Perception of stress pattern and word recognition:

Recognition of Dutch words with incorrect stress position

[Expanded version of paper J12 presented at the 110th Meeting of the Acoustical Society of America, 4 - 8 November, 1985, Nashville, TN, USA. Abstracted in *Journal of the Acoustical Society of America*, 78, 1985, S21.]

Vincent J. van Heuven
Dept. of Linguistics/
Phonetics Laboratory Leyden University
P.O. Box 9515
2300 RA Leiden
The Netherlands

July 1986

1. Introduction

1.1. Stress versus segments

In languages such as Dutch and English words can usually be recognized through identification of the constituent phonemes, without invoking the help of prosodic information such as stress. As a result of this, none of the current models for human word recognition explicitly considers the possible role of stress or rhythmic patterning in narrowing down the set of candidates. All these models map the incoming acoustic segments onto the stored lexical items as the segmental information enters the auditory system in its "left-to-right" order, and continue to do so until one of the stored items is sufficiently and uniquely compatible with the input segment string. Also automatic word recognition systems typically proceed on the basis of segmental information, matching spectral characteristics of the input signal to those of stored templates while leaving prosodic information out of consideration.

Yet it would appear feasible, for instance, to partition the lexicon of the language into a number of rhythm types, based on the number of syllables and the position of the stress within the array. Naturally, this information by itself is hopelessly insufficient do narrow down the number of competing recognition candidates to just one, but it would certainly help to limit the number of alternatives. Segmental information, on the other hand allows for a far greater number of lexical distinctions, but these are predominantly carried by rather subtle spectral differences that easily get distorted or masked in averse speech conditions. Rhythmic information, in contrast, is expressed by slowly varying prosodic parameters, and is therefore much more robust.

On the basis of this view we assign prosody a role of primary importance in the process of word recognition. However, under good speech conditions this importance does not surface, but remains dormant or latent. The true importance of stress and rhythm type will only come to light when speech quality deteriorates, as for instance in synthetic speech.

1.2. Effects of stress on word recognition

Nooteboom & Doodeman (1985) found recognition scores at about 70% for a set of Dutch 3-syllable words synthesized from diphones without prosodically marked stress position. However, when the stress position was marked by a pitch excursion and/or relative lengthening, recognition of the same words rose to about 85%.

In experiments like these it is impossible to determine exactly when and how the availability of prosodic information is used by the listener in the on-line recognition process. On the basis of some word recognition models (e.g. Marslen-Wilson, 1980) one would expect the listener to exploit any bit of information as early as possible, rather than await the end of the word. In order to trace the effect of prosodic information on recognition as the acoustic stimulus develops in time, Nooteboom & Doodeman (1985) adopted the gating method of presentation (cf. Grosjean, 1980). The listener first heard the initial CV-combination of a stimulus word and had to guess what word would eventually be presented. On successive presentations an ever larger portion of the word was made audible, until the listener was able to correctly determine the identity of the word. Nooteboom & Doodeman found that stimulus words could be recognised from shorter gated fragments when the stress position was prosodically marked (by a pitch-accent) than when the stimuli were prosodically uncorrected concatenated diphones. The advantage of stress marking was strongest for words with medial stress, weak for finally stressed words, and absent for words with initial stress.

1.3. Stress bias

In a subsequent error analysis of the responses obtained in these and similar gating experiments, Van Heuven (1984) found that listeners assume stress on the first syllable of the target word, irrespective of the actual stress position in the word. The proper stress position is not reflected in the error responses until the actually stressed syllable (in medial or final position) has been made audible. The overwhelming bias for initial stress was, however, suppressed when the segmental information was of good quality (i.e. significantly better than that of synthetic speech); also embedding the poor quality target (either synthetised speech from diphones, or LP-filtered natural speech) in a short, fixed carrier reduced the initial bias @somewhat,@ presumably because this provides a frame of reference within which the weight of the initial target syllable can be evaluated.

bias: effect or artifact?

One may argue, of course, that the bias for stress on the first syllable is an artifact of the gating procedure. For one thing, the subject is forced to respond with complete words. It may then be the case that initially stressed words more readily spring to mind. On the other hand, there are good reasons to believe that the bias is perception-based rather than response-based.

An initial stress bias has been observed in numerous experiments that did not involve word recognition tasks. Van Katwijk (1974) reports that the first of an array of three identical syllables /soesoesoes/ was invariably judged to be stressed by Dutch subjects. This bias could only be overcome by lengthening one of the other syllables, or by introducing a pitch movement there. Similarly, Berinstein (1979) synthesised "words" containing four identical syllables /bI/, and then varied the duration of each of the four vowels separately, while leaving the remaining three at a standard duration of 100 ms. When all four syllables had equal duration, English listeners heard stress on the first syllable. Lengthening one of the non-initial syllables in excess of 40 ms was needed for listeners to perceive a stress shift.

Word recognition based on bias

I submit that Dutch (or English) listeners proceed from a default recognition strategy that assumes the first syllable of a target to bear the stress. Their assumption will prove correct in the majority of the cases, but will be given up during the recognition of a word as soon as compelling counter-evidence comes available. This may occur at a very early stage if - in high quality natural speech - segmental information, e.g. vowel and consonant reduction, points towards an unstressed initial syllable. In poor quality speech the default stress assumption will be upheld until the true stress position is revealed in due course by the presence of a conspicuous pitch movement, or by lengthening, or both in one of the later syllables.

This strategy would lead us to predict an asymmetrical effect on word recognition of deliberately misplaced stresses on word recognition. If an initial stress is shifted away to a later position, the listener would still start his word isolation process from the biassed assumption that the initial syllable is stressed. When during the second or third syllable the true stress position is detected, the number of likely candidates has already shrunk to the point where the wrong stress is harmless. However, when a word with lexical stress in the second or third position is incorrectly pronounced with initial stress, the recognition process will be strongly impeded. The prosody will trick the listener into believing unconditionally that a word with initial stress is being spoken. Consequently, that part of the mental lexicon will be de-activated that contains words with non-initial stresses. At no point during the remainder of the stimulus word will the listener receive prosodic information signalling his erroneous decision, so that correct word recognition will often fail.

1.4. Approach

The viability of this account was provisionally tested in two small, related experiments. Crucially, we examined the inhibiting effect on word recognition of incorrectly placed stress (pitch accents). This approach has been adopted from Cutler & Clifton (1983) who found that word recognition (as measured by a semantic decision task) was about 15% slower when an English di-syllabic word was pronounced with stress on the wrong syllable. Counter to our prediction, their results show no interaction between lexical stress position and correct versus incorrect stress placement, at least not when the decision latencies were corrected for word duration. In our experiments we mainly adopted techniques that do not, or not exclusively, rely on reaction time measurement.

In the first experiment we used the gating method of presentation. Since this method is still open to criticism, a real-time recognition task was used in the second experiment. We argue that both experiments reveal the same type of (predicted) effects. In both tests we used synthetic speech, so as to obtain correct and incorrect exemplars of the same word, without affecting other factors such as segmental quality.

2. Experiment I: gating

2.1. Method

Stimuli were 20 di-syllabic Dutch nouns from the low frequency brackets of the lexicon, with a uniform segmental build-up CVCVC. Ten words had lexical stress on the first syllable, 10 more on the second (for a full listing of the set see appendix I). The unstressed syllable always contained a full vowel (i.e. no schwa). The words were synthesised from diphones using a Philips MEA8000 speech synthesiser (Brueck & Van Teuling, 1982) controlled by an Apple IIe microcomputer. Diphones are parametrised stretches of speech running from about the centre of one phone until about the centre of the following phone, as spoken by a human speaker in fluent speech.

In our system (Elsendoorn & 't Hart, 1982, 1984) the diphones are extracted from the originally accented syllables in nonsense words of the type /C@'CVC@/. Of each word, two exemplars were synthesised, one with an accent on the first syllable and one with accent on the second. Accents were implemented as a 5 semitone rise from and subsequent fall to the declination line. The pitch peak was placed 32 ms after the vowel onset. The declination was set at 5 semitones per second, and the pitch changes during the rise/fall were 75 semitones per second. Vowels in non-stressed initial syllables were shortened to 80% of their original duration.

The 2*20 words were presented to 2 groups of 10 Dutch listeners, such that each subject heard each lexical word only once, with 10 correct and 10 incorrectly stressed words in random order. Each word was presented 5 times with 7 s. intervals in between successive presentations. At the first presentation only the initial CV-combination was made audible until the centre of the vowel. On each of the following presentations the audible word fragment was lengthened by one diphone, until on the fifth gate the entire word was audible. After each fragment the subjects had to write down the complete word of which they believed they had just heard the initial portion, with an unlimited choice from the Dutch lexicon.

2.2. Results and discussion

Figure 1 plots per cent correctly recognised words as a function of the audible fragment's length, with separate curves for correct and incorrectly stressed versions, and with separate panels for lexically initial (A) and final (B) stresses.

here figure 1 A&B

The results are very much as predicted. Words with lexical stress on the first syllable do not suffer much from incorrect stress placement: after completion of the word, per cent correct is about equal for correct and incorrect versions (58 vs. 57%, respectively). However, during the development of the stimulus the recognition scores for the incorrect exemplars consistently remain below those of the correct versions. This may have been caused by the shortening of the initial syllable, which may have degraded its segmental quality.

When words with lexically final stress are correctly produced, their recognition is, again, on the order of 60%. As predicted, however, shifting the stress here to the wrong position has a clearly negative effect, resulting in some 20% lower recognition on completion of the stimulus presentation.

Finally, a rhythmic analysis was made of the error responses to the first syllable, i.e. accumulated over the first two gates. The results are as indicated in figure 2.

----
here figure 2

As is characteristic of poor quality speech, there appears a strong bias towards perceiving stress on the first syllable throughout, irrespective of the presence or absence of a prosodically marked stress: some 75% of the responses has initial stess, 10% has final stress, and for an other 15% the responses were ambiguous with respect to stress position.

3. Experiment II: real-time word recognition@

As we said in our introduction, one may legitimately object that this apparent bias is an artifact of the gating method. It may well be the case that instantaneous stimulus presentation would prompt the listener to postpone any use of prosodic information until either a clear stress is perceived, or even the end of the word has been reached. As a consequence, listeners might never go through a stage of excluding part of their lexicon on the basis of early information on the non-stressed nature of the initial syllable(s). If, on the other hand, the word isolation process is truly reflected in the gating task, the results of other, instantaneous recognition tasks should run parallel.

We therefore set up a second experiment in which the subject was simply asked to repeat the stimulus word, presented to him just once, as quickly as possible. Dependent variables are the correctness of the responses, and the repetition latency. This time tree-syllable words were used so as to provide a greater range of possible stress misplacements, which would allow us to test the differential effect of frontshifts and backshifts more criticially.

3.1. Method

Twenty-four morphologically simplex Dutch words of low frequency of occurrence were selected, evenly distributed over types with lexical stress in initial, medial, or final position. Appendix II lists the full set of words. Words were synthesised using the same procedure ᄀnd equipment as in the previous experiment.

Of each word three exemplars were synthesised, one with correct stress placement, and two with wrong stress position. Stresses were implemented by generating a pitch accent on the stressed vowel, executing a 30Hz pitch rise during 48 ms, followed by a 36 Hz fall for another 48 ms, such that the pitch peak occurred 32 ms after the vowel onset. The accent was superposed on a declination line that fell 1 Hz every 32 ms. The duration of the unstressed syllables was shortened to 70% of their original values, as copied from a naturally produced accented exemplar (see experiment I). Final syllables, however, were never shortened.

Three tapes were prepared such that each contained every word only once with equal distribution of words with lexical and actual stress in initial, medial, and final position. Eight words on each tape were correctly stressed, 16 had stress in a wrong position, again evenly distributed over the two possibilities.

The three tapes were presented to three groups of four subjects, who (after some practice with similar items) repeated the words as quickly as they could. Stimuli and responses were recorded on separate tracks of audio tape.

3.2. Results

Per cent correctly repeated words was determined, after excluding responses with latencies in excess of 3 seconds. Repetition latency, defined as the time lag between the onsets of stimulus and the corresponding response words, were collected using a Devices Digitimer D4030, and rounded off to the nearest 10 ms.

Table I presents the recognition scores in per cent correct, broken down by lexical and actual stress positions. Correctly stressed stimuli lie along the main diagonal in the matrix.

----
here table I

Correctly stressed words were recognised at about 70% correct, with a clearly better score for words with medial stress (81%). Misplaced stress exerts a very detrimental effect on word recognition in this type of task: no more than 37% of these stimuli were correctly recognised on average.

Crucially, a backshift of a lexically initial stress causes a relatively slight drop in recognition scores: 66% for correct stress versus 44 and 56% for incorrect stress in medial and final position, respectively. This amounts to an average drop of 11% for lexically initial stresses.

Words with lexically non-initial stresses suffer, as *predicted, very much more from incorrect stress placement,* with an average drop from 72% to a mere 31% correct.

The repetition latencies are given in table II. Here only those data have been processed that were collected for correctly recognised words with latencies below 3 seconds.

here table II

Words with correct stress patterns are repeated with a mean latency of 1480 ms, those with incorrect stress with 1660 ms. However, wrong stress position does no longer interact with lexical stress in the predicted way: the recognition of words with frontshifts of non-initial stress is delayed by 165 ms on average, but words with backshift of an initial stress are delayed even more (225 ms). Finally, we notice the odd (and so far inexplicable) effect that words with medial lexical stress are repeated faster when the stress in incorrectly placed in final position than when the stress is correct.

4. General discussion

By and large the results obtained in the two experiments, provide strong support for the essential correctness of our account of the role of stress bias in the recognition of spoken words. The predicted asymmetrical effects of back-shifting an initial stress (small drop in scores) versus front-shifting a non-initial stress (large drop in scores) were obtained in both experiments.

This asymmetry, to me, seems related to the asymmetrical behaviour of affixes in Dutch, and presumably in English as well. The position of the stress in Dutch stem morphemes is often backshifted under the influence of a suffix, which may either bear the stress itself, or attract the stress to a syllable one or two position before the suffix, as in English f'inal - fin'al+ity. Prefixes, however, (and affixes in general) never cause the stress to shift towards the beginning of a word, and are typically unstressable themselves.

It would appear that the role of stress and the observed position bias has to be explicitly accounted for in models of spoken word recognition. Clearly, the perception of a stress prompts a listener to reject (or de-activate) a large number of recognition candidates that do not share their stress position with that of the stimulus. However, leading (i.e. pre-stress) unstressed syllabes are not generally used to eliminate recognition candidates that begin with a stress.

Our results also underline the importance of the gating method as a research tool: the results obtained in this non-real-time task were essentially the same as those of the instantaneous recognition task. It could be objected, of course, that (correctly stressed) words in the instantaneous task were recognised some 10% better than in the gating task. This discrepancy is, quite probably, not a task effect, but caused by the greater word length (3 versus 2 syllables) in experiment II. Longer words are lexically more redundant, and will therefore be better recognised.

Finally, we may observe that measuring repetition latencies is not susceptible to all the types of effects that were predicted. Cutler & Clifton (1983) found effects of incorrect stress placement on the same order of magnitude in a semantic decision task (concrete vs. abstract referents of nouns), but likewise failed to uncover the predicted interaction with lexical stress pattern. Similarly, in our experiment II, the repetition latencies could not provide a basis to distinguish the predicted asymmetry of frontshifts and backshifts of stress position.

We see latency data as secundary evidence only. Reaction times are typically the result of complex processes involving msny unknown sources of variability. In the types of tasks used by Cutler & Clifton word recognition as such is followed by both a semantic decision and a motor activity (pressing a button); our own experiment involved at least a speech motor activity (viz. pronouncing the word, after correcting the stress position when applicable). We therefore take the view that the observed percentages of correct word naming, obviously involving word recognition (or else the stress pattern would not have been corrected), provide a much more reliable source of information.

Acknowledgements

Experiment I was carried out by Peter Hagman and Ludmila Menert, experiment II by Louise Muller and Hannie Nederlof. The technical support given by the Institute for Perception Research (IPO) at Eindhoven is gratefully acknowledged, with special thanks to Ing. Th.A. de Jong and Dr. B.A.G. Elsendoorn for writing the concatenation programme, and making the diphone data set available to us in Apple IIe readable format. The parameter editing programme for the Apple IIe was developed by Ing. J.J.A. Pacilly of our own laboratory.

References

Berinstein, A.E. (1979). A cross-linguistic study on the perception and production of stress, UCLA Working Papers in Phonetics, 47, 1-59.

Brueck, H.D. van, Teuling, D.J.A. (1982). Integrated voice synthesizer, Electronic Components and Applications, 4, 72-79.

Cutler, A., Clifton, J. (1983). The use of Prosodic information in word recognition, in H. Bouma, D. Bouhuis (eds.): Attention and Performance, 10, Erlbaum, London, 183-196.

Elsendoorn, B.A.G., Hart, J. 't (1982). Exploring the possibilities of speech synthesis with Dutch diphones, IPO Annual Progress Report, 17, 63-65.

Elsendoorn, B.A.G., Hart, J. 't (1984). Heading for a diphone speech synthesis system for Dutch, IPO Annual Progress Report, 19, 32-35.

Grosjean, F. (1980). Spoken word recognition and the gating paradigm, Perception and Psychophysics, 28, 267-283.

Heuven, V.J. van (1984). Segmentele versus prosodische effecten van klemtoon op de woordherkenning [Segmental versus prosodic effects of stress on word recognition], Verslagen van de Nederlandse Vereniging voor Fonetische Wetenschappen, 159/162, 22-38.

Katwijk, A.F. van (1974). Accentuation in Dutch, and experimental linguistic study, Van Gorcum, Assen.

Marslen-Wilson, W.D. (1980). Speech understanding as a psychological process, J.D. Simon (ed.): Spoken language generation and recognition, Reidel, Dordrecht, 39-67.

Nooteboom, S.G., Doodeman, G.J.N. (1985). Cues for lexical stress recognition of polysyllabic words, synthesized from diphones and presented in isolation, paper presented at the 109th Meeting of the ASA, Austin, TX, April 8-11, 1985.

Table and figure captions

Table I: Per cent correctly repeated words broken down by lexical stress position and actual stress position. Correct stress patterns lie on the main diagonal. Off-diagonal cells represent stimuli with incorrect stress patterns. Reponses with latencies longer than 3 seconds are excluded.

Table II: Repetition latency (in ms) for correctly repeated words broken down by lexical and actual stress position. Correct stress patterns lie on the main diagonal; off-diagonal cells represent stimuli with incorrect stress patterns. Responses with latencies exceeding 3 seconds are excluded.

Figure 1: Per cent correctly completed (recognised) words as a function of the number of diphones made audible from the word onset. Stimuli with correct stress position are indicated with open symbols, words with incorrect stress by filled symbols. Panel A presents the data for words with lexically initial stress, panel B for lexically final stresses.

Figure 2: Frequency distribution of perceived stress patterns as apparent from the error responses in a gating task after hearing the initial syllable of a word, broken down by lexical and actual stress position ("1": perceived initial stress, "2": perceived non-initial stress, "?": response ambiguous).

Appendix I: stimulus words used in experiment I

Lexical stress initial		Lexical stress final	
toeval	'coincidence'	piloot	'pilot'
virus	'virus'	hotel	'hotel'
middag	'noon'	loket	'ticketwindow'
bizon	'bison'	rumoer	'din, noise'
paling	'eel'	moraal	'moral'
datum	'date'	tomaat	'tomato'
divan	'couch'	seizoen	'season'
sieraad	'piece of jewelry'	totaal	'total'
humor	'humour'	konijn	'rabbit'
motor	'engine'	minuut	'minute'

Appendix II: Stimulus words used in experiment II

initial stress		medial stress		final stress	
cavia	'guinea pig'	propeller	'propeller'	boulevard	'boulevard'
carnaval	'carnaval'	pantoffel	'slipper'	kapitein	'captain'
paprika	'green pepper'	kabouter	'gnome'	peloton	'platoon'
tombola	'tombola'	kastanje	'chestnut'	testament	'will'
kolibri	'humming-bird'	benzine	'petrol'	tolerant	'tolerant'
dominee	'parson'	kanarie	'budgy'	canape	'couch'
kandelaar	'candle-stick'	tentamen	'test'	kapitool	'capitol'
piccolo	'bell-boy'	piano	'piano'	terpertijn	'turpentine'

		ACTUAL STRESS ON		
		1 ST SYLL.	2 ND SYLL.	3 RD SYLL.
LEXICAL STRESS ON	1 ST SYLL.	66 %	44 %	56 %
	2 ND SYLL.	34 %	81 %	31 %
	3 RD SYLL.	34 %	25 %	63 %

TABLE I

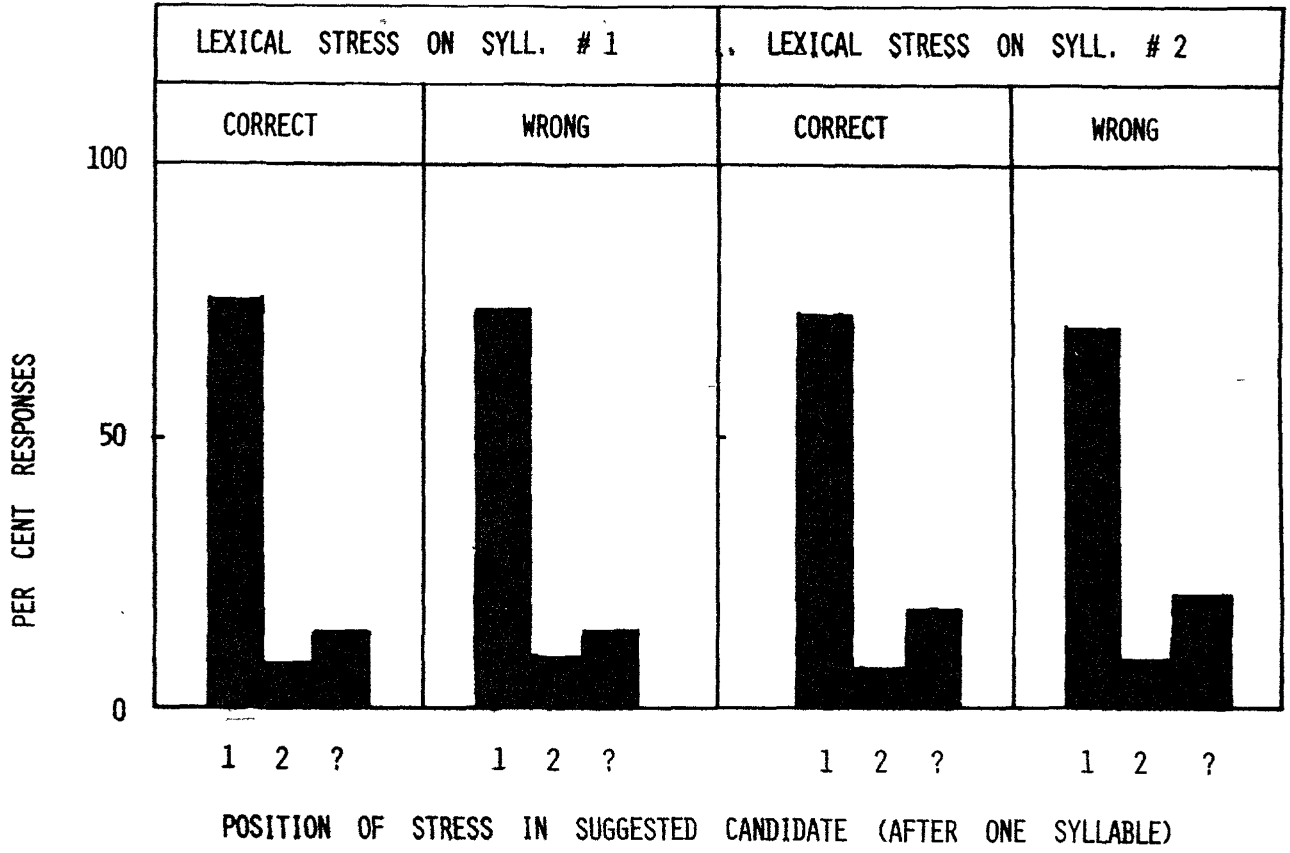
LEXICAL STRESS ON SYLL. # 1
LEXICAL STRESS ON SYLL. # 2
CORRECT
WRONG
CORRECT
WRONG
100
50
0
PER CENT RESPONSES
1
2
?
1
2
?
1
2
?
1
2
?
POSITION OF STRESS IN SUGGESTED CANDIDATE (AFTER ONE SYLLABLE)

FIGURE II

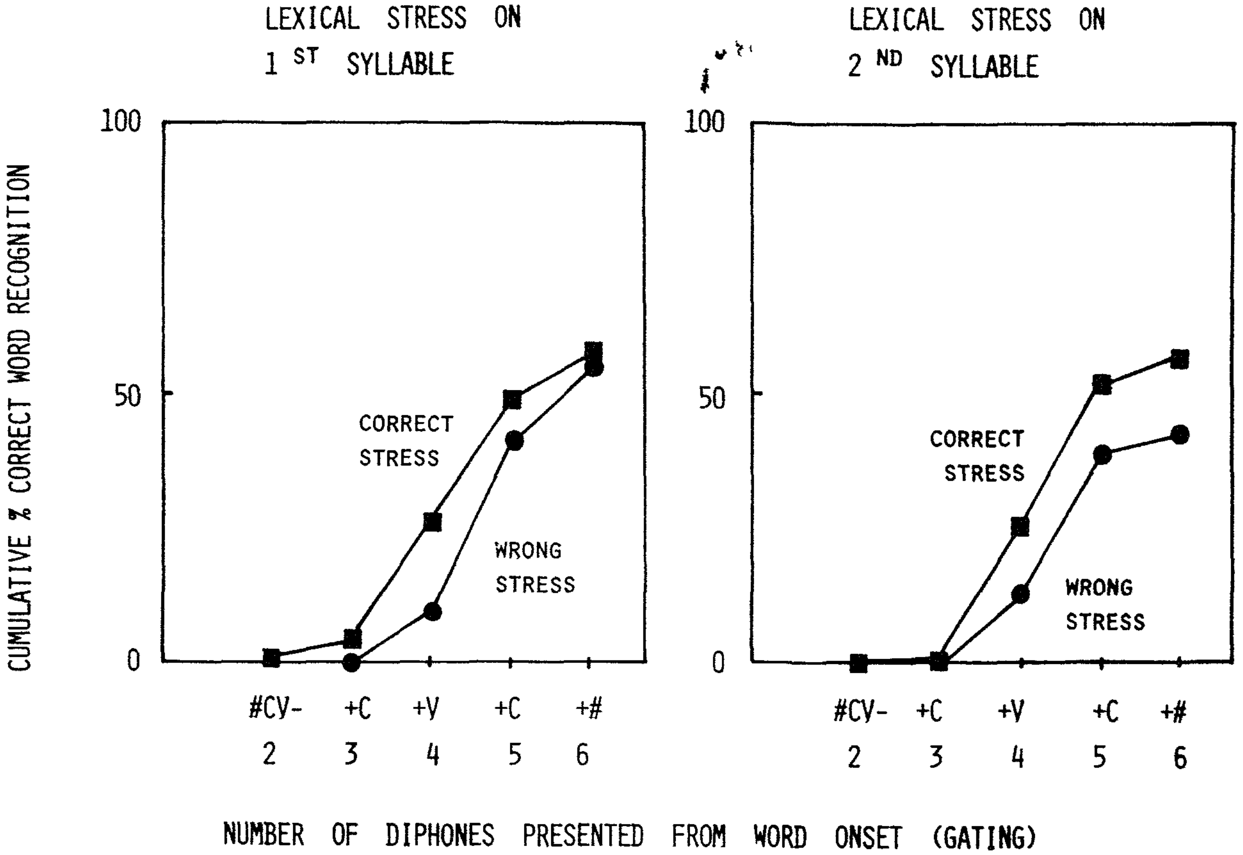
LEXICAL STRESS ON
1 ST SYLLABLE
LEXICAL STRESS ON
2 ND SYLLABLE
CUMULATIVE % CORRECT WORD RECOGNITION
100
50
0
100
50
0
CORRECT STRESS
WRONG STRESS
CORRECT STRESS
WRONG STRESS
#CV- +C +V +C +#
2 3 4 5 6
#CV- +C +V +C +#
2 3 4 5 6
NUMBER OF DIPHONES PRESENTED FROM WORD ONSET (GATING)

FIGURE I