ISCA Archive
http://www.isca-speech.org/archive

3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

PERCEPTION OF WORD BOUNDARIES BY DUTCH LISTENERS

Monique van Zon (1)
Beatrice de Gelder (1, 2)

Cognitive and neurosciences unit, Tilburg University, The Netherlands (1)
Laboratoire de psychologie experimentale, Université Libre de Bruxelles, Belgium (2)

10.21437/Eurospeech.1993-167

ABSTRACT

For understanding speech listeners spontaneously use a strategy that segments the speech stream into units. Recent work based on English suggests that speech segmentation is triggered by strong syllables. This was shown by a paradigm in which word boundary misperceptions were elicited in a task in which English native speakers listened to faintly audible fragments of sentences. In the present study we examined whether similar processing strategies can be observed for Dutch native speakers using the same paradigm. The results for Dutch are as predicted by the Metrical Segmentation Strategy.

Keywords: speech perception, stress, segmentation strategy

1. INTRODUCTION

Speech presents itself as a continuous stream and since - at least for the purposes that concern us here - individual words are the vehicles of meaning, the task of the listener is to spot words, in other words, to segment the stream into discrete perceptual units so that lexical representations can be accessed. Most straighforwardly, how are words spotted? Words derive their semantic identity from unique consonant/vowel patterns and consist of unique combinations of otherwise combinable parts. From the point of view of the listener paying attention to the primitives of these two interlocked combinatorial routines is as logical a route for watching words as spotting feathers is for watching birds. Properties of words and strategies of word spotting are two sides of the same coin. However, a quick glance at linguistic theories of consonant/vowel patterns and of word constituents brings out immediately that there is much room for phonological differences between languages. Consequently, there is room for the notion that listeners tend to spot words in a way that is partly at least language specific.

As far as phonetic and phonological properties are concerned a major heuristic for initial classification between languages is the contrast between 'stress-timed' and 'syllable-timed' languages (e.g., Dauer, 1983; Bertinetto, 1989). As Dauer (1983) points out, the terms 'stress' versus 'syllable-timed' are related to phonological properties. As the story goes, English is a stress-based language and to that extent its phonological properties are different from those of so called syllable-based languages. Dutch seems to be a language which is more closely related to a stress-timed language like English than to a syllable-timed language like French. The differences between stress-timed and syllable-timed languages have to do with a variability in syllable structure, vowel reduction, and the phonetic realization of stress (Dauer, 1983). In stress-timed languages more variation in syllable structure is allowed, e.g., in English vowels can be followed by 4 consonants, while in a syllable-timed language like Spanish, a vowel can be followed by not more than 2 (medial) consonants. Another property that distinguishes stress-timed languages from syllable-timed languages lies in the centralization of unstressed vowels. Users of stress-timed languages strive for maximalizing differences between stressed and unstressed syllables. This can be done by reduction to schwa, or - in specific languages - by reduction to other vowel variations. In English for instance the weak form of "him" (/hIm/) is reduced to (/Im/). Moreover, in English an unstressed syllable preserves its syllabicity. In contrast, in syllable-timed languages reduction in unstressed position mostly does not lead to other reduced vowel variants than schwa, or results to deletion of the relevant syllable (e.g., in French). As concerns the phonetic realization of stress most stress-timed languages have lexically free stress while most syllable-timed languages have no lexical stress at all (e.g., French). Dutch resembles English as far as the above mentioned phonological properties are concerned. However, a difference between English and Dutch is that in English vowel reduction is obligatory in unstressed syllables, while in Dutch it is optional (Trommelen & Zonneveld, 1989). Shortly, in Dutch unstressed syllables may be realized reduced or unreduced.

Starting from the importance of the syllable in linguistic theories Mehler, Dommergues, Frauenfelder and Seguī (1981) used a syllable monitoring task and found evidence for a syllable-based listening strategy in native speakers of French. The central issue was the critical importance of the syllable boundary (e.g., BA in BAlance or BAL in BALcon was responded to faster than BA in BAlcon or BAL in BALance). Mehler et al. concluded that native French speakers apply a syllable-based speech segmentation routine and on this basis suggested that syllables are likely candidates for the primitives of speech processing and for lexical access. One critical issue was the interaction between the properties of the language and the strategies shown by its speakers. In a follow up experiment carried out by Cutler, Mehler, Norris, Seguï (1986) with English subjects and English stimuli no such syllable effect was found. Explaining this result by appealing to the contrasting phonological properties of French and English, Cutler et al. (1986) argued that for a stress-based language like English a syllable-based strategy would be less efficient. As a matter of fact, a stress-based segmentation strategy appears to suggest itself when one considers the statistical properties of the English lexicon which has more syllables with (primary or secondary) stress on the first syllable (Cutler & Carter, 1987).

Research on listener's strategies in English has since yieled evidence for a segmentation strategy based on metrical weight more than on syllable structure. In a word-spotting task (Cutler and Norris, 1988) listeners were asked to detect real words embedded in nonsense bisyllables. CVCC-words were converted to nonsense words by adding a final VC sequence. The vowel in this final sequence could be either strong or weak (e.g., 'mint' in 'mintayf' or in 'mintef'). Results showed that detection of the embedded word was slower in a bisyllable with two strong syllables than in a bisyllable with a strong and a weak syllable (in other words detecting 'mint' in 'mintayf' took longer than in 'mintef'). Cutler & Norris (1988) take this result as evidence for the notion that in the beginning of the second strong syllable a new word spotting process is initiated and leads to an increase in reaction time.

Further evidence for the role of word initial strong syllables was obtained with a paradigm designed to elicit misperceptions of word boundaries (Cutler and Butterfield, 1992). Given the lexical statistics of English just mentioned, a plausible pattern of auditory misperceptions or slips of the ear would be one where strong syllables not occuring at the beginning of a word are mistakenly taken as signalling word beginnings. This was indeed what Cutler and Butterfield observed.

What are the similarities between English and Dutch and what are the reasons to believe that in Dutch listeners' strategies turn around the inital strong syllable? Samples from naturally occurring word boundary misperceptions are suggestive for the similarities between the two languages. For example, in Dutch the utterance "...iemand vergist zich..." ([somebody makes a mistake...]) was perceived as "...iemand heeft vis gegeten..." ([somebody ate fish]). In this slip of the ear a word boundary insertion occurred before a strong syllable: /ver-gist/ was replaced by /heeft vis/. Wordcounts performed on Dutch and English CELEX lexicons show major similarities between Dutch and English (see Vroomen and de Gelder, 1993, for details). Thus, evidence from lexical statistics of Dutch is suggestive for the privileged role of strong syllables in word spotting strategies. In Dutch vowel reduction of short vowels is not as compulsory as it is in English but it does occur in some specific cases and in casual or fast speech (Trommelen and Zonneveld, 1989). Thus the criterion for weak syllables in Dutch would be more stringent as it would only include syllables with a schwa. Using this strict criterium for Dutch a very low percentage of words having a weak word-initial syllable was observed, i.e., 12.4 percent which is even lower than in English (18.1 percent). Moreover, strong syllable initial words have a much higher frequency.

So far we have outlined the situation for English speakers given the facts about the English and Dutch and concluded that the same hypothesis might be formulated concerning segmentation strategies in Dutch as were found to exist in English speakers. If so, language dependent differences would offer a window on the generality of the proposed theory of speakers segmentation. The observation of differences that are related to phonological differences in the language would possibly contribute to understanding which aspects of the Metrical Segmentation Strategy are language specific (different for English vs. Dutch) vs. the ones that are potentially universal (same for English and Dutch).

In the present study we used the word boundary paradigm used by Cutler & Butterfield (1992). For Dutch similar predictions are made as for English namely, more word boundary insertions are expected before strong syllables than expected by chance, and more word boundary deletions are predicted to occur before weak syllables than expected by chance. Additionally, for English it was found that word boundary insertions before strong syllables lead to lexical words, and boundary insertions before weak syllables produce grammatical words. The question is whether similar effects can be observed for Dutch.

METHOD

Subjects. The subjects consisted of a group of 16 university students who were all native speakers of Dutch. None of them reported hearing deficits. The subjects were paid for their participation.

Materials. 54 experimental sequences were constructed containing mono- and bisyllabic words and consisting of a total of six syllables with alternating stress patterns. Taken by themselves the sequences were semantically ambiguous. Furthermore, for the pretest, a Dutch newspaper article and a list of isolated words was used. This word list consisted of 36 spondees (i.e. bisyllabic words having full vowels in both syllables). Half of the list had primary stress on the first syllable (e.g., "kaasboer"), the other half on the second (e.g., "motief").

All materials were recorded on two channels of a Philips

850 Digital Audio Tape-recorder. The utterances were digitized with 12 bits at 20 kHz. Maximum peak levels were approximately equal on the VU meter of the DAT-recorder. This was decided on basis of visual (waveform display) and auditive criteria. The sequences were spoken by a male native speaker.

An attenuator was calibrated on the basis of a 1 kHz signal. This was done by plugging the attenuator both in a 10 steps potentio-meter and in a Fluke 8922A dB-meter. The dB-meter was connected with MDR CD450 Sony headphones. In this way tension (in dBm) could be measured directly from the headphones. Measurement of small steps on the attenuator (1 step equals about .25 dBm in our scaling) was done.

Procedure. All materials were played in a sound-proof booth over Sony MDR CD450 headphones which were plugged in an attenuator. This attenuator was connected to the DAT-recorder. The experimental sequences were presented at the level of an individually estimated speech reception threshold. Previously to each sequence and in order to focus attention, the number of the trial to be presented was played some dB louder than the sequence itself. Subjects were told they were going to listen to speech presented "as if a radio was on a low volume". They were instructed to write down what they thought had been said and to approximate as well as they could what they had heard. They were told to write a dash where they thought there had been a syllable they were unable to report. They also were instructed to mark whether this dash (i.e. syllable) belonged to another word or stood on its own. Testing lasted about 50 minutes. Before testing individual speech reception threshold was estimated. Herefore we firstly asked the subject to listen to a Dutch newspaper article. The starting level of presentation of the newspaper article was equal for each subject and was set at a comfortable volume. The subjects were instructed to adjust the volume during listening to the lowest level at which they could still understand the speaker. Subjects were questioned afterwards in order to allow the experimenter to assess whether they had understood the text at the level they had chosen. This level was taken as the starting level for the isolated words list. The subject was asked to repeat each word. During this, the experimenter reduced the volume by 3 steps for each three items until one or more of three successive words was repeated incorrectly. The volume was increased by one step until an item was repeated correctly. The level on which the subject responded correctly to 50% of the words was regarded as the estimated speech reception threshold.

RESULTS

A total of 864 responses (16 x 54 input sequences) was given but only the sequences that fullfilled the following criteria were analyzed: (a) the number of syllables of a response sequence had to be the same as in the input (i.e. six); (b) response sequences had to have the same rhythmic pattern as in the input as far as concerns reduction, e.g., only those syllables were analyzed in which reduced syllables were replaced by reduced syllables and in which unreduced syllables were replaced by unreduced syllables. Thus correct responses, misses, responses with less than or more than six syllables and responses with deviant rhythmic patterns were not analyzed. Also errors on the first syllable of a sequence were left out. The total number of errors that remained for analyzing was 270 (for examples of misperceptions see Appendix). These errors were not restricted to specified sequences: in almost all sequences word boundary misplacements occurred.

We were interested in comparing the expected boundary misplacements with the observed misplacements. Therefore we first had to calculate (weighted) expected boundary misplacements based on the sequences in the input. Expected frequencies for insertions before a strong syllable could be calculated because insertions can only occur when the input sequence (in which an error was made) contained a strong syllable in a non-word-initial position and so on. We calculated weighted expected frequencies for each cell by multiplying the total number of the observed frequencies over 4 respectively 2 cells by the expected frequency of a cell (e.g., the category 'insertion before strong') divided by the total number of expected frequencies over 4 respectively 2 cells. The expected frequencies used were counted on basis of the input.

Table 1 shows the distribution of the observed and the (weighted) expected boundary misplacements across insertions versus deletions before strong (i.e. unreduced) versus weak (i.e. reduced) syllables. The predicted interaction is significant (with correction for continuity, chi-square (1 df) = 48.67, $p < .001$). In binominal tests syllable strength difference is significant both for insertions $z = 1.92$, $p < .03$) and for deletions ($z = 5.43$, $p < .001$). In other words, more word boundary insertions did occur before strong syllables than before weak syllables, and more deletion errors did occur before weak than before strong syllables.

Word boundary insertions were further analyzed in word class differences of the word following the insertion. Insertions followed by nonsense responses or dashes (in total 58 of the 126 insertions) were excluded from the analysis.

Table 1
Word Boundary Insertions and Deletions before Strong versus Weak Syllables
(weighted expected frequencies in parentheses)

	Before Strong	Before Weak
Insertions	94 (82)	32 (36)
Deletions	60 (105)	84 (47)

Table 2
Occurence of Lexical versus Grammatical Words following Insertions Before Strong versus Weak Syllables
(weighted expected frequencies in parentheses)

	Before Strong	Before Weak
Lexical	47 (47)	6 (20)
Grammatical	4 (0)	11 (1)

Table 2 shows the distribution of the observed and the (weighted) expected insertions before strong (i.e. unreduced) versus weak (i.e. reduced) syllables. As the table shows, insertions before strong as well as before weak syllables were followed both by lexical words and by grammatical words (or by nonsense words or dashes). Although an interaction can not be calculated here (because of a zero in the expected values), a binominal test on insertions before strong syllables shows that insertions before strong syllables are more often followed by a lexical word than by a grammatical word ($z = 5.88$, $p < .001$).

DISCUSSION

Using a paradigm developed by Cutler & Butterfield (1992) that was designed to elicit word boundary misperceptions we have shown that Dutch speakers listening to faintly audible fragments of sentences, tend to insert word boundaries before strong syllables and to delete word boundaries before weak syllables.

The present study took into account the similarities as well as the differences between English and Dutch and examined whether similar processing heuristics can be observed in the processing strategies of Dutch speakers. The results show- that boundary misperceptions in Dutch do occur as predicted by the Metrical Segmentation Strategy.
(a) As is the case in English for word boundary insertions more erroneously inserted errors are made before strong than before weak syllables; (b) for word boundary deletions we find that more word boundary deletions are made before weak than before strong syllables; (c) a word class effect observed in English is also found for Dutch namely strong syllables are more often followed by a lexical word than by a grammatical word.
Taken together these findings lend support to the conception that an interaction between strategies of segmenting the continuous stream of speech and processing strategies of its listeners exists.

REFERENCES

/1/: Bertinetto, P. M. (1989). *Reflections on the Dichotomy "Stress" versus "Syllable-timing"*. R.P.A., 91-92-92, 99-130.

/2/: Cutler, A. & Butterfield, S. (1992). Rhythmic Cues to Speech Segmentation: Evidence from Juncture Misperception. *Journal of Memory and Language, 31,* 218-236.

/3/: Cutler, A., & Carter, D.M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2,* 133-142.

/4/: Cutler, A., Mehler, J., Norris, D., & Seguï, J. (1986). The syllable's differing role of segmentation of French and English. *Journal of Memory and Language, 25,* 385-400.

/5/: Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113-121.

/6/: Dauer, R.M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11,* 51-62.

/7/: Mehler, J., Dommergues, J., Frauenfelder, U., & Seguï, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20,* 298-305.

/8/: Trommelen, M., & Zonneveld, W. (1983). *Klemtoon en Metrische Fonologie,* Muiderberg: Coutinho.

/9/: Vroomen, J., & De Gelder, B. (1993). *The role of strong syllables for spoken word recognition in Dutch,* European Society for Cognitive Psychology. Elsinore, Denmark, 11-15 September.

APPENDIX

Examples of Elicited Word Boundary Misperceptions (in Bold Cursive).

	Before Strong	Before Weak
Insertions	..vroeger bracht **gezang** ons.. (..formerly singing made us..) -> *vroeger bracht **de zang** ons*	..de zieke **eerder** kramp.. (..the sick sooner cramp..) -> *de zieke **eer bekranst***
Deletions	..uw leeftijd kreupe**l loopt**.. ..your age walks with a limp.. -> *in leeftijd **kreukeloos***	..je eerder **zelf** beweerd.. (..you earlier claimed by yourself..) -> *die eerder **zelfbeheer***