

THE ROLE OF LEXICAL STRESS IN THE RECOGNITION OF SPOKEN WORDS: PRELEXICAL OR POSTLEXICAL?

Willy Jongenburger and Vincent J. van Heuven
Dept. Linguistics/Phonetics Lab., Leiden University/
Holland Institute of Generative Linguistics, The Netherlands

ABSTRACT

In this study we investigate to what extent lexical stress information is used to narrow down the cohort of potential word candidates. Our gating data on Dutch minimal stress pairs showed that lexical stress information is not used in the activation phase of the word recognition process, but does contribute to the prelexical selection stage.

INTRODUCTION

In stress-accent languages such as English, German and Dutch the position of the stressed syllable varies from one word to the next. Information on the position of the stressed syllable might contribute to the human word recognition process. Very few studies investigated to what extent lexical stress narrows down the cohort of potential word candidates and so far, the experimental data present a confusing picture. There is evidence that the context preceding an accented monosyllabic word contains prosodic cues about the stress pattern of this word: the melodic and rhythmic organisation of the preceding context tells the listener when to expect an accented syllable [1,2].

Dutch listeners performing a gating task with isolated words under optimal conditions only need the first syllable of the target word to know whether this syllable is (lexically) stressed or not. In LP filtered speech (750Hz, -48dB/oct), however, there was a strong bias for initially stressed responses, both for initially stressed targets and for initially unstressed targets [3]. Gating and shadowing experiments showed that stressed versus unstressed realisations of otherwise identical word-initial full syllables effectively narrowed down different cohorts of recognition candidates [4]. A control experiment [5] justified the conclusion that lexical stress realised on the target words is used in

the early word recognition process, and that the relevant cues are provided by the first syllable of the target string, rather than by the prosody of the preceding context. So, prosodic information, notably the difference between stressed and unstressed but segmentally identical word onsets, is used in the word recognition process.

These findings contradict claims for English that the effects of stress are located in the postlexical phases of the word recognition process only [6]. During the prelexical activation and selection phase minimal stress pairs in English proved to be functional homophones in an on-line cross-modal priming experiment.

Since gating provides information about the cohort of word candidates at different stages in the word recognition process, we ran gating experiments with Dutch minimal stress pairs in context, using not only high-quality speech but also segmentally degraded (LP filtered) speech. In segmentally degraded speech, the relative contribution to word recognition of segmental and prosodic information changes. Typically slowly varying prosodic information is more resistant to distortion of the speech signal than relatively fast varying segmental information. In LP filtered speech the time-span of the prelexical phase is increased, so that stress information gets a better chance of contributing to the prelexical recognition phases.

GATING STUDY

Our gating study included two conditions: hifi speech and LP filtered speech. In order to obtain speech of poor quality that is still sufficiently intelligible, a pilot study was carried out to establish the appropriate cut-off frequency for LP filtering (-48dB/oct). Individual cut-off frequencies were established for each target word, so as to

guarantee that the two members of a minimal stress pair are equally (un)intelligible. Individual cut-off frequencies varied from 1250 Hz up to 3000 Hz.

METHOD

A male speaker of standard Dutch recorded seven Dutch minimal stress pairs. The two members of each pair were embedded in a non-biased (semantically neutral) sentence as in:

Ze dacht dat haar vriend CAnon/kaNON opzocht, she thought that her friend canon / gun looked up

With the aid of a waveform editor these utterances were cut into fragments of increasing length, under visual and auditory control. For both quality and stress conditions the same truncation points were chosen. The first gate consisted of the preceding context plus the initial consonants and the first vowel onset of the target word. Each next fragment contained one diphone more, i.e. the second fragment included the initial consonant(s), vowel and onset of the next consonant, until the whole word and even the beginning of the next word had been gated. For each speech condition two stimulus series were prepared with the stress pattern of the target words counterbalanced; each series contained one member of each minimal stress pair.

Forty subjects participated; each stimulus series was presented on-line over headphones to ten subjects. Subjects were instructed to write down and say aloud after each fragment, the word they thought was being presented. They also had to indicate on a 10-point scale how confident they were as to the correctness of their response.

RESULTS AND DISCUSSION

In order to investigate to what extent lexical stress helps the listener to narrow down the cohort of potential word candidates, written responses were analysed. In cases where the orthographic responses did not allow us to unambiguously establish the stress pattern of the responses, the audio recordings were analysed instead. Monosyllabic content words were considered initially stressed, mono-

syllabic function words as initially unstressed. An Anova on the isolation point data shows only a large main effect of speech quality condition $F(1,214)=15.1, p<0.001$. Subjects need more acoustic information to isolate the target in LP filtered speech than in high-quality speech; 5.0 versus 4.0 gates. This means that the time span of the prelexical phase in LP filtered speech is indeed increased. Figure 1 presents the cumulative distributions of percent correct word responses and of initially stressed error responses as a function of gate length, broken down by the stress pattern of the target, for high-quality and LP filtered speech.

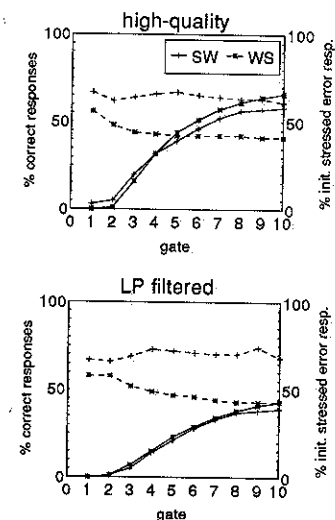


Figure 1. Percent correct word responses (solid lines) for high quality (upper panel) and LP filtered (bottom panel) speech and initially stressed error responses (dotted lines), as a function of gate length, broken down by the stress pattern of the target (SW versus WS).

For high-quality speech the percentage of initially stressed error responses obtained for initially stressed (SW) targets is

clearly higher than for initially unstressed (WS) targets at all gates. The difference in the proportion of initially stressed error responses between SW and WS targets is statistically significant at gates 3, 4 and 5 ($\chi^2 \geq 5.1$, $p < 0.02$). At earlier gates the differentiation is insignificant, at later gates the number of cases is too small to run statistical tests.

As to LP filtered speech it appears again that the proportion of initially stressed error responses is considerably larger for SW targets than for WS-targets. From the first gate onwards, stressed and unstressed word beginnings lead to different distributions of error responses. The differentiation assumes statistical significance (χ^2 between 5.1 and 19.2 for gates 3 through 8, $df=1$, $p < .05$) from gate 3 onwards, but, crucially, it is again insignificant during the first two gates, i.e. during the presentation of the initial syllable.

The observation that lexical stress is not heard during the first two gates suggests that prosodic information is not used during the activation phase. From the third gates onwards, however, listeners might use lexical stress in the selection phase. One may ask why the present results show only moderate differentiation of stressed and unstressed word beginnings, whilst much stronger differentiation was reported in the literature [2,3]. Several answers spring to mind. First, in the present experiments low-predictability words were embedded in uninformative contexts, whereas more easily available words were used in the earlier experiments, where they occurred either in isolation or in a slot in a carrier phrase. Second, in the earlier experiment with carrier sentences, listeners were provided with a typed version of the sentence up to the critical word, whereas our subjects had no information about the position of the target's onset: i.e. listeners did not know beforehand that the fragment ended with a word onset.

We will now try to show that lexical stress information is actually used in the selection phase of the word recognition process. If at the gate preceding the isolation of the target, i.e. one gate before the subject produces the target

word without subsequently changing his mind, the proportion of rhythmically correct error responses is considerably larger than at the same gate for all cases where isolation of the target does not follow, this would be an indication that prosodic information does in fact constrain the cohort of word candidates. Figure 3 presents the rhythmically correct error responses at the gate immediately preceding the isolation point for all cases where the listener did in fact reach an isolation point (+iso) as well as the corresponding rhythmically correct error responses at the same gate number when no subsequent isolation of a target followed (-iso).

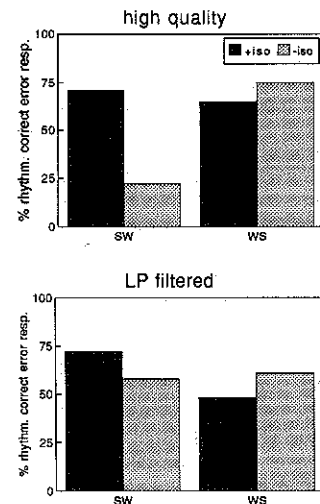


Figure 2. Percent rhythmically correct error responses for high-quality (upper panel) and LP filtered speech (lower panel) at the gate preceding isolation (+iso) and at the same gate position without isolation following (-iso).

The proportion of correctly stressed error responses in high-quality speech at the gate preceding isolation (+iso) for SW targets is considerably larger than at the

same gate when no isolation of the target follows (-iso); $\chi^2=7.7$, $p=0.006$. This means that listeners use prosodic information in the early phases of word recognition. The proportion of rhythmically correct error responses evoked by high-quality WS-targets does not differentiate between +iso and -iso. Although in LP filtered speech the proportion of correctly stressed error responses evoked by SW targets is larger for +iso than for -iso, this difference is insignificant.

In LP filtered speech the proportion of correctly stressed error responses evoked by WS targets in the -iso condition is statistically the same as in the +iso condition. This means that although listeners hear an unstressed initial syllable, they still are willing to reconsider a stressed onset syllable. These observations are in accordance with earlier findings [7] that unstressed syllables are not generally used by Dutch listeners to eliminate recognition candidates that begin with a stressed syllable, but that hearing stressed onset syllables effectively block access to that part of the mental lexicon containing initially unstressed words.

CONCLUSIONS

We will now recapitulate the main points of this paper. 1) Listeners proved unable in a gating task to differentiate between stressed versus unstressed beginnings of minimal stress pairs as long as no larger onset portion of the target was made audible than the first syllable. 2) Differentiation increased after the first syllable. Moreover, the subjects' word recognition was shown to be facilitated when the rhythmic pattern was correctly perceived before the isolation point was reached. We suggest, on the basis of these findings that lexical stress information is not used in the activation phase of the word recognition process, but still contributes to the prelexical selection stage.

Since the validity of the gating paradigm as a simulation of the on-line recognition process is subject to discussion, it is difficult to interpret whether these data falsify Cutler's claim that stress information does not play any

role at all in lexical access. Therefore we are currently running (on-line) cross-modal priming experiments [8] with the same experimental material used in this gating study.

ACKNOWLEDGEMENT

We acknowledge the financial support given by the Netherlands Organisation for Research (NWO) through the Foundation for Speech, Language and Logic, under project # 300-173-023.

REFERENCES

- [1] Cutler, A. (1976), Phoneme-monitoring reaction time as a function of preceding intonation contour, *Perception & Psychophysics* 20, pp. 55-60.
- [2] Cutler, A. & Darwin, C.J. (1981), Phoneme monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics* 29, pp.217-224.
- [3] Heuven, V.J. van (1984), Verslagen van de Nederlandse Vereniging voor Fonetische Wetenschappen, No. 159-162, pp 22-38.
- [4] Heuven, V.J. van, (1988), Effects of stress and accent on the human recognition of word fragments in spoken context: gating and shadowing, *Proceedings of the 7th FASEL/Speech-88 Symposium*, Edinburgh, pp 811-818.
- [5] Dalen, J. van & Noorloos, J. van (1989), De rol van klemtoonmarkering bij de lexicale access, unpublished paper, Leyden University.
- [6] Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access, *Language and Speech*, 29, pp. 201-220.
- [7] Heuven, V.J. van, (1985). Perception of stress pattern and word recognition: Recognition of Dutch words with incorrect stress position, *Journal of the Acoustical Society of America*, 78, s21.
- [8] Jongenburger, W. & Heuven, V.J. van, (1995), The role of linguistic stress in the time course of word recognition in stress-accent languages, *Proceedings of Eurospeech '95* (to appear).