reduction. The initial syllables of *SUBject* 'subject' and *subSIdie* 'subsidy' are segmentally the same, despite the stress difference. Dutch has many syllables with full vowels but without primary stress; thus paying attention explicitly to stress, via suprasegmental information, is of use in Dutch [2]; only the suprasegmentals allow Dutch listeners to tell *sub*- from *subject* versus *subsidie*.

In English, misstressing has little effect on word identification as long as vowel quality is not changed [3,4,5], but in Dutch there are large adverse effects of misstressing [6]. On-line experiments show that Dutch listeners also use suprasegmental information efficiently to resolve lexical competition between alternative words [7,8], while English-speakers make less use of stress for this purpose [9]. English listeners show greater bias effects from the predominance in the lexicon of initial stress than Dutch listeners [6,9].

English has some stress contrast in segmentally identical syllables – e.g., in the initial syllables of *MUsic* and *muSEum*. English listeners perform poorly at identifying such syllables in isolation [9]; Dutch listeners, who score very highly on the same task in their own language [7], actually outperform native listeners in this task in English [9]. In the present study we examined how Dutch listeners use suprasegmental cues to outperform native listeners. We report follow-up analyses and further investigations with the same materials used in two experiments in [9]. First, we report acoustic measures of the initial syllables of the pairs such as *mus(ic)/mus(eum)*. We will refer to the stress levels of the initial syllables of these two words as primary versus secondary stress (ignoring possible intermediate levels, e.g., as in *musiCOLogy*). Next, we analyse item mean RTs from two experiments in [9] as a function of the acoustic measures. One experiment used an on-line task (cross-modal priming), revealing listeners' use of stress cues in resolving lexical competition; the other used an off-line task (2AFC: two-alternative forced-choice identification), revealing listeners' ability to exploit available information with no time pressure. Finally, we report a new gating study in which Dutch listeners attempted to identify the words given fragments of the initial syllables. The gating task falls between the other two tasks; like 2AFC, it is offline and allows the listener full opportunity to exploit all available information, but with its incremental presentation, it offers, like priming, a view of gradually unfolding word processing.

2. ACOUSTIC ANALYSES

2.1. Materials

The materials for the study were 21 English word pairs; in each pair one initial syllable had primary, the other secondary stress. The two first syllables always contained the same segments, and the onset of the second syllable was likewise matched. The 21 pairs were: *booking, bouquet; campus, campaign; carton, cartoon; cashew, cashier; convent, convex; distance, distinct; district, distress; diver, divert; harpist, harpoon; humid, humane; impact, impress; influence, inform; liquid, liqueur; massive, masseur; motive, motel; music, museum; mystic, mistake; robot, robust; ruler, roulette; typhus, typhoon; union, unique*.

Each word was recorded in two contexts by a female speaker of Australian English. The contexts were short non-constraining sentences, of the type *They both approved of the...*, or *We were sure the word was...*. The recording was made on Digital Audio Tape with a sample rate of 44.1 kHz, downsampled to 16 kHz and stored on computer.

For the two word recognition experiments, each word was truncated, leaving only the first syllable plus part of the onset of the second syllable. This method of truncation maximised the material on which listeners could base word recognition decisions. Further, in words with initial primary stress and a weak second syllable with a single intervocalic consonant (e.g., *music*, *diver*), that consonant is held to be ambisyllabic and hence to belong simultaneously to both syllables, so that including information about that consonant avoids potential presentation of an incomplete syllable. Truncation was carried out using xWaves software.

2.2. Results

Using initially xWaves software, and later PRAAT, we computed for each of the truncated syllables the duration, the minimum, maximum and mean F0 in Hertz, and the standard deviation of this mean, the mean and standard deviation rms amplitude, and the spectral tilt (also called spectral balance). The duration was computed over the whole syllable, while F0 and amplitude measures were calculated over voiced samples only. For spectral tilt we compared the relative energy in the frequency regions containing the first, second and third formants versus the region containing the fundamental, using the method suggested in [10].

Table 1: Mean values on eight acoustic measures for initial syllables with primary (Str 1, e.g., *mus-* from *music*) versus secondary stress (Str 2, e.g., *mus-* from *museum*), with significance level of the difference between them (** = p< .001; * = p < .05) and effect size (Cohen's *d*) of the difference.

	Str 1	Str 2	sign.	*d*
duration (ms)	381	350	**	.71
min F0 (Hz)	183	161	**	1.30
max F0 (Hz)	224	202	**	.78
mean F0 (Hz)	208	180	**	2.69
sd F0 (Hz)	12.9	11.3	n.s.	.23
mean rms	641	511	**	.96
sd rms	229	174	*	.46
spectral tilt	.909	.202	**	.82

Table 1 shows the mean values averaged across the 42 (21 pairs x two contexts) syllables of each type. As can be seen, the primary-stressed syllables are longer, louder, have higher pitch, are more variable in F0 and amplitude, and have a greater proportion of energy in the higher frequency bands of the spectrum. There were significant statistical differences between the two syllable types, with $p < .001$, on all measures except standard deviations of amplitude ($p < .05$) and of F0 (n.s.). Effect size qualified as "large" (> 1) for F0 measures only.

Each word had been recorded twice; the two recordings of each word correlated positively on all measures, in nearly all cases with $p < .001$.

3. CORRELATIONS WITH LISTENING

3.1. 2AFC identification task

26 Dutch-speaking university students took part in this study, reported in [9]. The materials were all first syllables of the stimulus words above, without sentence contexts: 168 fragments in all (21 pairs x 2 words x 2 contexts, each heard twice). Listeners chose for each fragment, heard over headphones, the word (e.g., MUSIC-MUSEUM) they judged to be the fragment's source; overall, 72.34% of their choices were correct. (Overall percent correct for native listeners in the same task was 59.17%.)

Correlations compared acoustic measures for each item against the item's mean percent correct response choices. Higher values on any measure should associate with more correct responses to primary-stressed syllables, lower values with more correct responses to secondary-stressed syllables; greater differences between the two members of a pair should increase all correct responses. We report all effects in line with these predictions with a significance of at least $p < .1$.

For primary-stressed syllables, there were more correct responses with higher mean rms amplitude (r [41] = .304, p = .05), and with greater difference between the members of a word pair in duration (r [41] = .26, p = .098), in maximum F0 (r [41] = .283, p = .07), and F0 standard deviation (r [41] = .27, p = .084). The same maximum F0 effect was visible for secondary-stressed syllables (r [41] = .259, p = .097). There were no further significant effects on any measure. Notably, no effect was significant in the unpredicted direction.

These results, all in the predicted direction, show that these Dutch listeners are sensitive to the suprasegmental features of English primary versus secondary stress. F0 information seems to be used more consistently than other acoustic properties.

3.2. Cross-modal fragment priming task

56 native Dutch speakers from the same population took part in this study, reported in [9]. Truncated words, with preceding sentence context (e.g., *We were sure the word was mus-*), served as primes. Lexical decisions to visually presented words (e.g., MUSIC) were compared as a function of whether the prime had come from the same word, from the other member of the pair, or from a control word.

Since the influence of the spoken primes is here inferred indirectly from decisions to visual words, no direct relationship to acoustics is expected. Of particular interest, though, are mismatch conditions (e.g., MUSIC on the screen with a prime syllable from *museum*, or vice versa). We examined whether these responses reflected the degree of acoustic mismatch in stress pairs.This was the case for mean F0, but interestingly, only in the case of secondary-stressed words such as MUSEUM. The greater the difference in mean F0 between two members of a pair, the longer Dutch listeners took to respond yes to MUSEUM given a prime that was actually from *music* (r [20] = .523, p = .015).

3.3. Gating task

17 Dutch subjects from the same population took part in a gating study using the same materials. The words were presented incrementally, each step 50 ms longer than the step before. Listeners guessed the word after each increment. The mean number of correct responses by the 17 listeners across items after 40%, 60%, 80% or all of the initial syllable is shown in Table 2. Any response with correct segments and stress pattern is included (the response *musician* for fragments of *museum*, etc.).

Table 2: Gating experiment: Mean number of segmentally and suprasegmentally correct guesses across 17 listeners, as a function of proportion of initial syllable heard, for initial syllables with primary (Str 1) vs. secondary (Str 2) stress.

	Str 1	Str 2
40 %	5.93	4.29
60 %	6.02	4.74
80 %	6.81	6.93
100 %	7.79	7.71

Listeners were better at guessing words with initial than with non-initial stress; this could be due to differences in acoustic clarity, but it could also be a bias to respond with the more typical words, as noted earlier. However, even by the end of the first syllable, less than 50% of responses fell into this generously defined 'correct' category. Gated speech materials appear to constitute a very hard recognition task for L2 listeners.

Comparison to the acoustic measures showed more correct responses to longer syllables than to shorter syllables in the later gates. This may actually reflect amount of phonetic information (i.e., the first syllable in *campus* is longer than in *robot*, but also contains more different phonemes).

There was also influence of F0. For primary-stressed syllables, greater within-pair difference in maximum F0 led to more correct guesses at the 100% gate (r [41] = .429, p <.005) and in standard deviation of F0 at 60% (r [41] = .31, p < .05), 80% (r [41] = .405, p < .01) and 100% (r [41] = .524, p < .001). For secondary-stressed syllables, greater within-pair difference in minimum F0 gave more correct guesses at the 60% (r [41] = .314, p < .05) and 100% gates (r [41] = .335, p <.05) and a small effect in mean F0 at 60% (r [41] = .28, p < .075). There were no effects of amplitude or spectral tilt.

3.4. Comparison with native listening

For the results described in sections 3.1 and 3.2, comparison data from native listeners was also available. None of the six correlations reported above for Dutch listeners (five in 3.1, one in 3.2) appeared in the native data. In cross-modal priming no acoustic effect at all reached significance in the native data. In 2AFC, five significant effects of duration, amplitude and spectral tilt were in the opposite direction to that predicted. Effects of F0 appeared in the predicted direction, however. More correct choices on primary-stressed syllables were made the higher the maximum F0 (r [41] = .344, p < .03) and the greater the difference within a pair in mean F0 (r [41] = .398, p < .01). For secondary-stressed syllables, more correct choices were made with lower mean F0 (r [41] = -.276, $p < .08$), or minimum F0 (r [41] = -.283, $p < .08$). Contrary to predictions, the greater the within-pair difference on F0 standard deviation (the one insignificant acoustic difference), the fewer the correct choices for secondary-stressed syllables.

4. DISCUSSION

Several conclusions are motivated by our analyses. First, in the present materials, and by extension in English words in general, primary- and secondary-stressed syllables are different on every relevant suprasegmental dimension: F0, duration, amplitude and spectral tilt. The suprasegmental differences between the two syllable types are there for any listener to make use of in spoken-word recognition.

The correlation analyses of the prior word recognition results showed that the Dutch listeners had not been selectively sensitive to any single acoustic dimension, although F0 information was, across the various analyses, more consistently exploited than the other dimensions. The new gating results confirmed the stronger role of F0.

In 2AFC, listeners must base their decisions on acoustic properties of the syllables (otherwise they would just be guessing). Enough time is available for decisions. The native listeners proved to have also exploited F0 information in 2AFC. However, their judgements reflected incorrect interpretations of the acoustics as often as correct interpretations; the correlations were inconsistent, with roughly half going in the predicted direction and half against predictions. The correlations displayed by the Dutch listeners, however, were uniformly in the predicted direction. Our second conclusion, then, is that Dutch listeners use suprasegmental information more effectively than English listeners do, even suprasegmental information in English.

Indeed, contrary results for use of spectral tilt in stress judgements in Dutch and English [11,12] may reflect listener tendencies more than intrinsic informativeness of spectral tilt in each language.

The basis of Dutch listeners' better performance is presumably their accrued experience of effective exploitation of suprasegmental cues to word stress in their native language. Our third conclusion, then, is that the present results are fully consistent with all the behavioural evidence that suggests that Dutch listeners make more use of stress in spoken-word recognition [6,7,8] than English listeners do [3,4,5]. In the online cross-modal priming task, consistent with this, no correlation with the acoustics appeared in the native data. In the Dutch data, however, responses to targets like MUSEUM given a mismatching prime were affected by the acoustics. We suggest that Dutch listeners have, based on their listening experience, an idea of the suprasegmental form of unstressed syllables with full vowels (as in *mus*[*eum*]). The suprasegmental form of a primary-stressed syllable (as in *mus*[*ic*]) fails to match this template, so that lexical decision responses are slowed. English-speakers lack such a template, and so incur no such mismatch effect.

5. ACKNOWLEDGEMENTS

Thanks to our speaker (Frances from Brisbane). This research was supported by NWO-SPINOZA.

6. REFERENCES

[1] Trommelen, M., Zonneveld, W. 1999. Word-stress in West-Germanic: English and Dutch. In: Hulst, H. van der (ed), *Word Prosodic Systems in the Languages of Europe* Berlin: Mouton, 478-515.

[2] Cutler, A., Pasveer, D. 2006. Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. *Proc. Speech Pros. 2006*, Dresden, 237-400.

[3] Small, L.H., Simon, S.D., Goldberg, J.S. 1988. Lexical stress and lexical access: Homographs versus non-homographs. *Perception & Psychophysics* 44, 272-280.

[4] Slowiaczek, L.M. 1990. Effects of lexical stress in auditory word recognition. *Language and Speech* 33, 47-68.

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

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

[7] Cutler, A., Donselaar, W. van 2001. *Voornaam* is not a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech* 44, 171-195.

[8] Donselaar, W. van, Koster, M., Cutler, A. 2005. Exploring the role of lexical stress in lexical recognition. *Quarterly J.Experimental Psychology* 58, 251-273.

[9] Cooper, N., Cutler, A., Wales, R. 2002. Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech* 45, 207-228.

[10] Ortega, M., Prieto, P. in press. Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In: Prieto, P., Mascaró, J., Solé, M.-J. (eds), *Segmental and Prosodic Issues in Romance Phonolog*y. Amsterdam: John Benjamins.

[11] Sluijter, A.M.C., Heuven, V.J. van, Pacilly, J.J.A. 1997. Spectral balance as a cue in the perception of linguistic stress. *J. Acoust. Soc. Am.* 101, 503- 513.

[12] Campbell, N., Beckman, M.E. 1997. Stress, prominence and spectral tilt. In: Botinis, A., Kouroupetroglou, G., Carayiannis, G. (eds), *Intonation: Theory, Models and Applications (Proc. ESCA Workshop,1997).* Athens.