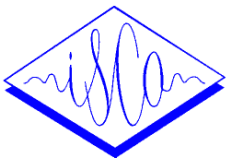

4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

THE ROLE OF LINGUISTIC STRESS IN THE TIME COURSE OF WORD RECOGNITION IN STRESS-ACCENT LANGUAGES.

Willy Jongenburger & Vincent J. van Heuven
e-mail: jongenburger@rullet.leidenuniv.nl; heuven@rullet.leidenuniv.nl
Dept. Linguistics/Phonetics Laboratory, Leiden University
Holland Institute of Generative Linguistics, The Netherlands

ABSTRACT

In stress-accent languages minimal stress pairs occur, i.e. identical segment sequences that are lexically distinct only by virtue of a difference in stress position. In these languages stress information might contribute to the human word recognition process. In this study we investigate the exact locus of stress effects in the human on-line word recognition process by means of a cross-modal priming study. The results are inconclusive.

1. INTRODUCTION

Models of the human on-line word recognition process typically distinguish a number of time-ordered phases [1].

i) Activation. On the basis of the first 200 ms of a word an initial cohort of recognition candidates is activated whose internal specifications are sufficiently compatible with the acoustic information heard so far.

ii) Selection. On the basis of following bottom-up information, as wel as of top-down information from contextual constraints, the set of candidates is narrowed down to one, at which point ("isolation point") lexical access takes place.

iii) Verification. All lexical information for the accessed word is made available to the recognition system, and checked against further bottom-up information.

If lexical stress information is used in the early activation phase, only word candidates that match the stressed/unstressed character of the acoustic input are activated. In the case that lexical stress information is not used during activation, word candidates are activated on the basis of segmental cues only, regardless of prosody. When lexical stress information is used during the selection phase, word candidates that no longer agree with the stress pattern of the growing acoustic input are de-activated, i.e. word candidates are selected on their stress pattern. If on the other hand, lexical stress is used only in the verification phase, minimal stress pairs are functionally homophonous during the prelexical phases of the on-line word recognition process. Cutler [2] used the cross-modal priming paradigm developed by Swinney [3] as a diagnostic test for homophony. Subjects made lexical decisions to visually presented words, related to one of the members of a minimal stress pair. Subjects listening to sentences containing either the initially stressed (SW, e.g. *FORbear*) or the initially unstressed (WS, *forBEAR*) member of a minimal stress pair, showed lexical decision facilitation for words related both to the SW member (*ancestor*) and to the WS member (*tolerate*), presented at the offset of the prime word in the sentence ("early probe position"). When the targets for lexical decision occur 750 ms after the offset of the prime word ("late probe position"), only targets related to the prime word are facilitated. Cutler concluded that English minimal stress pairs are homophones: lexical stress does not constrain lexical access.

It is surprising that such conspicuous properties as stress position should be systematically neglected during the prelexical stages of the word recognition process. Data obtained from gating experiments provided some evidence that listeners are able to take prosody into account [4,5]. Listeners performing a gating task on the members of Dutch minimal stress pairs embedded in a non-biased (semantically neutral) sentence, showed prelexical stress effects in the selection stage [6]. Listeners proved unable 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. Differentiation did not increase until *after* the first syllable. Moreover, in the case that a target word was isolated, subjects' responses at one and two gates immediately preceding the isolation point contained significantly more initially stressed responses for SW targets than for WS targets [7]. At the corresponding gates for items in case they were not isolated, the distribution of initially stressed responses is essentially the same for both SW and WS targets. On the basis of these findings we suggest 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 questionable, it is a moot point whether these data falsify Cutler's claim that stress information does not play any role in the prelexical phases of the word recognition process.

A second reason to re-examine the role of lexical stress in the prelexical recognition stages, is that Cutler's findings apply to optimal speech listened to under ideal circumstances. For segmentally degraded speech, however, the relative contribution to word recognition of segmental and prosodic information changes. Prosodic information is more resistant to distortion of the speech signal than segmental information, so that the relative importance of lexical stress information increases as speech quality and listening circumstances deteriorate. In segmentally degraded speech the time-span of the prelexical phase of the word recognition process is increased, so that the slowly varying prosodic information has more time to contribute to the prelexical recognition phases.

We ran (on-line) cross-modal priming experiments with Dutch minimal stress pairs in context sentences, in two speech quality conditions: hifi speech and segmentally degraded (Low Pass filtered) speech. Notice that exactly the same

experimental material was used in our earlier gating study [6].

2. PRELIMINARY EXPERIMENTS

In this section a series of preliminary experiments are discussed which were run in order to ensure a well chosen set of primes, targets, controls and non-biased context sentences.

2.1 Related and control words

In order to get words that are reliably associated with the members of the minimal stress pairs, an association experiment was conducted. One hundred and twenty subjects participated in the experiment and were asked to write down the very first word that entered their minds while listening to the stimulus word. Seven stress pairs appeared suitable for inclusion in the priming experiment. For each related word, i.e. associated to one member of a minimal stress pair, a control word was selected that was matched, as far as possible, in terms of frequency of occurrence, subjective familiarity (rated by the same listeners), number of syllables, stress pattern, length in letters and word class. Control words have no semantic relationship with either of the members of a stress pair. The related and control word of the SW prime *CAnon* 'song' are *zingen* 'sing' and *bellen* 'phone' respectively. The related and control word of the WS prime *kaNON* 'gun' are *oorlog* 'war' and *vinger* 'finger'.

2.2 Baseline lexical decision times

The priming effect is defined as the facilitation of lexical decision (in ms) of related words as compared to their control words. Therefore, it seemed necessary to control for possible intrinsic differences in lexical decision times between related and control words. Related and control words were presented in an unimodal lexical decision experiment to 40 subjects. An analysis of variance on the decision times by stress pattern, type of stimulus (related/control) and minimal stress pair produced no significant main effects. Nevertheless, lexical decision times in the bimodal priming experiment (§3) will be adjusted for the baseline response times by subtracting from each response time the mean baseline decision time for the same item. Unimodal lexical decision times are presented in table 1 (next page).

2.3 Non-biased sentences

One neutral, semantically non-biased sentence was constructed for each minimal stress pair of the type:

Ze dacht dat haar vriend CAnon / kaNON opzocht.

'she thought her friend canon / gun looked up'

Thirty-five subjects participated in a completion task. Subjects were presented with the written initial portion of the seven sentences up to the prime word, (i.e. member of a minimal stress pair) and were asked to choose one of the two members of a pair as the beginning of a completion or they could choose the option that both words were equally viable options. In 64% of all cases both words were considered as equally viable options. Subjects had a preference for SW words in 20% and a preference for WS words in 17%. We concluded from this experiment that our context sentences introduce no bias favouring one member of the minimal stress pair over the other.

2.4 Low Pass filter frequency

Cut-off frequencies for LP filtering (-48dB/oct) were established for each individual prime word so as to guarantee that the two members of a minimal stress pair are equally (un)intelligible. Twenty subjects listened to the sentences filtered at various cut-off frequencies, and wrote down what they heard. Word-individual cut-off frequencies for the primes varied from 1250 Hz up to 3000 Hz.

3. CROSS-MODAL PRIMING STUDY

In Cutler's study [2] visual targets for lexical decision in the *early probe condition* were presented at the offset of the auditive prime word in the sentence, assuming that this probe position is suitable to tap into the early phases of word recognition. However, to investigate the early prelexical phases of the word recognition process, it is necessary to probe at positions in the prime word before the isolation point. Because of intrinsic differences between the prime words in terms of the amount of acoustic information necessary for recognition, we decided to establish an individual probe position for each stress pair in each speech quality condition, cf. [8].

Probe positions were based on the gating data with the same experimental material as collected in [6]. For each speech condition (hifi, LP filtered) an early probe position was defined as the earliest isolation point for either of the members of a minimal stress. A third, late probe (hifi) condition was included in order to be able to tap into the verification phase, after recognition has taken place: targets appeared 750 ms after the prime word's offset. Summarizing, the priming study included three conditions:

- hifi speech with early targets (probes)
- LP filtered speech with early targets (probes)
- hifi speech with late targets (probes)

3.1 Method

Seven Dutch minimal stress pairs embedded in their non-biased sentence had been recorded by a male speaker of standard Dutch. Two lists were prepared, each containing seven non-biased sentences with the stress pattern of the prime words equally divided between, and counterbalanced across, the two lists. These sentences were interspersed with seven filler sentences. For each prime word (sentence) there was a set of four possible visual lexical decision targets:

- the word related to the SW member
- its control word
- the word related to the WS member
- its control word

Four lists with lexical decision targets were prepared. Each list contained one target word of the set for each prime. Type of target word was blocked and counterbalanced across lists. The seven filler sentences were combined with non-word targets. Experimental trials were preceded by 48 practice trials (24 words and 24 non-words).

Timing and data collection were under the control of a Silicon Graphics computer. Each condition was presented to 40 subjects, yielding a total of 120 subjects. Attention to and intelligibility of the auditory stimuli were tested by a sentence completion task immediately following each stimulus. The sentence the subjects had just heard was now presented

Table 1: Mean baseline LD-times (in ms) for related words and their control words, analysed for the stress pattern of the source prime word, broken down by minimal stress pair.

minimal stress pair	SW-related	LD-time	SW-control	LD-time	WS-related	LD-time	WS-control	LD-time
VOORkomen-voorKOmen 'appearance-prevent'	*uiterlijk* 'appearance'	652	*werknemer* 'employee'	589	*genezen* 'cure'	599	*bevelen* 'command'	601
SERvisch-serVIES 'Serbian-chinaware'	*oorlog* 'war'	586	*vinger* 'finger'	579	*kopje* 'cup'	645	*motor* 'motor'	588
VOORruit-voorUIT 'windscreen-forward'	*auto* 'car'	549	*orde* 'order'	575	*opschieten* 'hurry up'	604	*kwijtraken* 'lose'	676
CAnon-kaNON 'song-gun'	*zingen* 'sing'	611	*bellen* 'phone'	566	*oorlog* 'war'	586	*vinger* 'finger'	579
DOORlopen-doorLOpen 'move on-pass through'	*opschieten* 'hurry up'	604	*kwijtraken* 'lose'	676	*school* 'school'	526	*vriend* 'friend'	554
PLAto-plaTEAU 'Plato-plateau'	*filosofie* 'philosophy'	679	*amerikaan* 'americain'	648	*berg* 'mountain'	571	*tong* 'tongue'	598
Avonduur-avonTUUR 'evening hour-adventure'	*laat* 'late'	587	*goed* 'good'	530	*spannend* 'exciting'	597	*dwingend* 'compelling	686
mean		610		595		590		612

visually with the target word replaced by three dots. Subjects were asked to name the missing (prime) words. Lexical decision times of targets which were followed by incorrect naming were excluded from the dataset (7.7% of all data). Note that responding the prime's stress partner was not considered as a naming error (this situation occurred in 3.0% of the cases).

3.2 Results

In table 2 the mean lexical decision (LD) times are presented. Here LD-times have been adjusted for baseline reaction times (§2.2) by subtracting from each individual response time the mean baseline LD-time for the same item. We notice, first of all, that LD in the cross-modal priming experiment takes considerably (some 250 ms) longer than in the unimodal task, indicating higher processing load in the cross-modal priming design. In table 2 LD-times are broken down by the stress pattern of the prime word (SW and WS) and the target type (word related to prime word in sentence, its control word, word related to prime word's stress partner and its control word). In figure 1 the mean priming effect, defined as the adjusted mean LD-time for control words minus the adjusted mean LD-time for related words (i.e. Δ in table 2) is presented, both for words related to the prime in the sentence ("prime") and for words related to the prime word's stress partner ("partner"), broken down by stress pattern.

The results show no effect of priming at all. In each of the three probe/quality conditions the LD-times for targets semantically related to a prime word (or to the prime's stress partner) are *longer* than those obtained for the non-related control words. This result indicates that the on-line recognition of the semantically related words is not facilitated by the presence of the auditory prime. This is in contrast with previous runs of this experiment, in which we did find (small) priming effects as predicted [7]. Although the present experiment is rendered inconclusive by this general result, we will proceed with an analysis of the results broken down by probe/quality condition.

The results for *hifi speech with early targets* show that neither SW nor WS primes facilitate the recognition of their related targets: although SW primes do slightly facilitate LD of words related to the stress partner (14 ms), WS primes do not. If lexical stress is used in the prelexical stage, only the recognition of the LD-target related to the prime, and not the target related to the prime's stress partner, should be facilitated. This effect clearly does not obtain. If stress is not used during the prelexical stage SW and WS words are functionally homophonous, so that both the LD targets related to the prime word and to its stress partner should be facilitated. This effect is not found, either.

The effects found with *LP filtered speech with early probes* are qualitatively the same as those obtained in the corresponding hifi condition, when the LD targets are semantically related to the prime word, be it that the WS primes are more distracting in filtered than in hifi speech. However, the difference between SW and WS primes is neutralized when the visual LD targets are semantically related to the auditory prime's stress partner. This latter finding renders these results uninterpretable.

As for the *hifi speech condition with late probes*, it appears that neither SW nor WS primes facilitate the recognition of their related targets, nor do they facilitate the LD of words related to the prime's stress partner. If lexical stress is used in the postlexical stage, only LD of words related to the prime should be facilitated. This effect is not found.

4. DISCUSSION

With the present Dutch experimental material we were not able to replicate Cutler's findings for hifi speech. The effects found were not as predicted. Yet, in earlier gating studies [6,7] we did find an early stress effect, viz. in the prelexical selection phase (see §1). Probably, the results obtained in the gating study set an upper boundary to what listeners are able to do with bottom-up information in on-line word recognition. In additional experiments [7] we explicitly instructed Dutch listeners to determine the stressed versus unstressed nature of

Table 2: Mean lexical decision times (in ms), adjusted for baseline lexical decision times, broken down by stress pattern of the prime word (SW and WS) and the target type, for three experimental conditions.

condition	stress pattern of prime word	LD target related to prime word			LD target related to prime's stress partner		
		related target	control target	Δ	related target	control target	Δ
hifi, early probes	SW	279	274	-5	199	213	14
	WS	254	205	-49	266	211	-55
LP filtered, early probes	SW	267	267	0	284	239	-45
	WS	300	217	-83	253	235	-18
hifi, late probes	SW	264	225	-39	245	211	-34
	WS	284	261	-23	267	246	-21

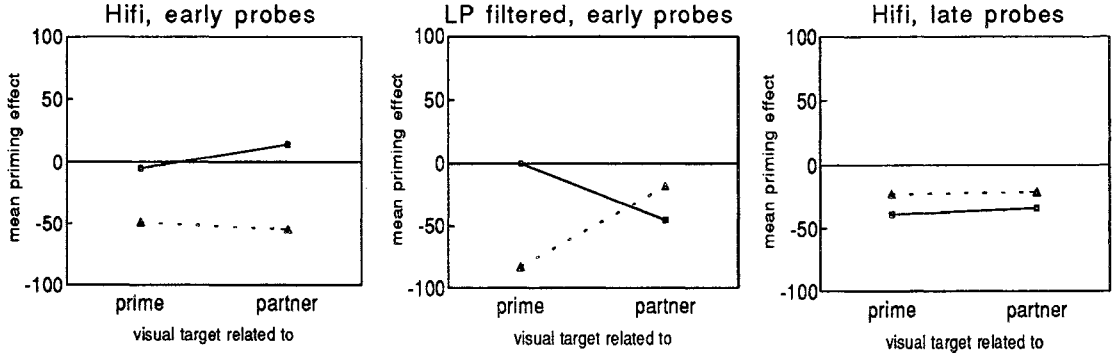

Figure 1: Mean priming effect (adjusted LD-time controls minus adjusted LD-time related words) for words related to the auditory prime ("prime") and for words related to the prime word's stress partner ("partner"), broken down by stress pattern.

the initial syllables in the members of our minimal stress pairs presented in LP filtered conditions, in a binary forced choice categorization task (i.e. not a word recognition task). Our subjects proved perfectly able to determine the difference between stressed versus unstressed word onsets *from the first gate* onwards. So the perceptual cues for the contrast between stressed and unstressed word beginnings are available to the listener even during the activation stage, but they are not (fully) used in the word recognition process. They are only used in the selection stage during off-line gating, and they are only used post-lexically (i.e. during the verification stage) in on-line (cross-modal priming) tasks. This state of affairs runs counter to the widely accepted idea that the human listener, when performing on-line word recognition, uses any source of information available in the speech signal at the earliest possible moment. At this time we cannot offer a solution to this dilemma. Nevertheless, it is obvious that some insight into this matter should be provided sooner or later by any comprehensive theory of the auditory word recognition process.

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] Frauenfelder, U.H. & Tyler, L.K. (1987). The process of spoken word recognition: An introduction, *Cognition*, 25, pp. 1-20.

[2] Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access, *Language & Speech*, 29, pp. 201-220.

[3] Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, pp. 645-659.

[4] Heuven, V.J. van (1984), Segmentele versus prosodische effecten van klemtoon op de woordherkenning, Verslagen van de Nederlandse Vereniging voor Fonetische Wetenschappen, No. 159-162, pp. 22-38.

[5] 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 FASE/Speech-88 Symposium*, Edinburgh, pp. 811-818.

[6] Jongenburger, W. & Heuven, V.J. van (1995). The role of lexical stress in the recognition of spoken words: prelexical or postlexical?, *Proceedings of the 13th International Congress of Phonetic Sciences* (in press).

[7] Jongenburger, W. The role of lexical stress in spoken word recognition: prelexical or postlexical, doct. dissertation, Leiden University (in preparation).

[8] Zwitserlood, C.M.E. Words and sentences: the effects of sentential semantic context on spoken word processing, doct. dissertation, Nijmegen University, 1989.