Journal of Phonetics (1992) **20,** 331–350

Durational cues for word segmentation in Dutch

Hugo Quené

Research Institute of Language and Speech, Rijksuniversiteit Utrecht, Trans 10, NL-3512 JK Utrecht, The Netherlands

Received 26th December 1990, and in revised form 5th February 1992

Word segmentation is a prerequisite for word recognition and subsequent comprehension of a connected speech utterance. Linguistic and contextual information provides insufficient means to this end. In this paper, the contribution of acoustic-phonetic word boundary markers to perceived word segmentation is investigated. In the first experiment, phonetically ambiguous word combinations were presented; linguistic and contextual information could not contribute to word segmentation of these stimuli. Yet, the observed 80% accuracy shows that phonetic boundary cues are used effectively. *Post-hoc* acoustic analysis suggests that three durational parameters may affect perceived segmentation, viz. pre-boundary vowel decay time, intervocalic consonant duration, and post-boundary vowel rise time. In the second experiment, the latter two of these parameters were manipulated. Results show a significant effect of consonant duration on perceived segmentation, but no effect of vowel rise time. Uncontrolled boundary markers (presumably, boundary segments and allophonic variants) also provided positive boundary cues for /C # V/ context. In conclusion, some word boundary markers contribute to the perceived word segmentation. Nevertheless, results also suggest that non-sensory (linguistic and contextual) information is more important for retrieving the intended word segmentation.

1. Introduction

A word in connected speech often differs considerably from its canonical (isolated) sound form (see e.g. Shockey, 1974; Klatt, 1980). This is yet another example of the well-known problem of *phonetic variance*: the sound form of a word is subject to variations, which are partly due to the sound forms of adjacent words, as in (1) below.

(1a) zat je [zɑt # jə] ("sat you")→[zɑcə]
(1b) would you hit it to Tom? →[wʊdʒəhɪrɪtətʰɑm]
(Klatt, 1980, p. 249)

The effect of these changes is that the isolated word forms cannot be deduced unambiguously from the speech utterance as a whole. In other words, the utterance becomes "phonetically ambiguous" (Spencer & Wollman, 1980): it may correspond to a number of messages at the linguistic level, only one of which is the intended one. Words excerpted from conversational speech are more difficult to identify, due

0095-4470/92/030331 + 20 $03.00/0

to this degradation of the isolated sound form (Pickett & Pollack, 1963). Additional evidence for this ambiguous character of speech utterances with respect to its constituting words is provided by misperceptions (2):

(2a) utterance: there's some ice tea made
perceived: there's a nice team mate
(Garnes & Bond, 1980, p. 235)
(2b) utterance: ik voel'm aan [. .vul # əm # an] "I feel him (on)"
perceived: volle maan [vɔlə # man] "full moon"
(Utrecht corpus of speech errors)

In spite of this phonetic ambiguity, correct understanding of an utterance requires that discrete words are identified, i.e., correct word segmentation of the speech signal. For example, in (2a) above, the listener failed to interpret the [z] sound as being the result of voicing assimilation and subsequent degemination between **there/zs/ome**, thus giving rise to the alternative interpretation **there/z/a**. Thus, incorrect word demarcation drives the listener to an incorrect understanding of (the remainder of) the speech message. In normal speech communication, however, listeners usually arrive at a correct understanding of the speech message. By implication, the utterance is correctly divided into its constituting words. Little is known, however, about the perceptual strategies employed in order to retrieve the intended word segmentation from a connected speech utterance.

This paper investigates the perceptual division of a continuous speech signal into discrete words. This *word segmentation* is a task of utmost importance in speech comprehension, since words are the smallest independent elements conveying the meaning of an utterance. Probably, word segmentation has a higher perceptual priority than the demarcation of linguistic elements of lower (e.g., syllable, phoneme) and higher levels (e.g., clause, phrase). For reasons to be explained below, the primary focus of this paper is on acoustic-phonetic aspects of word segmentation. Specifically, it will be investigated whether certain acoustic-phonetic markers of word boundaries are present within the speech signal (Section 2) and whether these cues are perceptually relevant for word segmentation (Section 3).

However, there is considerable evidence that word segmentation is not achieved solely on the basis of acoustic-phonetic information. It may also result from successful recognition of the previous word, which can be achieved before the end of that word (Marslen-Wilson & Welsh, 1978, p. 57). If this happens, the listener can anticipate the following boundary and word onset. Example (3a) illustrates this strategy. The word **elephant** can be identified after hearing **eleph-**; the listener can then anticipate the remainder of the word, the word boundary, and the subsequent word onset (printed in capital):

(3a) an elephant # Is a big animal
(3b) an elephant # tEnds to be big
(3c) a bat # is a small animal

However, Frauenfelder (1985) has argued that this anticipation strategy fails first if the speech utterance contains an ambiguous word boundary, such as example (3b); second after short words, which can only be recognized after their acoustic offset, such as example (3c); and third if a word is embedded in a longer word (e.g. Dutch

kan/s "can/chance", **kar/dinaal** "cart/cardinal", **bank/et** "bank/banquet", **ree/gen** "roe/rain"; English **cap/tain, car/d/iologist**).

In addition, word segmentation may result from *phonotactic restrictions*. If an illegitimate consonant combination is detected, then the listener must assume a word boundary somewhere within this combination. For example, the combinations [rmdr] in Dutch or [mgl] in English yield legal offset and onset clusters only in segmentation as [rm # dr] and [m # gl] respectively. However, such constraints are usually not very restrictive. On the basis of analyses of transcribed corpora of English texts, Lamel (1984, p. 66) reported that 20% of all distinct consonant combinations containing a word boundary may also occur word-internally; this figure increases to 2/3 if frequency occurrences of tokens are taken into account. These analyses suggest that phonotactic constraints are ineffective for word segmentation in two out of three intervocalic consonant combinations.

The contribution of *stress* to word recognition and segmentation is clearest in languages where stress is fixed on the first syllable of a word (e.g. Finnish). Upon hearing a stressed syllable, a listener may safely assume a word boundary preceding this syllable, and hence treat the stressed syllable as the onset of a new word to be recognized. Taft (1984), Grosjean & Gee (1987) and Cutler (1989) argue that this strategy is also employed in languages with varying stress patterns (lexical stress; e.g. English and Dutch). Presumably, a lexical access is attempted at every stressed syllable. The efficiency of this strategy depends greatly on the distribution of stress patterns across word types and tokens. This word segmentation strategy is bound to fail with words having a non-initial stressed syllable (e.g. English **patrol, succeed,** Dutch **muziek** "music", **agenda** "agenda"). In Dutch texts, such words have a pooled token frequency of 10%, according to the CELEX lexical database (Quené, 1992).

Since word boundaries cannot always be derived unambiguously from such higher-order phonological information, we assume that more direct *acoustic-phonetic markers* also contribute to word segmentation. In other words, we assume that word segmentation is (at least partly) influenced by acoustic-phonetic aspects of the speech signal. These word boundary phenomena provide cues which are relevant for retrieving the intended word segmentation of the connected speech signal. In this respect, three types of acoustic-phonetic word boundary cues may be discriminated:

(1) Boundary segments, with none or irregular vocal cord vibration (glottal stop, laryngealization, vocal fry). These result in a silent interval, or in speech fragments with reduced formant structure (Lehiste, 1960, 1965; Gårding, 1967; Nakatani & Dukes, 1977). These boundary segments can only correspond to a (word or syllable) boundary, but not to any lexical phoneme.

(2) Spectral phenomena, where the location of the word boundary may influence allophonic variation (Trubetzkoy, 1939) and the degree of coarticulation and assimilation between speech fragments (Daniloff & Moll, 1968; Moll & Daniloff, 1971; Su, Daniloff & Hammarberg, 1975; Slis, 1985; Loots, 1983).

(3) Durational phenomena, where the duration of a speech segment varies with the position of the segment with respect to the word boundaries (e.g. Hoard, 1966; Nooteboom, 1972; Carlson, Granström, Lindblom & Rapp, 1972; Lindblom & Rapp, 1973; Oller, 1973; Umeda & Coker, 1975; Klatt, 1975, 1976).

These three types of word boundary markers may be ordered from "strong" to "weak". Boundary segments provide the strongest cues, since they cannot correspond to lexical speech segments; they can only be interpreted as phonetic correlates of linguistic boundaries. Durational boundary phenomena provide the weakest segmentation cues, since the primary function of the relevant carrier segment is to convey lexical information. Hence, it may be argued that "phonetic ambiguities" [such as (2) above] in which boundary segments and spectral boundary cues are present, are not ambiguous at all at the phonetic level. A native listener will never misinterpret such a fragment (even when heard without context), if the fragment contains boundary segments, distinct allophones and/or audible (absence of) coarticulation. This claim is supported by the perceptual relevance of these word boundary markers, as reported repeatedly over the last decades (Lehiste, 1960; Gårding, 1967; Nakatani & Dukes, 1977; Nakatani & Schaffer, 1978; Harris, Umeda & Bourne, 1981; Barry, 1981, 1984; Cohen, 1987).

In the research reported here, we will therefore concentrate on the influence of the weakest boundary cues on the perceived word segmentation, viz. *durational* word boundary markers. For each word combination, several hybrid versions are created; these versions are identical but for the manipulated durational boundary markers. If durational markers are indeed found to be perceptually relevant, then we have corroborated the available evidence that sensory information contributes to perceived word segmentation. It is predicted that listeners can locate word boundaries in connected speech on the basis of durational cues alone, even if no obvious correlates of word boundaries (viz. boundary segments) are present in the speech signal.

To complicate matters, however, there is evidence that word segmentation is a language-specific task. The relative importance of sensory and higher-order information (Lehiste, 1965), as well as the relative (acoustic and perceptual) importance of various phonetic boundary cues (Barry, 1984) varies between languages. For example, if the glottal stop or vowel laryngealization has phonemic status, then these phenomena are less likely to be used as boundary cues. Since we preferred to use Dutch as the target language, we had to be establish first whether the durational cues of interest do indeed occur as word boundary phenomena in Dutch. The first experiment (described in Section 2) aims at identifying which durational phenomena vary systematically with the intended word segmentation.

2. Experiment 1

2.1. *Introduction*

The aim of this experiment is (a) to investigate whether and to what extent listeners can correctly locate word boundaries, on the basis of exclusively acoustic-phonetic cues, and (b) to investigate which durational boundary markers may have been used to this end. Consequently, no higher-order cues to the intended segmentation should be available in the stimulus material. As in previous research, two-word combinations were used as stimuli, with an ambiguous position of the intermediate word boundary (e.g./no(#)n(#)oʃən/ yielding **known#ocean** or **no#notion**; Nakatani & Dukes, 1977; Cohen, 1987). Since our primary interest is in the segmentation of connected speech, these word combinations are to be embedded in a longer (meaningful) speech utterance.

The first question will be answered by evaluating listeners' perceived word segmentation of these stimuli. If their accuracy deviates from chance, then acoustic-phonetic markers probably contribute to retrieving the intended word segmentation. The second question will be answered by measuring durations of speech portions, and identifying portions with systematically varying durations between the two contrastive (intended) segmentations. Subsequent comparison with the obtained responses may indicate which variations affect the segmentation process.

2.2. *Method*

2.2.1. *Stimulus word combinations*

In order to construct two-word utterances with ambiguous word boundary position, an inventory was made of all Dutch monosyllabic words with an ambiguous boundary, by means of tables describing all Dutch monosyllabic words (Bakker, 1971), as well as by means of rules describing the phonemic distribution of Dutch (Trommelen, 1983). Included words yield an existing word if the final or initial consonant(s) is (are) removed as a consequence of an alternative segmentation. Only monomorphematic words were selected. The numbers of words in this inventory, broken down by number of ambiguous consonants and by initial *vs.* final boundary ambiguity, are given in Table I.

Stimuli were constructed by combining a word with ambiguous offset and one with ambiguous onset, thus producing a minimal pair of (potentially meaningful) two-word utterances (e.g.,/dip # ɪn/"deep in" and/di # pɪn/"that pin"). For stimuli with a *single* intervocalic consonant, two minimal pairs of utterances were constructed for each of the pivotal consonants/p, t, k, f, s, x, m, n, l, r, j, ʋ/ which may occur both word-finally and word-initially. However, glides /j, ʋ/ each contributed to one minimal pair of utterances, since no more words with boundary ambiguity could be found. Since only two possible segmentations are allowed (excluding segmentations involving geminates), stimuli with *two* intervocalic consonants were divided in two categories: (1) with ambiguous offset $/..V_1C(\#)C(\#)V_2../$, and (2) with ambiguous onset $/..V_1(\#)C(\#)CV_2../$. For each of

TABLE I. Numbers of words in an inventory of Dutch monosyllabic, monomorphematic words with ambiguous boundary, broken down by *length* of the ambiguous fragment (number of ambiguous consonants) and by *position* of the ambiguous fragment. Marginal totals may be less than the sums of cells, since words may contribute to more than one cell

	Position		
Length	Final	Initial	Total
1	248	402	581
2	161	283	441
3	11	12	23
Total	420	697	990

these categories, 15 minimal pairs of utterances were constructed. The resulting (22 + 15 + 15 =) 52 minimal pairs are listed in the Appendix.

Each member of a minimal pair was embedded in a meaningful sentence, which allowed only one of the two possible segmentations of the stimulus utterance. In other aspects, the two sentences for the two contrastive members of a minimal pair had to be as similar as possible with respect to number of syllables and words, fragment position in sentence, etc. Care was taken that no important syntactic or prosodic break occurred within or immediately preceding or following the stimulus fragment. In order to obtain meaningful sentences, however, it was sometimes necessary to extend the second of the two words making up a stimulus utterance with an inflexional morpheme [ə]; if necessary, this was done in both contrastive members of that pair. Example (4) shows a sample stimulus utterance

(4) consonant: /p/
stimulus utterance: /di(#)p(#)ɪn/
/p#/ sentence: stop deze zaden niet te **diep in** de grond
"put these seeds not too deep in the ground"
/#p/ sentence: deze bom ontploft als je **die pin** eruit trekt
"this bomb explodes if you that pin out pull"

Subsequently, these (2 × 52 =) 104 sentences were randomized in a list, which was preceded and followed by 10 filler sentences.

2.2.2. *Speakers and procedure*

The sentence list was read by two male and two female native speakers of Dutch. All were staff members of the Utrecht Institute of Phonetics, participating voluntarily. They were unaware of the aim of the experiment, and were not informed about the occurrence of contrastive two-word stimulus utterances within the sentences.

Speakers were instructed to read the list of sentences at a subjectively fast speech rate, in order to avoid breaks or pauses within the sentences. In addition, they had to accentuate those syllables which were underlined. This was done in order to obtain identical prosodic patterns in the two contrastive members of a minimal pair of two-word utterances.

Recordings were made in a sound-proofed booth, using high-quality recording equipment. The speakers had to repeat the list of sentences until the experimenter judged that each sentence was spoken at least once in accordance with the instructions. This took about two trials for each speaker.

Afterwards, speakers were asked to mention any striking aspects of the sentences. None of them had noticed the stimulus utterances with contrastive segmentations. Even after this contrast had been pointed out to them, they were all unable to point out any stimulus fragment in the sentence list.

2.3. *Segmentation responses*

2.3.1. *Stimulus material*

For each member of a minimal pair of two-word utterances, the realization with the least evidence for a boundary segment was selected. The presence of such boundary segments (pausing, glottal stop, laryngealization, vocal fry) was established by means of a speech editing program with visual (waveform) and auditory feedback. The selected realization was digitized (at 10 kHz sampling frequency, 4.5 kHz low-pass filtering, 12 bits) and excised from the carrier sentence; cuts were made at zero crossings; no windows were applied.

A simple design was chosen, where each of the four speakers accounted for only $104/4 = 26$ items of the total stimulus set (see Appendix). Since we were not interested in any possible speaker effects, a latin square design with balancing between speakers was considered unnecessary.

Stimulus utterances were converted back to an analog signal and recorded on audio tape in pseudo-random order, with an inter-stimulus-interval of 4 s. The stimuli were preceded and followed by 10 filler items. Total time of the test tape was approximately 10 min.

2.3.2. *Subjects and procedure*

Nineteen (native Dutch) subjects participated voluntarily; all but one were students in Phonetics or Linguistics courses. They were unaware of the aim of the experiment.

Subjects individually listened to the test tape over (binaural) headphones in a sound-proofed booth. They were instructed to choose between two possible contrastive segmentations, which were transcribed (in Dutch orthography) on their response sheets. Relative order of the intended and contrastive segmentation responses was randomized. It was emphasized that they should abstract from (sometimes rather different) orthographies of the two possible responses (e.g., **zeis om** *vs.* **zij som**).

Responses in accordance with the intended segmentation were scored as "correct", alternative responses as "wrong".

2.3.3. *Results and discussion*

For each stimulus item, the percentage of correct responses was calculated across 19 subjects. Average accuracy (percentage correct) was 80.4% across all 104 stimulus fragments. Detailed results are given in Table II.

These results show a clear response bias in the first two stimulus categories, towards responses involving a word-initial vowel /#V/. This bias is most likely due to the presence of a "positive" boundary marker in /C#V/ stimuli (e.g. laryngealized vowel onset). The absence of this "positive" boundary signal in the contrastive /#CV/ stimuli has provided a weaker perceptual cue. According to Gårding (1967), who observed a similar bias in Swedish, this may be due to the fact that the realized /#CV/ syllabification is also found for intended /C#V/ divisions in faster speech. Secondly, the fact that /C#V/ stimuli were not involved in the third stimulus category might partly explain the lower overall accuracy for this condition.

The observed accuracy was tested against the binomial distribution ($p = 0.5$; $N = 19$) by means of the Kolmogorov–Smirnov test (Siegel, 1956). The resulting test value was significant at $p < 0.001$, indicating that the observed accuracy differs significantly from chance. Hence, word segmentation can indeed be accomplished fairly accurately, even if only acoustic-phonetic information is available.

TABLE II. Number of stimuli, mean boundary detection accuracies, and number of tokens with an observed boundary segment for each contrastive version of the three stimulus categories used in Experiment 1

Stimulus category		*N*	Accuracy (% correct)	Number with boundary segments
Single cons.	$VC\#V$	22	89	14
	$V\#CV$	22	76	1
Mean			82	
Two cons. final	$VC_1C_2\#V$	15	93	9
	$VC_1\#C_2V$	15	77	1
Mean			85	
Two cons. initial	$VC_1\#C_2V$	15	67	2
	$V\#C_1C_2V$	15	80	1
Mean			73	
Grand mean/total		104	80	28

The excerpting of the stimulus fragments may have decreased subjects' response accuracy, since such excerpts are perceived less accurately than identical fragments spoken in isolation (Pickett & Pollack, 1963; Amerman & Parnell, 1981). However, it obviously did not prevent listeners from perceiving the intended boundary. Despite the fact that stimulus fragments were originally embedded in connected speech utterances, they apparently contained effective boundary markers. This is remarkable since speakers were not aware of any phonetic ambiguities, and therefore did not intentionally disambiguate the stimulus fragments in the original sentences.

The exact nature and identity of the relevant acoustic-phonetic word boundary markers can only be established, however, if perceptual results are combined with acoustic analyses of the stimulus material. Presumably, markers were most effective in those stimuli which were perceived with the highest accuracy. In addition, stimuli perceived with an accuracy below chance level presumably contain markers which indicate a segmentation *opposite* to the intended one. The necessary acoustic analyses of the stimulus material are described in the following section.

2.4 *Acoustic analyses*

2.4.1. *Boundary segments*

Although the present experiment primarily aims at describing durational boundary cues in Dutch, we were also interested in the presence of (perceptually "stronger") boundary segments in the stimulus material. To this end, each of the 2×52 stimuli used in the listening experiment was inspected visually and auditorily with regard to the presence of a boundary segment (pause, glottal stop, laryngealization, vocal fry). A silent portion preceding the noise burst of a plosive consonant was regarded as the closure portion of the plosive (rather than as a pause or glottal stop).

The resulting frequencies of occurrence (see Table II above) support the explanation for the observed response bias, as discussed above. Most of the /C#V/

stimuli contain a boundary segment, which may have served as a "positive" marker. By contrast, boundary segments are rare in the /C#C/ and /V#C/ stimuli.

These data show that even in /#V/ contexts, realization of a glottal stop is not obligatory in Dutch. Similar data were reported by Jongenburger & Van Heuven (1991), who observed that glottal stops were realized in 56% of the /#V/ contexts (by one professional speaker). In their experiment, glottal stops were realized equally often at ambiguous and unambiguous word boundaries (39% *vs.* 37%).

These results suggest that the presence of absence of a boundary segment yields insufficient information to account for subjects' accuracies in perceived word segmentation. Apparently, listeners use additional sensory information to retrieve the intended boundary position. Since our primary interest is in (perceptually less salient) sub-phonemic cues to word segmentation, durational word boundary phenomena are investigated in the following section.

2.4.2. *Durational phenomena*: *measurements*

Vowel and consonant durations are reportedly affected by the intended word boundary position (cf. Section 1). This durational variation is essentially due to the difference in intended syllable (and hence word) affiliation of the pivotal consonant. In order to establish whether such variations also occur in Dutch, the durations of these segments were included in our analyses.

It was assumed that this variation would be largest in the stimuli with a single intervocalic consonant, because in these cases there is least articulatory similarity between the two contrastive versions of a minimal pair. For example, in stimuli with an ambiguous two-consonant offset, the first consonant does not change in syllable affiliation between the two contrastive intended segmentations. Consequently, one would not expect variations in the degree of articulatory overlap between the nuclear vowel and the first consonant of the ambiguous offset; hence no durational differences are to be expected for the /V_1C/ portion of these /V_1C(#)C(#)V_2/ stimuli. For stimuli with an ambiguous two-consonant onset, the same applies to the second consonant and the following nuclear vowel.

For this reason, only stimulus fragments with a *single* boundary consonant were selected for subsequent acoustic analysis. Of these, the two word combinations involving /j, ʋ/ were discarded, because the word-final and -initial variants of these phonemes could not be considered as phonetically related allophones.

Variation in the degree of articulatory overlap between the two nuclear vowels and the intermediate pivotal consonant probably corresponds to durational variations. Presumably, the greater overlap in /V_1C#V_2/ as compared to /V_1#CV_2/ would result in a shorter V_1 offset duration in the former case (as well as in a higher rate of formant transition between V_1 and C). For V_2 onset duration, the reverse argument applies. In order to achieve a more detailed analysis of these crucial transitions, the durations of the V_1 *offset* and V_2 *onset* portions were also included in the acoustic analyses.

In order to neutralize possible biases in the balancing of stimuli across speakers, realizations of all four speakers, of both members of the 20 minimal pairs of utterances, were included in the acoustic analysis. As a result, the material consists of 4 × 2 × 20 word combinations. For comparisons between durational parameters and perceptual accuracies, however, only the subset of stimulus fragments used in the listening experiment is available (2 × 20 combinations).

Durations of the relevant speech portions were measured, using a computer program with auditory and visual (waveform) feedback. Segment boundaries were determined according to criteria derived from Klatt (1976). In addition, vowel decay times were defined as the interval durations from the pitch period with maximal amplitude to the vowel end; for vowel rise times the reverse applies. In this connection, it should be noted that in Dutch, there is no aspiration of word-initial plosives. The observed segmental durations were each fed into a separate analysis of variance, with (intended) Boundary Position and Speaker as main effects (fixed and random, respectively).

2.4.3. *Durational phenomena*: *results and discussion*

The observed *durations* of V_1, the nuclear vowel of the first word (preceding the ambiguous word boundary position), show no significant difference between contrastive boundary positions, nor between speakers. The interaction effect was also insignificant. The absence of a boundary position effect might, however, be an artefact of the fact that nuclear vowels in the two words of a stimulus combination could not be balanced with respect to vowel identity (in order to arrive at meaningful combinations). Consequently, vowel durations are averaged over different vowel phonemes with different intrinsic durations (Peterson & Lehiste, 1960, p. 701; Lehiste, 1970, p. 18; Klatt, 1975). This leads to a high error variance (71%). In other words, intrinsic differences in V_1 duration are too large to observe any experimental effect.

The observed *decay times* of V_1 showed a significant difference between /V_1C#/ and /V_1#C/ [Table III (a); $F(1, 33) = 5.61$; $p < 0.05$]. Speaker and interaction effects were insignificant. Again, the large error variance (77%) may be an artefact of the intrinsic durational differences between vowels. Both phonologically short and long vowels were used in the stimuli; phonological length is known to affect vowel decay time (Cohen, Schouten & ’t Hart, 1962; Cohen, 1962). The direction of this difference is in line with the predicted degree of coarticulation between V_1 and

TABLE III. Average values in ms of three duration parameters: (a) pre-boundary vowel decay time, (b) pivotal consonant duration and (c) post-boundary vowel rise time, broken down by speaker and intended segmentation of the ambiguous two-word utterance. Standard deviations are given in parentheses; each individual mean is based on 20 utterances

Parameter	Boundary position	M1	M2	F_1	F_2	Mean
(a) Decay V1	C#	12 (12)	16 (10)	17 (11)	8 (6)	13 (10)
	#C	23 (3)	24 (8)	18 (10)	20 (9)	21 (8)
Difference		−11	−8	−1	−12	−8
(b) Consonant	C#	45 (11)	56 (21)	41 (12)	63 (15)	51 (16)
	#C	81 (14)	81 (20)	58 (20)	68 (18)	72 (20)
Difference		−36	−25	−17	−5	−21
(c) Rise V2	C#	21 (7)	20 (8)	22 (8)	10 (8)	18 (8)
	#C	14 (3)	18 (7)	13 (7)	8 (2)	13 (6)
Difference		7	2	9	2	5

the pivotal consonant. V_1 decay times are shorter if there is more coarticulation (articulatory overlap), viz. in /$V_1C\#$/ boundary position.

The observed *consonant durations* also showed a significant difference between /$V_1C\#V_2$/ and /$V_1\#CV_2$/ [Table III (b); $F(1, 33) = 14.86$, $p < 0.001$]. Speaker and interaction effects were insignificant. This difference between contrastive intended segmentations corroborates Nooteboom's (1972) results for Dutch. It exceeds the just noticeable difference (JND) for speech segment durations, which is usually reported to be in the range of 5% to 15% (Lehiste, 1970; Huggins, 1972). Hence the observed difference might constitute a primary (viz. sufficient) perceptual cue to the intended word segmentation.

The observed *rise times* of V_2, the nuclear vowel of the second word (following the ambiguous boundary position), also show a significant difference between /$C\#V_2$/ and /$\#CV_2$/ [Table III (c); $F(1, 33) = 4.64$; $p < 0.05$]. In addition, significant differences between speakers were observed [$F(3, 33) = 3.84$; $p < 0.05$], while the interaction effect was insignificant. Again, the direction of this difference corresponds with the predicted degree of coarticulation between V_2 and the pivotal consonant. V_2 rise times are shorter if there is more coarticulation (articulatory overlap), viz. in /$\#CV_2$/ boundary position.

The observed *durations* of V_2 show no significant difference between contrastive word boundary positions. Differences between speakers were insignificant. Again, the insignificance of the main effects is probably an artefact of pooling different vowels, with different phonological length and intrinsic durations. In this case, however, the situation seems to be more complex: all speakers produce such a difference, but they do so in opposite directions. Two speakers produce *longer* word-initial vowels, whereas two others produce relatively *short* word-initial vowels, as compared to non-word-initial vowels. This variable behaviour among speakers yields a significant interaction effect [$F(7, 33) = 3.10$, $p < 0.05$].

In summary, three parts of the stimulus fragments were produced with significantly different durations between /$V_1C\#V_2$/ and /$V_1\#CV_2$/ segmentations: (1) decay time of V_1, (2) consonant duration, (3) rise time of V_2. These parameters may have influenced subjects' perceived segmentation as acoustic-phonetic word boundary markers. Of these three, the difference in consonant duration (viz. word-initial consonant lengthening) is most significant. This phenomenon has already been reported for Dutch (Nooteboom, 1972; De Rooij, 1979) as well as for Swedish and English (Carlson *et al.*, 1972; Lehiste, 1960; Oller, 1973; Nakatani & Schaffer, 1978).

In addition, the results in Table III show considerable variation among speakers with regard to the produced differences between contrastive boundary positions, with possible compensation within each speaker between durational boundary phenomena with large and small variation. For example, speaker Female 2 hardly employs word-initial consonant lengthening as a boundary phenomenon. This is compensated for, however, by the differences in V_1 decay times she produces, which are the largest of all speakers. Likewise, the absence of V_1 decay time differences in the speech of Female 1 is compensated by her large differences in V_2 rise time.

2.5. *Discussion*

The acoustic analysis of the ambiguous word combinations indicates that three durational criteria may be used to discriminate between the contrastive word

TABLE IV. Pearson's correlation coefficients between perceived segmentation (rationalized arcsine of proportion of /#C/ responses), and three segmental duration parameters, broken down by speakers

Parameter	M1	M2	F_1	F_2	Pooled
Decay V1	0.49	0.47	−0.11	0.72*	0.34*
Dur. C	0.93**	0.93**	0.43	−0.23	0.60**
Rise V2	−0.54	−0.61	−0.19	−0.13	−0.37*

* $p < 0.05$; ** $p < 0.001$.

boundary positions. Although these parameters are boundary-dependent, they need not contribute to the perceived word segmentation. The first indication of their perceptual relevance can be obtained from the correlation between parameter values and perceived segmentation, as presented in Table IV. In these (and subsequent) statistical calculations, we used the rationalized arcsine of the proportion of $/V_1 \# CV_2/$ responses (Studebaker, 1985), rather than the raw percentages, because the former scale is linear, additive and normally distributed, while the latter is not.

Resulting coefficients show that the duration of the intervocalic consonant is the strongest durational correlate of the perceived word boundary position. In addition, both the decay time of V_1 and the rise time of V_2 constitute additional, weaker correlates.

Again, there are considerable differences among speakers. For example, responses correlate highly with consonant duration for stimuli realized by Male 1 and Male 2; this is not the case with stimuli from other (female) speakers. An interesting comparison with Table III shows that both male speakers produce the highest (absolute and relative) difference among speakers in consonant duration; speaker Female 2 produces the highest difference among speakers in V_1 decay time. The combined results of this experiment suggest that speakers can make their own selection from the potential word boundary markers. Moreover, subjects seem to pay attention only to those phenomena which are indeed effective segmentation cues for a given speaker (Quené, 1989).

However, it may not be concluded from the above correlations, that the observed word boundary dependent phenomena have indeed contributed to subjects' perceived segmentations. To this end, a stricter test is required, which is provided for by the following perception experiment.

3. Experiment 2

3.1. *Introduction*

In this experiment, the perceptual relevance of the three durational boundary markers is established. The central question here is whether these markers contribute to subjects' perceived word segmentation. Manipulation of boundary markers offers an appropriate method to answer this question. Under the hypothesis that such markers are indeed perceptually relevant, the perceived boundary position is predicted to shift or change as a function of the manipulation conditions (because different segmentations are indicated in the latter). In addition, we investigated which markers are perceptually most salient, by varying multiple boundary markers simultaneously. Perceived segmentation of the resulting hybrid stimuli (with conflicting markers) indicates which marker provides the strongest perceptual cue.

The three relevant boundary phenomena are established by variations in phonetic segment durations, with the "short pole" of the duration range corresponding to one intended segmentation and the "long pole" to the opposite segmentation. Hence, these boundary markers can be manipulated by means of shortening and lengthening of the relevant segment durations. The two manipulation conditions for each boundary marker yield a $2 \times 2 \times 2$ full factorial design. However, we wanted to avoid higher-order interactions between (stimuli containing) conflicting cues, since our primary goal is to establish the perceptual relevance of the boundary phenomena. For that reason, a 2×2 design with manipulation of only two of the three boundary phenomena was deemed more appropriate. For two reasons, the V_1 *decay time* was excluded as an experimental factor. First, Gårding (1967) provides evidence that pre-boundary markers are perceptually less efficient than post-boundary cues. Second, the correlation data from Experiment 1 (Table IV) suggest that on the whole, this cue is perceptually least efficient, albeit not for all speakers.

3.2. *Method*

3.2.1. *Stimulus material*

Considering the aim of the present experiment, as well as the observed differences in produced boundary markers between speakers (Table III), it seems reasonable to concentrate on the speaker who provides highest correlations between produced boundary phenomena and perceived segmentation. The relevant markers are likely to be present in this speaker's stimuli, and not be overshadowed by other, more salient boundary cues. On the basis of the correlation data from Experiment 1 (Table IV), speaker Male 2 was selected as the single speaker for the present experiment.

Each of the 20 minimal pairs of utterances by this speaker analysed in Experiment I were used to create a set of eight stimuli, yielding all possible combinations of (a) a "short" *vs.* "long" pivotal consonant, (b) a "short" *vs.* "long" V_2 rise time, and (c) two contrastive intended segmentations. The duration for the "short" and "long" portions were chosen to be 67% ("short") and 134% ("long") of their original duration. Using these ratios, changes in duration exceed the perceptual threshold (Lehiste, 1970), while the resulting segment durations are still acceptable. This design yields 2×2 conditions for each stimulus fragment.

The 40 digitized stimulus fragments were fed into a standard LPC analysis (10 filter poles; 25.6 ms window; 10 ms window shift). Durations were manipulated by changing the number of data-points (samples) per frame, to lower (shorter relative time) or higher (longer relative time) values for the relevant speech fragments. (The segment boundaries were located as closely as possible to those obtained in Experiment 1, considering the 10 ms integration time using LPC frames.) After LPC resynthesis, this procedure results in speech which is speeded up or slowed down, respectively (Vogten, 1983, p. 131). The remaining portion of the speech file was left unmanipulated (with original duration). F_0 was fixed at 100 Hz (monotone) in order to neutralize intonational differences which might function as word boundary markers. Such differences might be present in the original stimulus fragments as a result of lexical stress and sentence accents in the original sentences.

These $40 \times 4 = 160$ LPC speech files were then re-synthesized, DA-converted (at 10 kHz sampling frequency, 4.5 kHz low-pass filtering, 12 bits), and re-recorded in pseudo-random order onto audio tape, with an inter-stimulus interval of 2.5 s. Ten filler items preceded and followed the stimulus fragments. Total time of the test tape was about 11 min.

3.2.2. *Subjects and procedure*

Eighteen subjects listened to the test tape. Most of them were undergraduate students in various language and linguistics studies. Their participation was voluntary, but they received a small payment (ƒ5.=) for their services. The testing procedure, as well as subjects' instructions, were similar to those in Experiment 1 (see Section 2.3.2).

Subjects' responses were fed (manually) into a computer, which analyzed their choices and sorted these data according to the experimental conditions. The resulting dependent variable was the percentage of $/V_1\#CV_2/$ responses relative to all responses given for a particular stimulus fragment (and *not* the percentage of correct responses, as in Experiment I).

3.2.3. *Results*

The percentages of $/V_1\#CV_2/$ responses in each cell are presented in Table V. After rationalized arcsine transformation (Studebaker, 1985), these data were fed into an analysis of variance with the two manipulation variables (ConsDur and V2RiseDur), as well as (intended) Boundary Position as main fixed factors. Of the two word boundary markers, only manipulation of the intervocalic consonant duration yields a significant difference in the perceived word boundary position [ConsDur: $F(1, 152) = 16.1$; $p < 0.001$; V2RiseDur: $F(1, 152) = 2.3$; n.s.].

In addition, a significant effect of the (intended) Boundary Position was observed [$F(1, 152) = 94.2$; $p < 0.001$]. This indicates that the stimulus fragments contain additional word boundary markers, which also contribute to the perceived segmentation. From the results in Table V, it may be deduced that such additional markers were more salient in /C#V/ stimuli; the absence of these "positive" markers in the /#CV/ stimuli provides a less salient cue (as in Experiment 1; Section 2.3.3).

TABLE V. Perceived word segmentation (percentages of /#C/ responses), broken down by (intended) word segmentation and manipulation condition. The first character of the manipulation condition represents the intervocalic consonant duration, the second character the V_2 rise time ("S" = short, "L" = long). Each cell represents 20 (stimuli) × 18 (subjects) = 360 responses

Intended segmentation	Manipulation				
	SS	SL	LS	LL	Mean
/V1C#V2/	22	17	33	29	25
/V1#CV2/	53	44	63	61	55
Mean	37	31	48	45	40

Despite equal balancing of the stimulus material over the two boundary positions, the overall percentage of /#CV/ responses is significantly lower than the expected 50% (Kolmogorov–Smirnov test, with $p < 0.01$; Siegel, 1956). Finally, no significant interactions between main effects were observed.

3.3. *Discussion*

The above results show that the duration of the intervocalic consonant in /V(#)C(#)V/ context affects subjects' perceived word segmentation. Apparently, this parameter is a perceptually effective word boundary marker, both with and without other conflicting word segmentation cues (viz. V_2 rise time and additional markers not controlled for in this experiment).

By contrast, the rise time of the post-boundary vowel (V_2) does *not* contribute to perceived word segmentation in this experiment. Nevertheless, manipulation of this parameter has some influence on word segmentation, yielding a 5% difference. Its influence is apparently overruled, however, by other, stronger word boundary markers. The insignificant interaction effect between V2RiseDur and (intended) Boundary Position [$F(1, 152) = 0.007$; n.s.] seems to support this explanation. Otherwise, the 5% difference would be found only between conditions where the manipulated and uncontrolled boundary markers indicate identical *vs.* opposite segmentations.

These results indicate that subjects' perceived word segmentation was also affected by other boundary markers not controlled for in this experiment. Especially boundary segments seem to be relevant here (rather than spectral cues). Their distribution shows that they provide a positive cue to the intended /$V_1C\#V_2$/ segmentation (cf. Table II); moreover, a boundary segment can only be interpreted as the phonetic correlate of an intended boundary (as explained in Section 1).

4. General discussion

The acoustic analyses in Experiment 1 show that three durational parameters are produced with systematic differences between the contrastive segmentations of an ambiguous /$V_1(\#)C(\#)V_2$/ fragment in connected speech, viz. V_1 decay time, C duration, and V_2 rise time. Hence, these durational phenomena may provide cues to the intended word segmentation. The results of Experiment 2 demonstrate that this is the case for the intervocalic consonant duration: manipulation of this parameter affects the perceived segmentation. Besides, additional boundary phenomena not controlled in this experiment also contributed to word boundary detection. Presumably, boundary segments are the most important of these uncontrolled phenomena. Hence, the experiments reported here show that acoustic-phonetic cues contribute to word segmentation, at least under conditions where no other information is available (as in the present experiments).

However, the generalization of this conclusion to normal speech communication situations is somewhat complex. First, the acoustic analyses showed a great variability among speakers with regard to the phonetic contrasts they produce between opposite segmentations (Table III). Presumably, a certain amount of compensation between boundary markers is possible. This implies that we have to investigate the whole complex of phonetic correlates of contrastive segmentation, rather than separate boundary markers (cf. Pisoni & Luce, 1987; Repp, 1988). None of the phonetic boundary phenomena alone provides a reliable segmentation cue.

Second, an 80% accuracy in perceived segmentation was observed in Experiment 1, where acoustic-phonetic boundary markers provided the only segmentation cues. These cues alone apparently provide insufficient information for perfect word segmentation. In order to achieve 100% accuracy, higher-order information seems to be necessary.

Under normal speech communication situations, this higher-order information is indeed available; it was absent in the experiments reported here. In our stimuli, (1) no anticipation was possible, since pre-boundary words were monosyllables, recognizable only after their offset; (2) the stressed-onset strategy was ineffective, since subjects knew that the second syllable belonged to the post-boundary word in all cases; (3) phonotactic constraints were ineffective, since none of the possible segmentations was restricted; and (4) the subjects' task drew their attention to the speech signal, rather than to its semantic content. The combination of these stimulus properties forced listeners to employ some emergency procedure, in which they pay greater attention than usual to acoustic-phonetic word boundary markers. Consequently, the observed perceptual relevance of durational and segmental boundary markers is artificially high. In normal speech communication, boundary markers probably play only a secondary role, while linguistic and contextual information is decisive for word segmentation.

This perceptual preference for higher-order information is also found in word recognition (obviously interlinked with word segmentation). In a word-by-word gating experiment, about 20% of word recognitions occurred only after presentation of successive, longer gates (Bard, Shillcock & Altmann, 1988). The authors argue that in these cases, recognition is possible only after subsequent words are recognized, and their semantic content becomes available. Although word segmentation is almost trivial in this experiment (by definition of "word gate"), semantic and contextual information is crucial to the "late recognition" of the target word. Sensory information alone is insufficient, as in the present experiments (cf. also Spencer & Wollman, 1980; Austin & Carter, 1988).

This secondary role of acoustic-phonetic word boundary markers raises the question as to the relative importance of acoustic-phonetic and higher-order segmentation cues. If the experimental situation bears a greater resemblance to normal listening situations, then word segmentation is expected to be based more on non-phonetic information. This question can be investigated by using (1) meaningful *vs.* nonsense speech, where anticipation is impossible in the latter, (2) natural *vs.* synthetic speech, where uncontrolled boundary markers are absent in the latter, (3) monosyllabic *vs.* polysyllabic words with various stress patterns, (4) varying degrees of phonotactic restrictions, (5) segmentation *vs.* comprehension tasks. Several of these factors are currently being investigated, and will be reported on in the future.

My sincere thanks are due to Toon Cohen, for his inspiring supervision of the research reported here. In addition, I wish to thank M. E. Beckman, E. Gårding, an anonymous reviewer, Sieb Nooteboom and Frank Wijnen for their valuable comments on earlier versions of this paper.

References

Amerman, J. D. & Parnell, M. N. (1981) Influence of context and rate of speech on stop-consonant recognition, *Journal of Phonetics,* **9**, 323–332.

Austin, N. J. & Carter, E. (1988) Nonacoustic contributions to the perception of fluent speech, *Language and Speech,* **31**, 21–41.

Bakker, J. J. M. (1971) Constant en Variabel: de fonematische structuur van de Nederlandse woordvorm. Dissertation Universiteit van Amsterdam.

Bard, E. G., Shillcock, R. C. & Altmann, G. T. M. (1988) The recognition of words after their acoustic offset in spontaneous speech: Effects of subsequent context, *Perception and Psychophysics,* **44**, 395–408.

Barry, W. J. (1981) Internal juncture and speech communication, *Arbeitsberichte Institut für Phonetik Universität Kiel (AIPUK),* **16**, 229–287.

Barry, W. J. (1984) Perception of juncture in English. In *Proceedings of the tenth international congress of phonetic sciences, Utrecht 1983* (M. P. R. van den Broecke & A. Cohen, editors), pp. 529–536. Dordrecht, Cinnaminson: Foris.

Carlson, R., Granström, B., Lindblom, B. & Rapp, K. (1972) Some timing and fundamental frequency characteristics of Swedish sentences: data, rules, and a perceptual evaluation, *Quarterly Progress and Status Report, Speech Transmission Laboratory, Royal Institute of Technology (KTH),* no. 4, 11–19.

Cohen, A. (1962) De tijd als parameter bij spraakperceptie, *Handelingen van het zevenentwintigste Nederlandse filologencongres,* **27**, 48–54. Groningen: Wolters.

Cohen, A. (1987) Juncture revisited, In *In honour of Ilse Lehiste* (R. Channon & L. Shockey, editors), pp. 7–18. Dordrecht: Foris.

Cohen, A., Schouten, J. F. & 't Hart, J. (1962) Contribution of the time parameter to the perception of speech. In *Proceedings of the fourth international congress of phonetic sciences, Helsinki 1961* (A. Sovijärvi & P. Aalto, editors), pp. 555–560. The Hague: Mouton. Janua Linguarum; 10.

Cutler, A. (1989) Auditory lexical access: where do we start? In *Lexical representation and process* (W. Marslen–Wilson, editor), pp. 342–356. Cambridge MA, London: MIT Press.

Daniloff, R. & Moll, K. (1968) Coarticulation of lip rounding, *Journal of Speech and Hearing Research,* **11**, 707–721.

Frauenfelder, U. H. (1985) Cross-linguistic approaches to lexical segmentation, *Linguistics,* **23**, 669–687.

Gårding, E. (1967) *Internal juncture in Swedish.* Lund: Gleerup. Travaux de l'Institute de Phonétique de Lund; 6.

Garnes, S. & Bond, Z. S. (1980) A Slip of the Ear: A Snip of the Ear? A Slip of the Year? In *Errors in linguistic performance: slips of the tongue, ear, pen, and hand* (V. A. Fromkin, editor), pp. 231–239. New York, London: Academic Press.

Grosjean,, F. & Gee, J. P. (1987) Prosodic structure and spoken word recognition, *Cognition,* **25**, 157–188.

Harris, M. O., Umeda, N. & Bourne, J. (1981) Boundary perception in fluent speech, *Journal of Phonetics,* **9**, 1–18.

Hoard, J. E. (1966) Juncture and syllable structure in English, *Phonetica,* **15**, 96–109.

Huggins, A. W. F. (1972) Just-noticeable differences for segment duration in natural speech, *Journal of the Acoustical Society of America,* **51,** 1270–1278.

Jongenburger, W. & Heuven, V. Van (1991) The distribution of (word initial) glottal stop in Dutch. In *Linguistics in the Netherlands 1991* (F. Drijkoningen & A. van Kemenade, editors), pp. 101–110. Amsterdam, Philadelphia: John Benjamins.

Klatt, D. H. (1975) Vowel lengthening is syntactically determined in a connected discourse, *Journal of Phonetics,* **3**, 129–140.

Klatt, D. H. (1976) Linguistic uses of segmental duration in English: acoustic and perceptual evidence, *Journal of the Acoustical Society of America,* **59**, 1208–1221.

Klatt, D. H. (1980) Speech perception: a model of acoustic-phonetic analysis and lexical access. In *Perception and production of fluent speech* (R. A. Cole, editor), pp. 241–288. Hillsdale NJ: Lawrence Erlbaum Ass.

Lamel, L. F. (1984) The use of structural constraints to determine word boundaries, *MIT Speech Communications Group Working Papers,* **6**, 65–72.

Lehiste, I. (1960) An acoustic-phonetic study of internal open juncture, *Phonetica,* **5,** supplement. 5-54.

Lehiste, I. (1965) Juncture. In *Proceedings of the fifth international congress of phonetic sciences, Münster 1964* (E. Zwirner & W. Bethge, editors), pp. 172–200. New York, Basel: Karger.

Lehiste, I. (1970) *Suprasegmentals.* Cambridge, MA: MIT Press.

Lindblom, B. & Rapp, K. (1973) Some temporal regularities of spoken Swedish, *Papers from the Institute of Linguistics, University of Stockholm,* **21**.

Loots, M. (1983) Syntax and assimilation of voice in Dutch. In *Sound structures: studies for Antonie Cohen* (M.P.R. van den Broecke, V. J. van Heuven & W. Zonneveld, editors), pp. 173–182. Dordrecht, Cinnaminson: Foris. Publications in Language Sciences; 13.

Marslen-Wilson, W. D. & Welsh, A. (1978) Processing interactions during word recognition in continuous speech, *Cognitive Psychology,* **10**, 29–63.

Moll, K. L. & Daniloff, R. G. (1971) Investigation of the timing of velar movements during speech, *Journal of the Acoustical Society of America,* **50**, 678–684.

Nakatani, L. H. & Dukes, K. D. (1977) Locus of segmental cues for word juncture, *Journal of the Acoustical Society of America,* **62**, 714–719.

Nakatani, L. H. & Schaffer, J. A. (1978) Hearing 'words' without words: prosodic cues for word perception, *Journal of the Acoustical Society of America,* **63**, 234–245.

Nooteboom, S. G. (1972) Production and perception of vowel duration: a study of durational properties of vowels in Dutch. Dissertation Rijksuniversiteit Utrecht.

Oller, D. K. (1973) The effect of position in utterance on speech-segment duration in English, *Journal of the Acoustical Society of America,* **54**, 1235–1247.

Peterson, G. E. & Lehiste, I. (1960) Duration of syllable nuclei in English, *Journal of the Acoustical Society of America,* **32**, 693–703.

Pickett, J. M. & Pollack, I. (1963) Intelligibility of excerpts from fluent speech: effects of rate of utterance and duration of excerpts, *Language and Speech,* **6**, 151–164.

Pisoni, D. B. & Luce, P. A. (1987) Trading relations, acoustic cue integration, and context effects in speech perception. In *The psychophysics of speech perception* (M. E. H. Schouten, editor), pp. 155–172. Dordrecht, Boston, Lancaster: Martinus Nijhoff. NATO Advanced Science Institutes, Series D: Behavioural and Social Sciences; 39.

Quené, H. (1989) The influence of acoustic–phonetic word boundary markers on perceived word segmentation in Dutch. Dissertation Rijksuniversiteit Utrecht.

Quené, H. (1992) Integration of acoustic–phonetic cues in word segmentation. To appear in *The auditory processing of speech: from sounds to words* (M. E. H. Schouten, editor). Berlin: Mouton de Gruyter.

Repp, B. H. (1988) Integration and segregation in speech perception, *Language and Speech,* **31**(3), 239–271.

Rooij, J. J. De (1979) Speech punctuation: an acoustic and perceptual study of some aspects of speech prosody in Dutch. Dissertation Rijksuniversiteit Utrecht.

Shockey, L. (1974) Phonetic and phonological properties of connected speech, *Ohio State University Working Papers in Linguistics,* **17**, iv–143.

Siegel, S. (1956) *Nonparametric statistics for the behavioral sciences.* New York: McGraw Hill; Tokyo: Kogakusha. International student edition.

Slis, I. H. (1985) The voiced-voiceless distinction and assimilation of voice in Dutch. Dissertation Katholieke Universiteit Nijmegen.

Spencer, N. J. & Wollman, N. (1980) Lexical access for phonetic ambiguities, *Language and Speech,* **23**, 171–198.

Studebaker, G. A. (1985) A "rationalised" arcsine transform, *Journal of Speech and Hearing Research,* **28**, 455–462.

Su, L.-S., Daniloff, R. & Hammarberg, R. (1975) Variation in lingual coarticulation at certain juncture boundaries, *Phonetica,* **32**, 254–263.

Taft, L. (1984) Prosodic constraints and lexical parsing strategies. Ph.D. thesis University of Massachusetts.

Trommelen, M. T. G. (1983) *The syllable in Dutch: with special reference to diminutive formation.* Dordrecht: Foris. Publications in Language Sciences; 15.

Trubetzkoy, N. S. (1939) *Grundzüge der Phonologie.* Prague. Travaux du Cercle Linguistique de Prague; 7.

Umeda, N. & Coker, C. H. (1975) Subphonemic details in American English. In *Auditory analysis and perception* (G. Fant & M. A. A. Tatham, editors), pp. 539–564. London: Academic Press.

Vogten, L. L. M. (1983) Analyse, zuinige codering en resynthese van spraakgeluid. Dissertation Technische Hogeschool Eindhoven.

Appendix

List of ambiguous stimulus phrases, with speaker identification for each realization (intended boundary position), as used in Experiment 1.

Type 1: *single pivotal consonant*

No.	Cons.	Phrase	Spk /C#/	Spk /#C/
1	p	/di p ɪn/	F2	F1
2	p	/rɛi p okər/	M2	M1
3	t	/bɛi t ɑl/	M2	M1
4	t	/la t ɔp/	F2	F1
5	k	/pʌy k al/	M1	F1
6	k	/rɛi k ʌyt/	M1	M1
7	f	/ru f ɔp/	M2	F2
8	f	/wɛi f ojə/	F1	M2
9	s	/mu s ɔp/	F1	F2
10	s	/zɛi s ɔm/	M2	M1
11	x	/bʌy x yrə/	M1	M2
12	x	/ma x aut/	F1	M1
13	j	/ma j arə/	F2	F2
14	ʋ	/me ʋ ɛikə/	M2	F2
15	m	/du m er/	F2	F1
16	m	/na m ɔp/	F1	F1
17	n	/zo n ajt/	M2	F1
18	n	/tu n ɑrə/	M2	M2
19	l	/ku l ɑœf/	F1	M2
20	l	/mɛi l et/	M1	F2
21	r	/mu r ok/	M2	F1
22	r	/pa r ʌycəs/	M1	F2

Type 2: *two boundary consonants, ambiguous offset* /*VC*(#)*C*(#)*V*/

No.	Cons.	Phrase	Spk /C#/	Spk /#C/
23	p	/hil p ɑs/	M2	M2
24	p	/stɔm p ɛx/	M2	F2
25	t	/byr t ɪn/	M1	F1
26	t	/zɪx t er/	M2	M1
27	t	/ɑl t ɔp/	M1	M1
28	t	/kɑn t ɑl/	F2	F2
29	k	/bɑl k ɛikə/	M2	F2
30	k	/pɛr k et/	M1	F1
31	s	/bɪt s ɔm/	F2	M1
32	s	/hɛk s ɔp/	F1	F1
33	s	/par s ɔp/	M1	M1
34	f	/wɔl f ojə/	F2	M1
35	x	/bɑl x an/	F2	F2
36	m	/wɑl m et/	M1	F1
37	m	/wir p ɑs/	F2	M1

Type 3: *two boundary consonants, ambiguous onset /V(#)C(#)CV/*

No.	Cons.	Phrase	Spk /C#/	Spk /#C/
38	p	/ze p rat/	M2	F2
39	p	/ku p lʌymə/	F2	F1
40	t	/ma t rɔsə/	M2	M1
41	t	/mɛi t wɪst/	M2	F2
42	k	/bʌy k let/	M1	F1
43	k	/wɛi k rakə/	F2	M1
44	k	/wi k wɛl/	M2	F1
45	k	/rɛi k nɑpə/	F1	M1
46	f	/di f lɛs/	F1	F1
47	x	/bʌy s xen/	M2	M2
48	s	/wɛi s mɑkə/	F2	F1
49	s	/rø s lʌykə/	F1	F1
50	s	/pa s narə/	M2	M1
51	x	/rɛi x rɔtə/	F2	F1
52	x	/wi x lɑs/	F2	F2