Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 19 (1995), 51-69.

HOW TRANSITIONS AND LOCAL CONTEXT AFFECT SEGMENT IDENTIFICATION*

R.J.J.H. van Son and Louis C.W. Pols

Abstract

Theories on the mechanisms of phoneme identification generally involve only the actual segment and the transitions to and from neighbouring segments. In a listening experiment we tested the importance for vowel and consonant identification of the presence of speech segments beyond the transition parts. The results clearly show that identification continues to improve when speech is added beyond the boundaries of the transitions to neighbouring phonemes. Adding speech in front of the target segment improved identification more than adding speech at the back of the target segment, even if the sound added in front was actually not part of the target phoneme itself. We also show that the identification of consonantal segments correlates with the correct identification of the vowel, and vice versa, in CV-type stimuli, but not in VC-type stimuli. From these results, we conclude that context beyond the CV- and VC-transitions is used for both consonant and vowel identification.

1. Introduction

Current theories that try to model vowel and consonant identification are centred around the segment proper and the transitions between segments. Papers that discuss these theories indeed do not mention speech beyond the nearest consonant-vowel transition (e.g., for vowels Strange, 1989; Fox, 1989; Andruski and Nearey, 1992; see also Van Son, 1993a,b; Harrington and Cassidy, 1994; for consonants Pickett et al., 1994). In general, these papers use a subdivision of the speech signal into "proper segments", whose articulation is dominated by a single underlying phoneme, and the transitions between these proper segments, whose articulatory properties are determined by a changing mix of the two flanking phonemes. In papers on vowel production and recognition, the "segment proper" is usually called the vowel kernel. The boundaries between the vowel kernel and the transitions are ill-defined (see, e.g., Benguerel and McFadden, 1989).

It has long been known that the transitions to and from neighbouring segments, especially vowels, are often essential for the correct identification of consonants. Many studies have investigated the contributions of acoustic cues from neighbouring vocalic transitions to consonant identification and its importance relative to cues from the consonantal segment proper (e.g., Cooper et al., 1952; Delattre et al., 1955; Ohde and Sharf, 1977; Pols and Schouten, 1978; Pols, 1979; Miller and Bear, 1983; Mack

* A preliminary version of this paper was presented at Eurospeech 95 in Madrid (Van Son and Pols, 1995).

and Blumstein, 1983; Tarter et al., 1983; Polka and Strange, 1985; Diehl and Walsh, 1989; Mann and Soli, 1991; Nossair and Zahorian, 1991; Ohde, 1994; Cassidy and Harrington, 1995). From these studies it becomes clear that human listeners are quite able to identify many consonants from vocalic transitions alone. It is also clear that, if present, listener do use cues from outside the consonantal segment proper to identify it. However, only rarely is the influence of the next vowel itself on consonant identification, as opposed to only the vocalic transitions, investigated or acknowledged (Ohde and Sharf, 1977; Mann and Soli, 1991; Ohde 1994). An exception might be made for voicing contrasts, where the duration and F_0 trajectory of the whole neighbouring vowel is generally used to explain identification results (Lisker; 1986; Van Santen, 1993).

There has been a long standing discussion about whether consonant-vowel transitions are used in vowel identification (both CV and VC). Many studies do not acknowledge a role for the transitions (Lehiste and Peterson, 1961; Nearey and Assmann, 1986; Miller, 1989; Nearey, 1989; Andruski and Nearey, 1992). However, other studies stress the importance of these transitions for correct vowel identification (Lindblom and Studdert-Kennedy, 1967; Strange et al., 1976; Gottfried and Strange, 1980; Strange et al., 1983; Pols et al., 1984; Verbrugge and Rakerd, 1986; Benguerel and McFadden, 1989; Di Benedetto, 1989; Fox, 1989; Strange, 1989a,b; Jenkins et al., 1994; see also Van Son, 1993a,b). Even these latter studies do not assess the importance of speech beyond the vocalic part of the transition.

From studies of speech production, it is clear that the effects of coarticulation and assimilation affect complete phoneme segments. The often profound changes induced by coarticulation do not seem to bother the listeners. Identification does not seem to deteriorate from coarticulation (Gottfried and Strange, 1980; Macchi, 1980; Strange and Gottfried, 1980; Strange, 1989a). Even the spread of features like rounding and nasality over other segments seems not to deteriorate identification (Manuel, 1995). This is remarkable, because influences are often strong and can range well beyond the neighbouring segments (e.g., Öhman, 1966, 1967; Keating et al., 1994). Complementing the variation in the pronunciation of individual phonetic segments are regularities in the interaction between phonemic segments that could be used in the "reconstruction" of the intended phonemes. For example, the duration of a vowel is linked to the voicing of a following consonant (Lisker, 1986; Van Santen, 1993). Vowels can interact across intervening consonant clusters (Öhman, 1966, 1967; Benguerel and McFadden, 1989), and the articulation of intervocalic consonants can be described as a perturbation of the vowel-vowel trajectory (Öhman, 1966, 1967; Keating et al., 1994). One of the more consistent regularities of Consonant-Vowel sequences can be described by Locus Equations that correlate F_2 values in the centre of the vowel realization with the values found at the start of the CV transition (e.g., Schouten and Pols, 1979a,b, 1981; Sussman et al., 1991, 1993, 1995).

All studies point towards a universal occurrence of across-segment regularities in speech production. It is natural to ask whether these regularities are used by listeners when they try to identify the individual segments.

From a recent review of the literature on vowel identification (Van Son, 1993a,b), it became clear that the experiments that are generally thought to support the importance of the transitions for vowel identification in fact could not distinguish between the effects of the vowel on- or offset transitions and the effects of the consonantal context itself (from Lindblom and Studdert-Kennedy, 1967 to Andruski and Nearey, 1992). Where such a distinction would have been possible (e.g., Pols and

Van Son, 1993; Van Son and Pols, 1993), the results could show a detrimental effect of the presence of (synthetic) transitions without an appropriate context. Studies on speaker-normalization do show an influence of sentence context (Verbrugge et al., 1976), although in this case too it seems that identification deteriorated from it.

Abstracting from specific theories on coarticulation and assimilation, the question is whether listeners can, and do, use contextual speech from beyond the nearest transitions to identify individual phonemes. In an experimental study, this translates into the question whether the presence or absence of (the original) context influences identification. Ohde and Sharf (1977, see also the comments in Pols and Schouten, 1978; Pols, 1979; and the thesis of Klaassen-Don, 1983) investigate this question using plosives and a few vowels in CV and VC tokens uttered in isolation. They do find a relatively small effect of context on the identification of vowels and consonants. However, due to the limited inventory and the fact that the (non-word) syllables were pronounced in isolation, it is possible that the context effects could be much more salient in "normal" speech with a considerably larger phoneme inventory and more coarticulation and reduction.

There are other studies that supply information about the influence of context on the identification of vowels (e.g., Benguerel and McFadden, 1989; Kuwabara, 1983, 1985, 1993; Huang, 1991, 1992) or consonants (Mann and Soli, 1991; Ohde, 1994). The results of these studies too suggest that listeners do indeed use cues from speech originating beyond the nearest transitions. These latter studies were not designed to answer this particular question so there are confounding factors that make extrapolating difficult. The most problematic factor generally being a lack of information on segmentation procedures.

In natural speech there are many, redundant, cues to segment identity. Experiments with synthetic speech can only assess a small subset of these cues and it is always difficult to ascertain whether these are indeed relevant to natural speech. Even in natural speech, there can be hyper-articulated words or syllables that contain atypical cues to segment identity (Lindblom, 1990; Moon and Lindblom, 1994). This can be expected when isolated or semantically empty "target" words are used. Again, it can be difficult to be sure whether the results apply to normal, or hypo-articulated speech (cf., Ohde and Sharf, 1977; Pols and Schouten, 1978; Pols, 1979; Klaassen-Don, 1983; Mann and Soli, 1991).

In the present study, we tried to avoid these problems by using connected read speech (from a long, meaningful text). The downside of this approach is that it is nearly impossible to control all factors. In a normal text, the distribution of individual phonemes and phoneme combinations is highly unbalanced and many phonotactically allowed combinations will be missed. Still, it is the best way to ensure that the results are relevant to natural speech situations.

Table 1: Formant excursion sizes used to select CVC segments (see text). Numbers inside the table reflect the number of CVC segments actually used in the identification experiments.

		Large		Small	
		$\Delta F_2 \leq -175$	$\Delta F_2 \geq 175$	$-85 \leq \Delta F_2 \leq 85$	Total
Large	$\Delta F_1 \leq -125$	1	-	-	1
	$\Delta F_1 \geq 125$	19	20	20	59
Small	$-65 \leq \Delta F_1 \leq 65$	20	20	20	60
Total		40	40	40	120

In natural speech, listeners are quite good at inferring the context of a segment. Very small fractions of neighbouring segments often are sufficient to identify the context with high reliability (Ohde and Sharf, 1977; Pols and Schouten, 1978; Pols, 1979). At the other hand, in a full sentence, or even in words or syllables, the intended words can often be guessed, even when individual segments are not intelligible. This "lexical" information can be used to "correct" the identification of the individual segments. In an experiment that aims at assessing the use of acoustical information from the context for the identification of individual phonemic segments, enough speech must be presented to allow for the identification of the context, but not enough to allow for the identification of the original syllable of word. This can be achieved by using fragments of the context beyond the transitions, and at the same time ignoring word and syllable boundaries.

In this paper we will try to find an answer the question whether listeners indeed use speech sounds beyond the consonant-vowel transitions to identify both the vowel and the consonant. As a first step in this direction we will limit ourselves to speech fragments from the nearest neighbours of the target segment. Our experimental design is comparable to the design used by Ohde and Sharf (1977). However, we will use a larger inventory of CVC fragments which are taken from continuous read speech.

2. Material and Methods

2.1. Stimulus material

Tokens were constructed using vowels and their context from a pre-existing corpus (Van Son and Pols, 1990). The segments were taken from two readings of a long, informative text (844 words), read by a single, professional speaker. The speech was

Table 2: Number of C_1VC_2 speech tokens containing specific consonants and vowels. Median vowel length is 132 ms. (+/-): Numbers between brackets indicate the number of tokens containing a vowel realization that does/does not carry sentence accent.

Cons	C_1	(+/–)	C_2	(+/–)	Vowel	V	(+/–)
f/v	14	(4/10)	11	(4/7)	ɑ	10	(4/6)
s/z	8	(3/5)	9	(5/4)	aː	34	(18/16)
ʃ	4	(3/1)	0	(0/0)	ɛ	16	(7/9)
x	3	(2/1)	5	(1/4)	i	19	(13/6)
h	5	(1/4)	1	(1/0)	u	1	(0/1)
p/b	10	(6/4)	2	(1/1)	oː	37	(14/23)
t/d	16	(7/9)	12	(8/4)	y	3	(3/0)
k	4	(0/4)	5	(3/2)			
m	12	(7/5)	1	(1/0)			
n	17	(11/6)	11	(4/7)			
r	10	(7/3)	35	(13/22)			
l	3	(3/0)	18	(10/8)			
w	7	(2/5)	6	(6/0)			
j	2	(1/1)	0	(0/0)			
ŋ	0	(0/0)	3	(1/2)			
#	5	(2/3)	1	(1/0)			
Total	120	(59/61)	120	(59/61)		120	(59/61)

recorded on a commercial Sony PCM recorder, low-pass filtered at 4.5 kHz and digitized at 10 kHz, with 12 bit resolution. Subsequent storage, handling, and editing were done in digital form only.

All realizations of a subset of the Dutch vowels were isolated and labelled for a grand total of 1178 realizations. Vowel boundaries were determined using audio-visual cues from a waveform display. Each vowel segment contained all pitch periods that could not be attributed to the neighbouring segments. As a result, almost all vowel segments contained audible cues about their context. Each vowel realization was isolated with 50 ms of the surrounding speech. The identity of the flanking segments was transcribed and the presence of sentence accent on the vowel part was marked.

For the current identification experiment, 120 Consonant-Vowel-Consonant realizations were selected from this corpus (either Consonant could be silence, indicated by #). To be able to extract a reasonably stable vowel kernel, only CVC realizations for which the vowel duration was 100 ms or longer were used. The vowel formant F$_1$ and F$_2$ excursion sizes (i.e., ΔF$_1$ and ΔF$_2$, see Van Son, 1993a) were used as an indicator of coarticulation strength (Krull, 1989). At most 20 realizations each were selected at random from combinations of low and high excursion sizes for F$_1$ and F$_2$ (see table 1). For some combinations, less than 20 realizations were available. The phonemic structure of the CVC segments that were used is presented in table 2.

The actual tokens presented to the listeners were constructed from these speech segments (see table 3). For the vowel identification experiment, the vowel kernel was represented by the central 50 ms of the vowel realization (Kernel). From the complete vowel segment (with a median duration of 132 ms), 10 ms was removed from both sides to eliminate audible traces of the consonant. This was the Isolated Vowel token (V). The CVC token was constructed by adding 10 ms of context to both sides of the original vowel realizations (20 ms with respect to V). CV and VC tokens were constructed by removing, respectively, the vowel off-glide or on-glide transitions from the CVC tokens (leaving the Kernel part, or the central 50 ms, intact).

For both consonant identification experiments, tokens were constructed around the

Table 3: Construction of stimulus tokens and their durations (ms). Tokens consisted of fragments of the original CVC speech segments. The vowel kernel (Kernel) is represented by the central 50 ms of the Vowel. The On- and Off-glide Transitions of the vowels are the parts outside the vowel kernel up to the boundary with the consonant. The symbol * marks tokens used in the Vowel identification experiment. Tokens used in one of the consonant identification experiments are marked with +.

	Pre-Voc.	On-Glide	Vowel	Off-Glide	Post-Voc.	Min.	Median
	Cons.	Transition	Kernel	Transition	Cons.	Duration	Duration
Kernel*	-	-	50	-	-	50	50
V*	-	≥15	50	≥15	-	80	112
CVC*	10	≥25	50	≥25	10	120	152
CT$^+$	10	≥25	-	-	-	35	41
CCT$^+$	25	≥25	-	-	-	50	56
CV^{*+}	10	≥25	50	-	-	85	91
CCV$^+$	25	≥25	50	-	-	100	106
TC$^+$	-	-	-	≥25	10	35	41
TCC$^+$	-	-	-	≥25	25	50	56
VC^{*+}	-	-	50	≥25	10	85	91
VCC$^+$	-	-	50	≥25	25	100	106

CV and VC boundary, respectively. The shortest tokens contained only the vowel on- or off-glide transition up to, but not including, the central 50 ms (i.e., excluding the Kernel part) and 10 ms of the consonantal context (CT and TC). Longer tokens were constructed by adding either the central 50 ms of the vowel to the transitions (CV and VC, identical to those used in the vowel identification experiment) or an extra 15 ms of the consonant (CCT and TCC, the CC indicates an extended, 25 ms, consonant fragment), or additions at both the vowel and the consonant side (CCV and VCC). Before being presented, all these fragments were windowed with a 2 ms Hanning window at both sides to smooth the on- and off-set of the sounds.

2.2. Subjects and procedure

All subjects that participated in these experiments were students and staff members of our institute. Participation was voluntary and no rewards of any kind were offered. None of the subjects reported hearing problems. None of the subjects had heard the stimuli before and none was acquainted with the structure or construction of the stimuli. Tokens were presented separately for vowel identification, consonant identification in pre-vocalic position (CV-type tokens), and in post-vocalic position (VC-type tokens). For each subject, there was always more than a week between experiments.

17 subjects participated in the vowel identification experiment and 15 in both the CV and VC consonant identification experiments. Of these, 14 participated in all experiments. Subjects heard each token twice in quick succession over closed headphones in a quiet room at a comfortable sound level. They had to select a response from the orthographic set of all Dutch monophthongs (vowel identification experiment) or consonants (both consonant identification experiments). Subjects had to select a response symbol on a CRT-screen with a mouse-driven cursor. Presentations of phonemes in orthographic form poses no problem in Dutch and no training was required. The accompanying computer was outside the room and was not audible. The experiment was forced choice and self paced.

There were 600 tokens in the Vowel identification experiment and 480 tokens in both consonant identification experiments. These tokens were presented in a pseudo-random order that was different for each subject. Each experiment was preceded by a sequence of 10 practice tokens, taken to be the last 10 tokens of the particular sequence of the subject.

3. Results

3.1. Vowel identification

The results of the vowel identification experiment are displayed in figure 1 for all vowel realizations pooled as well as for accented and unaccented vowels separately. All differences between token classes are statistically significant (Macnemars' test, $p \leq 0.01$, two-tailed), except for the difference between the CV tokens and either the V (for +/– Accent) or the CVC type tokens (for All and –Accent). The differences between accented and non-accented vowels are statistically significant for the V, CV, and CVC token types ($\chi^2 \geq 12$, $\nu = 1$, $p \leq 0.01$).

It is clear from figure 1 that the central 50 ms of the vowel tokens (Kernel) is largely inadequate for identification. Even when long/short vowel errors are ignored, almost one in three responses is incorrect. The high error rate for the Kernel-type of

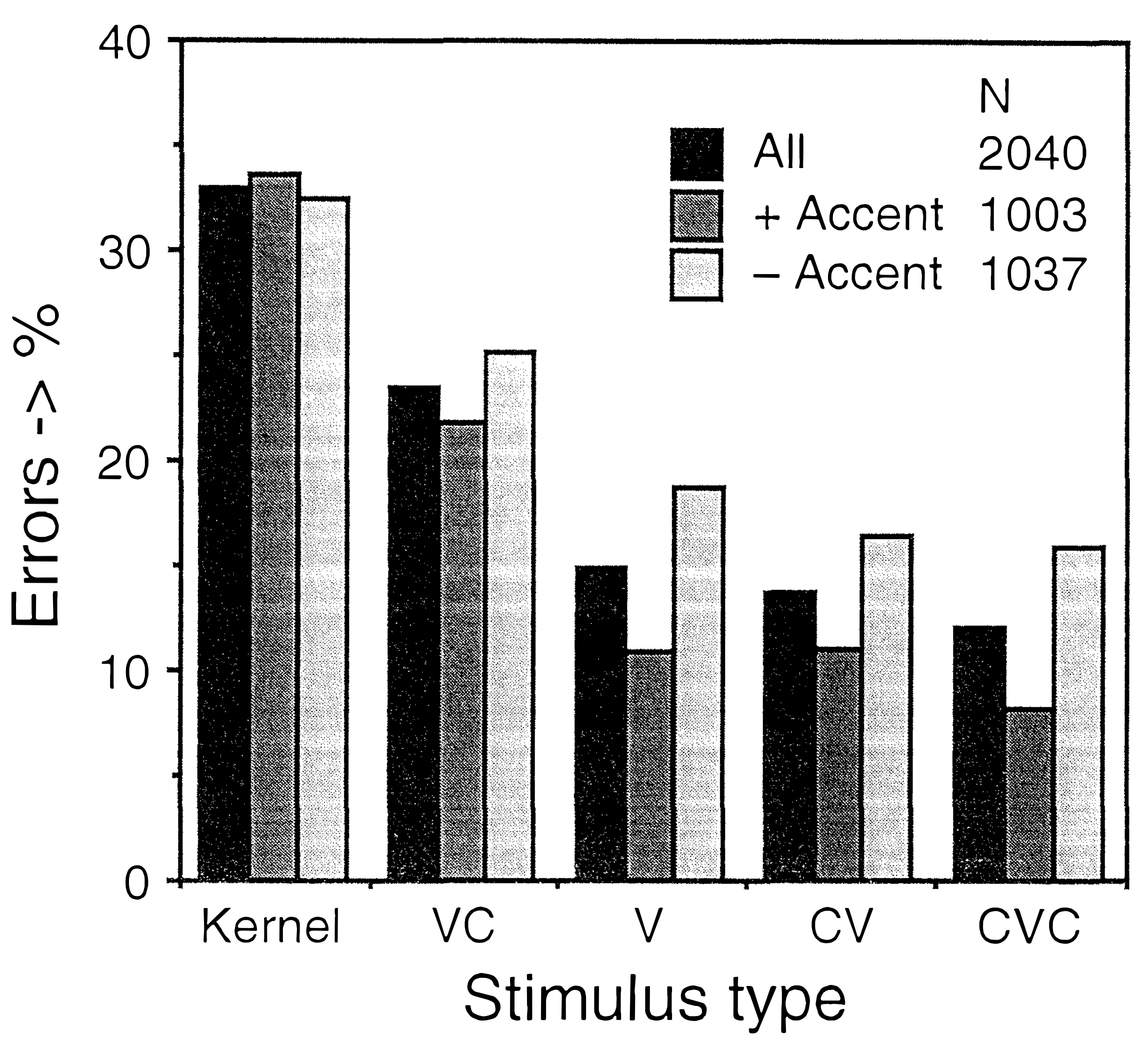

Figure 1. Error rates for vowel identification for the individual stimulus token types. Given are the results for all tokens pooled (All) as well as for vowels with and without sentence accent (+ and – Accent respectively) separately. Long-short vowel errors were ignored, i.e., /ɑ/-/aː/ and /ɔ/-/oː/ confusions in our experiment.

tokens is found for both the accented and the unaccented vowels, with no real difference between these two. When more of the vowel realizations and their context is added to the tokens, the differences between the error rates of accented and unaccented vowels grows. Therefore, it can be concluded that the effects of sentence accent on vowel intelligibility are to a large part mediated by the outer parts of the realizations and the context in which they are presented. In all other respects, the identification of both accented and unaccented vowel realizations benefits alike from the presence of context.

In figure 2 the results are split on long versus short-vowel stimuli (/aː oː/ versus /ɑɛiuy/). All differences between token classes are statistically significant (Macnemars' test, p≤0.01, two-tailed), except for the difference between the CV tokens and either the V- or the CVC-type tokens (for both long and short vowels) and between V- and CVC-type tokens (not significant for short vowels only). The differences between the error rates for long and short vowels are statistically significant for the Kernel and VC-type tokens only ($\chi^2 > 20$, $\nu = 1$, $p \leq 0.01$). For the Kernel-type stimuli, the errors are concentrated on the long-vowel realizations (/aː oː/). Still, there is a large error rate for the short-vowel tokens too (/ɑɛiuy/). The error rate drops sharply when a larger fraction of the vowel realizations is used. For the V-

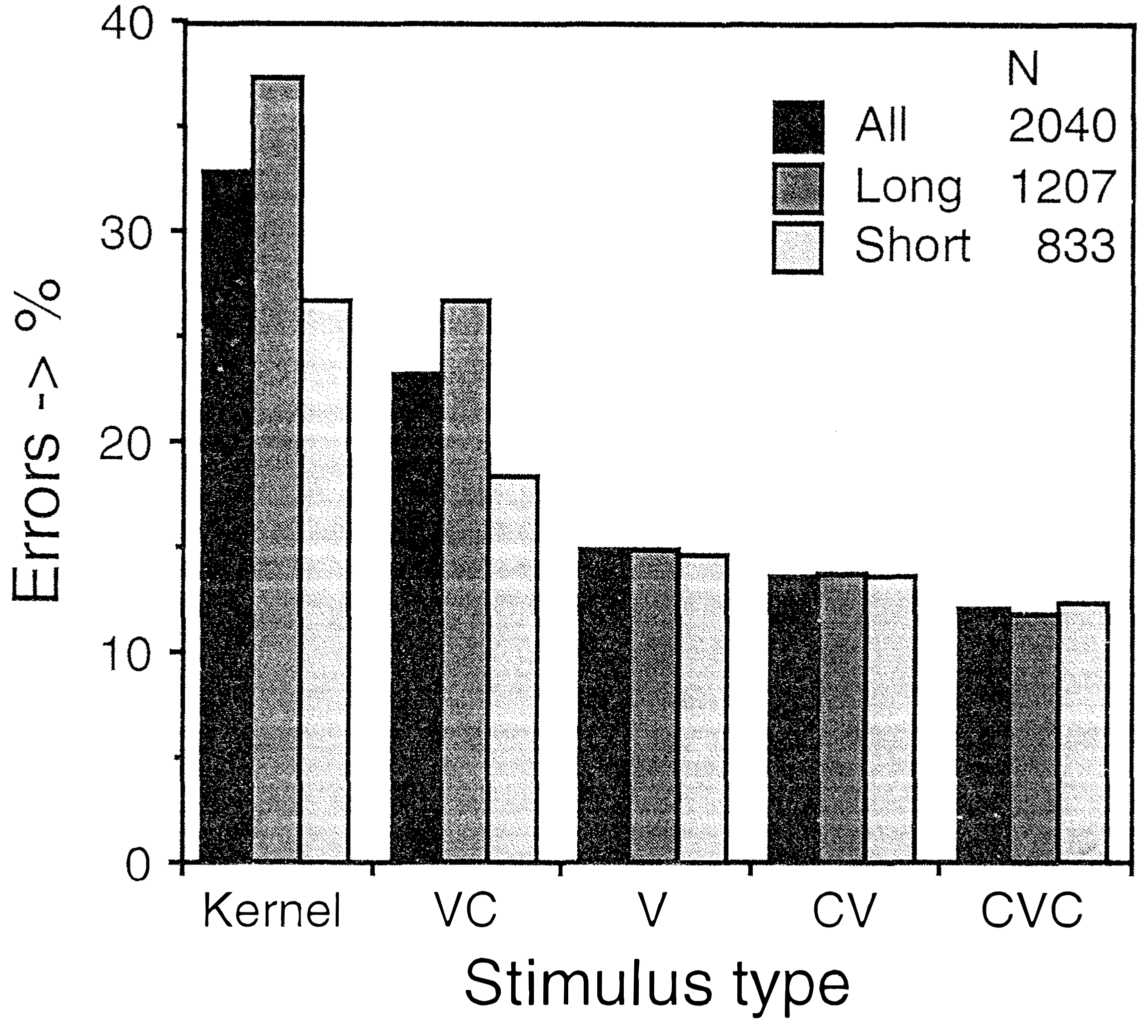

Figure 2. Error rates for vowel identification for the individual stimulus token types. Same data as figure 1 but now a subdivision is made in terms of intrinsically long and short vowels. Given are the results for all tokens pooled (All, identical to figure 1) as well as for long and short vowels (/a: o:/ and /ɑɛiuy/ respectively) separately. Long-short vowel response errors were ignored, i.e., /ɑ/-/a:/ and /ɔ/-/o:/ confusions in our experiment.

type of tokens, which include most of the vowel on- and off-glide, error rates drop to 15%. There is no relevant difference in error rates between long- and short-vowel stimuli vowels for the V and CVC type of tokens ($\chi^2 \leq 1$, $\nu = 1$, $p > 0.01$).

There is a large and statistically significant difference between the error rates for long- and short-vowel realizations in VC-type tokens. There is no such difference for the equally long CV-type tokens. This indicates that the (lack of) difference in intelligibility between long- and short-vowel realizations from different types of tokens cannot be attributed to only the duration of the tokens. The difference in intelligibility between long- and short-vowel tokens can most likely be attributed to diphthongization of the long-vowel realizations (see also Peeters, 1991; Andruski and Nearey, 1993). From figure 2 it is clear that the V-, CV-, and CVC-type stimuli contain enough dynamic "diphthong" information to blur the distinction in "intelligibility" between the realizations of long- and short-vowels. The Kernel- and VC-type stimuli are less adequate for the identification of such diphthongized vowels.

When the results of figures 1 and 2 are compared, it appears that the difference in intelligibility between accented and unaccented vowels is of a different type than that between long- and short-vowel realizations. The intelligibility differences with respect to sentence accent seem to be mediated by the "context", whereas the differences with respect to the long/short distinction seem to be segment (vowel) internal.

The V-type tokens already consisted of more than 80% of the duration of each vowel realization (85% median). However, the error rate still goes down when more speech is added. The error rate for CVC-type tokens is 12%, a fifth lower than for the V-type tokens. The difference in error rate between these two token types is statistically significant (Macnemars' test, $p \leq 0.01$, two-tailed). The drop in error rates is found for both accented and unaccented vowels (Macnemars' test, $p \leq 0.01$, two-tailed). For both conditions, the size of the difference is equally large (19% to 16% for unaccented and 11% to 8% for accented vowels). More or less the same holds for the long- and short-vowel realizations. However, probably due to the smaller number of short-vowel realizations, the difference is not statistically significant for this set.

Both CV- and VC-type tokens are better recognized than the central 50 ms alone (Kernel). It is clear that vowel identification benefits more from speech added in front of the kernel (CV-type tokens) than from speech added in the back of the kernel (VC-type tokens), with the error rate of the former almost half that of the latter. The intermediate position of the CV-type tokens between V and CVC-type tokens in the error rates suggests that the reduction of the error rates in the V- and CVC-type tokens is largely due to the added token onsets. The offset parts of the tokens seems to play only a minor role in reducing the error rate.

The vowel tokens were balanced with respect to the formant excursion sizes, ΔF_1 and ΔF_2 (see table 1). There was no detectable effect of first formant excursion size on the error rate (ΔF_1, not shown). However, there were large differences in identification errors due to differences in the second formant excursion size (ΔF_2, see table 4). Differences between the three sets of excursion sizes of the second formant were highly significant for all token types ($\chi^2 \geq 16$, $\nu = 2$, $p \leq 0.01$). For each excursion size, the tokens followed the same pattern of error-rates as was found for the vowels as a whole (3 rows in table 4, Friedman's $Q = 11.5$, $p \leq 0.05$, Kendal's concordance, i.e., mean rank correlation coefficient, $W = 0.956$). In general, the vowel realizations with large negative excursion sizes, $\Delta F_2 \leq -175$ Hz, induced the highest error rates, those with large positive excursion sizes, $\Delta F_2 \geq +175$ Hz, induced the lowest error rates. On average, the vowel realizations with small excursion sizes, $|\Delta F_2| \leq 85$ Hz, scored in between. This pattern was very consistent over token types (5 columns in table 4, Friedman's $Q = 6.4$, $p \leq 0.05$, Kendal's concordance $W = 0.64$).

From these results it is clear that the *absolute* size of the formant excursions (i.e., formant dynamics, either ΔF_1 or ΔF_2), was not related to vowel intelligibility in our tokens. The excursion size of the first formant had no effect whatsoever on the error rates. Large positive excursions of the F_2 were correlated to low error rates whereas large negative excursions were correlated to high error rates. Small F_2 excursion were in between. A possible explanation of this somewhat odd pattern can be found when the distribution of vowels over formant excursions is taken into account. The excursion size of the second formant (ΔF_2) correlates strongly with vowel height (Pols and Van Son, 1993; Van Son, 1993a). The strong and consistent effect of the F_2

Table 4. Vowel identification error rates (%) as a function of second formant excursion size (ΔF_2, Hz).

	Kernel	VC	V	CV	CVC	Mean
$\Delta F_2 \leq -175$	39	30	18	19	17	25
$\Delta F_2 \geq 175$	31	18	8	7	5	14
$\lvert\Delta F_2\rvert \leq 85$	29	22	19	16	14	20
Mean	33	23	15	14	12	19

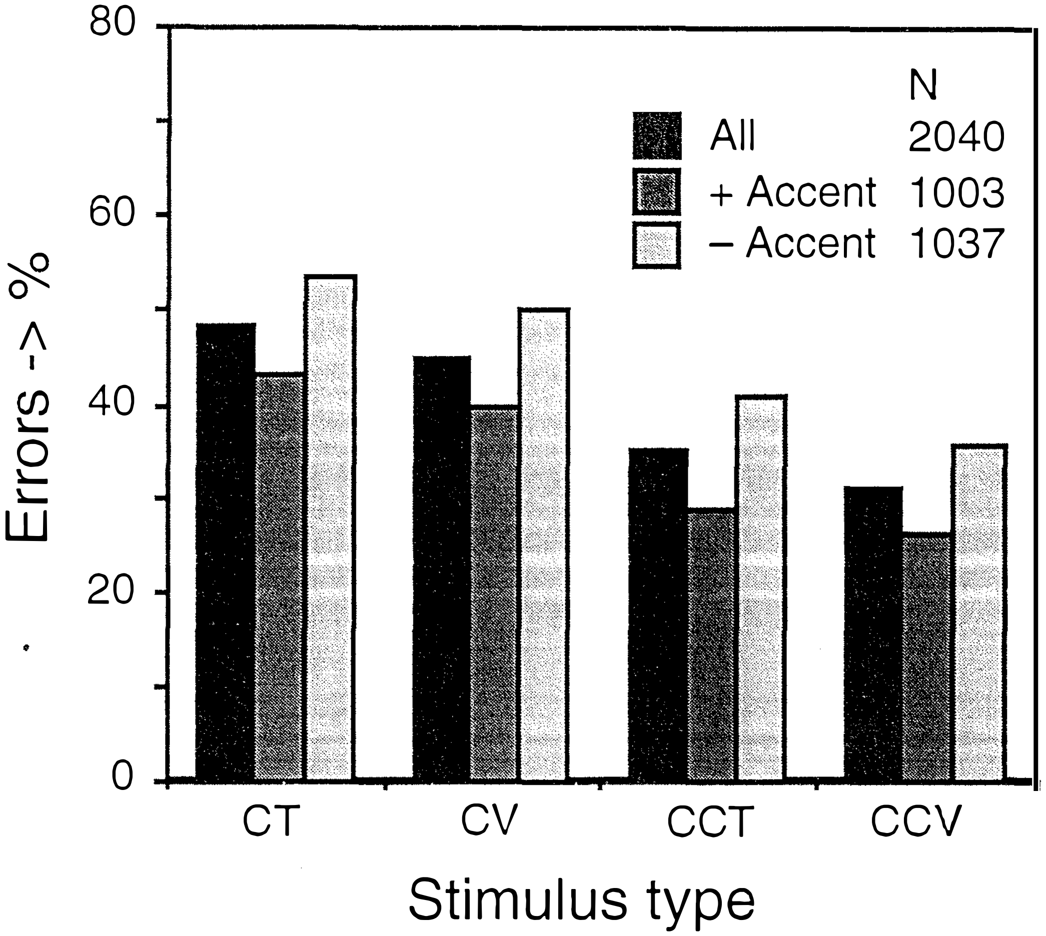

Figure 3. Error rates for the identification of pre-vocalic consonants (%). The error rates are calculated by ignoring voiced/voiceless confusions. Error rates are presented for all consonants pooled (All) as well as for Consonants preceding vowels carrying sentence accent (+Accent) and those preceding unaccented vowels (-Accent) separately.

excursion size on vowel identification can therefore be described as a correlation between vowel height and intelligibility. Higher vowels seem to induce less errors in our experiment.

3.2. Consonant identification

3.2.1. Consonant identification in pre-vocalic position

The consonant identification results were evaluated with respect to the correctness of the identification using different criteria. In figure 3, the error rates for consonant identification are presented, ignoring voicing errors. All differences between the different classes of tokens (CT, CV, CCT, and CCV) are statistically significant except for the consonants preceding accented vowels in CT- versus CV-type tokens and CCT- versus CCV-type tokens (Macnemars' test, $p \leq 0.01$, two tailed). The differences between consonants preceding accented and non-accented vowels are significant for all token classes ($\chi^2 \geq 18$, $\nu = 1$, $p \leq 0.001$). Both identification *per se* , and identification of only place or manner of articulation showed the same pattern of errors as identification based on ignoring voicing errors (not shown, Place: Labio-dental /fvpbmw/, Alveolar /sztdnl/, Palatal /ʃj/, Velar-Uvular /kxŋr/, Glottal /ɦ/;

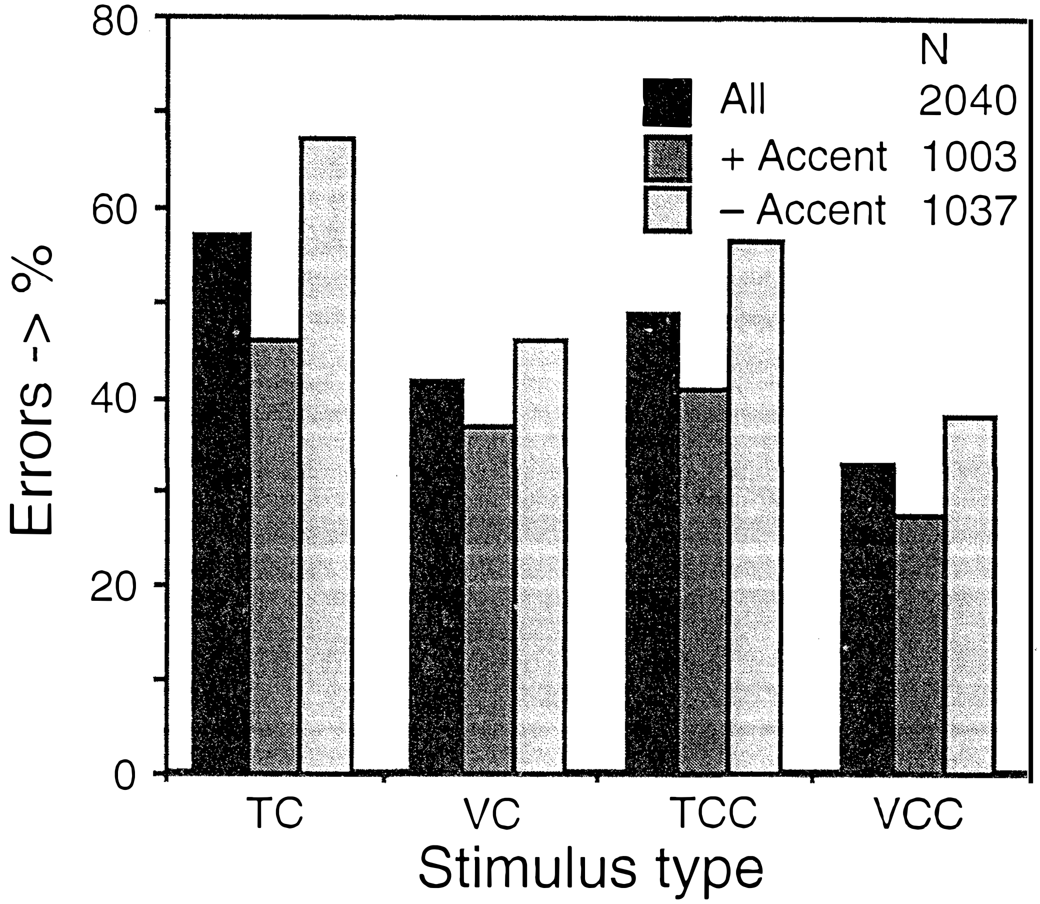

Figure 4. Error rates for the identification of post-vocalic consonants (%). The error rates are calculated by ignoring voiced/voiceless confusions. Error rates are presented for all consonants pooled (All) as well as for those following vowels carrying sentence accent (+Accent) and those following unaccented vowels (–Accent) separately.

Manner: Fricative /fvszʃgh/, Plosive /pbtdk/, Nasal /mnŋ/, Vowel-like /wljr/). The absolute error rates varied between the error criteria.

Voicing errors were relatively constant with respect to token types (between 2% and 3%). Error rates for both place and manner of articulation were high, around 36% and 19% respectively. Error rates for place and manner of articulation were strongly correlated for consonants preceding accented vowels ($r = 0.976$, $p \leq 0.025$, not shown) and somewhat less so for consonants preceding unaccented vowels ($r = 0.914$, $p > 0.025$ not significant, not shown). Therefore, confusions of manner and place of articulation both seem to play a similar role in determining the overall error rates. The more speech is added, the lower the error rates become in both categories (not shown). As with the vowel identification experiment, adding speech in front of a token (at the consonantal side) always reduces the error rate more than adding speech

Table 5. Pre-vocalic consonant identification error rates (%) as a function of second formant excursion size in the vowel (ΔF_2, Hz).

	CT	CV	CCT	CCV	Mean
$\Delta F_2 \leq -175$	51	48	42	38	45
$\Delta F_2 \geq 175$	34	30	25	19	27
$\lvert\Delta F_2\rvert \leq 85$	60	57	39	36	49
Mean	48	45	35	31	40

at the back of a token. That is, CCT-type tokens have a lower error rate than CV-type tokens, but both have lower error rates than CT-type tokens that are included in them. The lowest error rates are found for CCV type tokens, that completely overlap all other tokens.

For consonant identification too, the excursion size of the first formant of the vowel, ΔF_1, had no effect on consonant identification (not shown). There were large differences in consonant identification between tokens with different F_2 excursion sizes ($\chi^2 \geq 40$, $\nu = 2$, $p \leq 0.01$). All three groups of tokens with different ΔF_2 (see table 1) showed the same pattern of identification error-rates with respect to token classes (Friedman's $Q = 9$, $p \leq 0.05$, Kendal's concordance $W = 1$). Pre-vocalic consonants followed by a vowel with a large positive F2 excursion ($\Delta F_2 \geq 175$ Hz) always induced the lowest error rate.

3.2.2. Consonant identification in post-vocalic position

The pattern of identification errors for the post-vocalic consonants is similar to that of pre-vocalic consonants (figure 4). All differences between the token types are statistically significant except for the consonants following accented vowels in VC- versus TCC-type tokens (Macnemars' test, $p \leq 0.01$, two tailed). The differences between consonants following accented and non-accented vowels are significant for all token types ($\chi^2 \geq 14$, $\nu = 1$, $p \leq 0.001$). Again, there is no difference in the pattern of error rates for the different error types (i.e., including or excluding voicing errors and errors regarding manner or place of articulation, not shown). Error rates for both place and manner of articulation were high, around 36% and 31% respectively, and were strongly correlated for consonants following both accented and non-accented vowels ($r \geq 0.994$, $p \leq 0.01$, not shown).

It is clear that the error rate for post-vocalic consonants too reduces more when speech is added in front of the consonant than when it is added at the back. The VC-type stimuli induced lower error-rates than the TCC type stimuli. This is remarkable when it is considered that the difference between the TC and the VC-type tokens is the presence of the vowel kernel whereas the TCC-type tokens have a longer consonantal part. It seems that the position where speech is added to the minimal TC- or CT-type tokens is more important than the segment where this added speech originated.

In both figures 3 and 4 there is a strong effect of sentence accent (carried by the vowel) and consonant identification in the same token. This difference seems to be independent of the token type, i.e., the effect of sentence accent is even present when the vowel kernel is absent. Part of this difference is expected to be due to a difference in the consonant sets used (see table 2). However, the fact that both pre- and post-vocalic consonants (originating from different sets, see table 2) are identified worse when accompanying a non-accented vowel indicates that sentence accent is indeed an

Table 6. Post-vocalic consonant identification error rates (%) as a function of second formant excursion size of the vowel (ΔF_2, Hz).

	TC	VC	TCC	VCC	Mean
$\Delta F_2 \leq -175$	60	43	44	31	45
$\Delta F_2 \geq 175$	48	34	47	28	39
$\lvert\Delta F_2\rvert \leq 85$	62	48	55	39	51
Mean	57	42	49	33	45

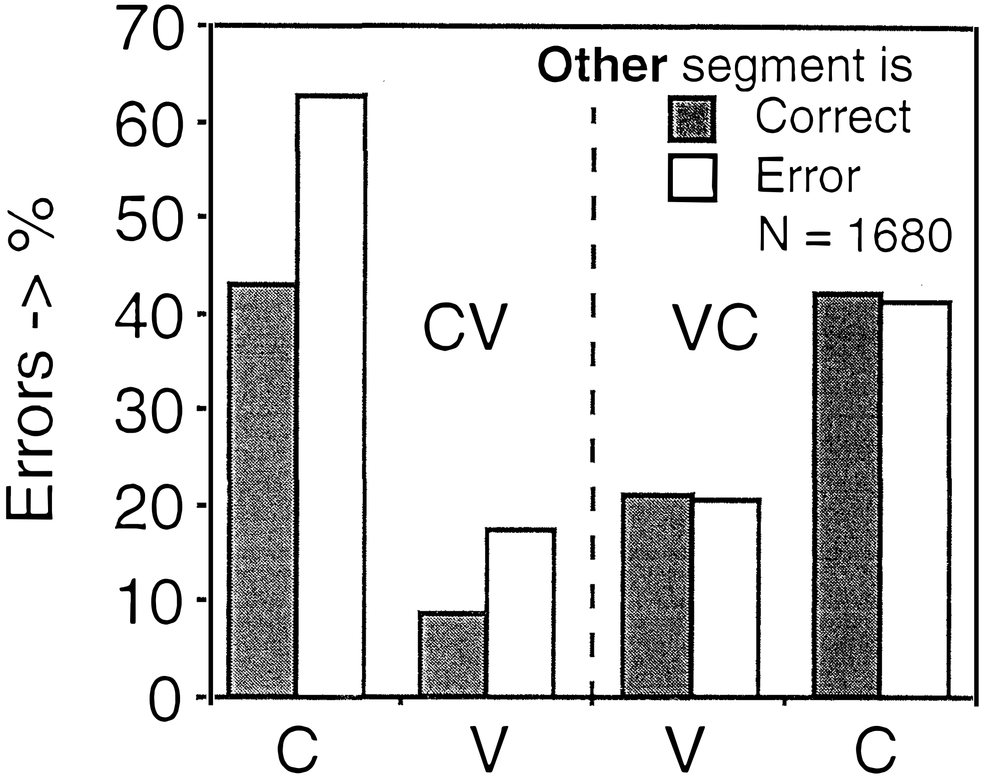

Figure 5. Error rates for vowel and consonant identification in CV- and VC-type tokens with respect to the correct and incorrect identification of the other segment in the same token. Voiced/Voiceless errors in consonants and Long/Short errors in vowels were ignored.

important factor for consonant identification. Again, this was found also for identification of place and manner of articulation.

For post-vocalic consonant identification too, the excursion size of the first formant, ΔF_1, had no effect on consonant identification (not shown). There were large differences in consonant identification between tokens with different F_2 excursion sizes ($\chi^2 \geq 14$, $\nu = 2$, $p \leq 0.01$, table 6). All three groups of tokens with different ΔF_2 (see table 1) showed the same pattern of identification error-rates with respect to token types (Friedman's Q = 9, $p \leq 0.05$, Kendal's concordance W = 1). Post-vocalic consonants preceded by a vowel with a small F_2 excursion ($|\Delta F_2| \leq 85$ Hz) always induced the highest error rate. Either of the other two groups with large F_2 excursion sizes could induce the lowest error rate.

3.3. Relation between Vowel and Consonant identification

In figure 5, the correctness of identification of one segment in a token is correlated to the correctness of identification of the other phoneme in the same token. The differences in error rate are statistically significant for the CV tokens only ($\chi^2 = 28.7$, $\nu = 1$, $p \leq 0.01$). For the VC-type tokens we see no relation at all. For the CV tokens the difference in error rate is large indeed. The vowel identification error rate almost doubles when the consonant is identified incorrectly with respect to when it is identified correctly. The results in figure 5 were obtained by ignoring long/short vowel and voicing errors (for vowel and consonant identification respectively). The

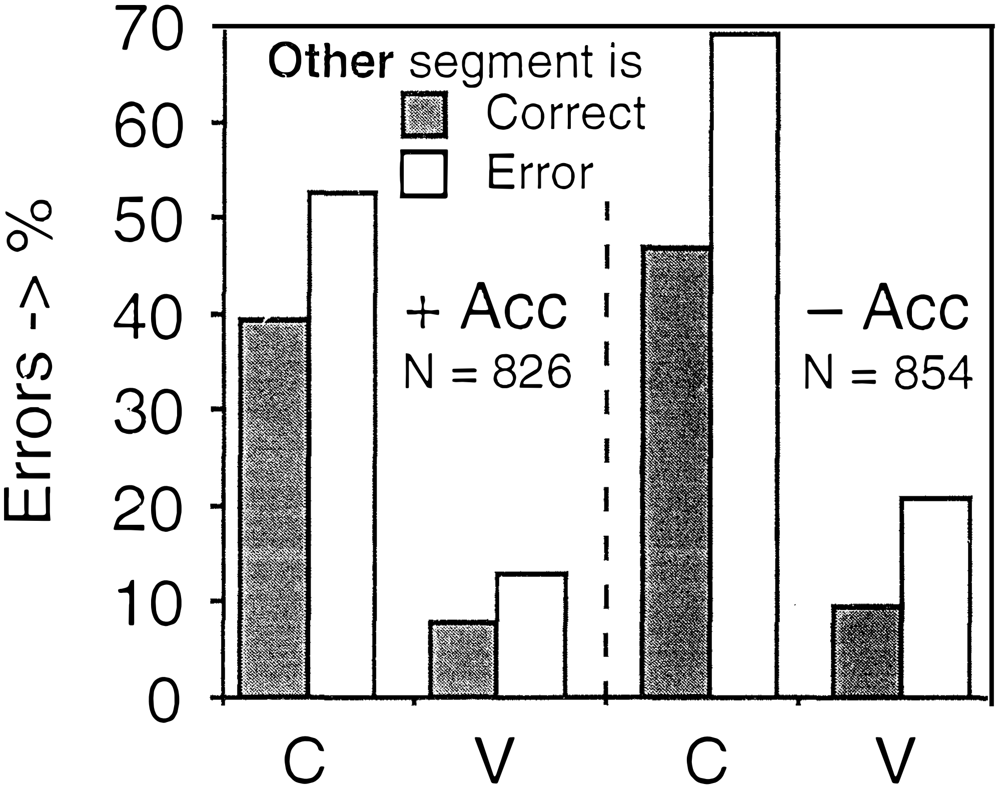

Figure 6. Error rates for vowel and consonant identification with respect to the correct and incorrect identification of the other segment in the same token. The data are a subset of those presented in figure 5. Presented are the results for CV-type tokens for which the vowels carried sentence accent (+Acc) or not (–Acc). Voiced/Voiceless errors in consonants and Long/Short errors in vowels were ignored.

same results were found when using consonant identification *per se*, or only errors in the place and manner of identification (not shown).

When the results for the combined identification are split on sentence accent, the pattern found is the same (see figure 6, CV-type tokens only). The error rates are higher when the other segment is identified incorrectly. However, these differences are only statistically significant for the tokens with unaccented vowels (–Acc, $\chi^2 = 21.9$, $\nu = 1$, $p \leq 0.01$).

4. Discussion

Our experiments are very much like those reported by Ohde and Sharf (1977). Contrary to this earlier paper, we do find strong effects of the presence of context on phoneme identification for vowels, pre-, and post-vocalic consonants. A likely explanation for this difference is a combination of two factors: First, in our experiments we used a larger inventory of consonants and vowels taken from CVC sequences, not only plosives and point vowels (/uia/) taken from CV- or VC-like combinations. Second, our stimuli were taken from a long, meaningful text read aloud instead of spoken in isolation. It is known that syllables spoken in isolation show less reduction and coarticulation than when reading aloud a long text. Therefore, it is to be expected that our stimuli show a larger variation in the strength and presence of cues to identity. This difference seems to force our subjects to rely more on contextual cues than the subjects of the earlier experiments.

From our results it follows that theories that try to explain vowel or consonant identification by the spectro-temporal structure of the segment proper (either the vowel "kernel" or the consonant without the transition) are incomplete. For vowels, it is shown that (diphthongized) long vowels and (monophthongal) short vowels alike benefit from context (figure 2). The Kernel-type tokens, consisting of only the central 50 ms of the original realizations, elicited very high error rates. These high error rates cannot be attributed to their short durations alone (only 50 ms). Experiments with synthetic vowels showed that Dutch subjects could identify synthetic vowels in isolation very consistently down to durations of only 25 ms (Pols and Van Son, 1993; Van Son, 1993a; Van Son and Pols, 1993). Comparable results were found for Japanese subjects who were generally able to identify natural vowel segments with durations down to 35 ms (Ohta et al., 1962). Even shorter, single pitch period stimuli, have been shown to be identified consistently (Van der Kamp and Pols, 1971; Fox, 1989). All this indicates that, instead of simply being too short, the 50 ms Kernel-type vowel fragments somehow lacked the information to be identified reliably.

When more of each vowel realization was present in the isolated vowel type tokens (V-type, containing around 85% of the original realization), the error rates decreased considerably (figures 1 and 2). This can be attributed to the presence of formant dynamics in the vowel realizations. Such a behaviour is expected if subjects use "target-overshoot" to compensate for the effects of coarticulation and reduction (Lindblom and Studdert-Kennedy, 1967; Strange, 1989a; Di Benedetto, 1989). However, irrespective of the token classes, the absolute size of F_1 and F_2 excursions, which are proxies for coarticulation, had no effect on error rates. This means that any "compensation" for coarticulation or any perceptual "target-overshoot" had to work equally well on the Kernel-type tokens as on the CVC-type tokens. But the former lack most, if not all, dynamic information. Therefore, an effect of the absolute formant excursion size in itself on our identification results seems implausible. Furthermore, it has been shown before that synthetic vowels with formant dynamics added without consideration for the central formant values or context, are identified "worse" (i.e., by "undershooting" instead of "overshooting" the target) than those without such synthetic formant dynamics (Pols and Van Son, 1993; Van Son, 1993a; Van Son and Pols, 1993).

There is undoubtedly a beneficial effect of the presence of transitions to vowel identification. However, "simple" target-overshoot related models cannot account for the fact that the absolute size of the formant excursions seems to be irrelevant. A possible explanation could be that the spectral change in the transition regions is a (redundant) independent cue to the identity of a vowel. That is, the spectral change in itself is recognized as a separate feature of a certain vowel and context.

For consonants, it was shown that identification benefited from adding the central part of the neighbouring vowel. For post-vocalic consonants, the effects of adding the vowel kernel in front of the token were larger than those of adding more of the consonant at the back. These results were found for both the place and the manner of articulation (not shown). This indicates that the high error rates were not only caused by a lack of structural, or "manner", information (e.g., plosives versus fricatives) but also from a lack of spectral, or "place", information. The high correlations between error rates calculated with respect to manner and place of articulation indicate that these two "dimensions" of consonant articulation are not independent from a perceptual point of view, at least not with respect to the manipulations we used to construct our tokens.

The presence or absence of sentence accent on the vowels had a strong effect on identification, both for the vowels themselves and for the consonants. For vowels, the difference between the error rates of accented and unaccented vowels increased with the amount of context (figure 1). This indicates that this difference in vowel intelligibility is due to the outer parts of the vowel realizations and possibly the context. For consonants, we could not find such a difference between token classes.

As with consonants, vowel identification benefited also from speech added at the periphery of the realization, crossing the segment boundaries. For the CVC-type tokens, the consonants were audible and this in itself seemed to have helped the listeners identifying the vowels, as was shown by the strong relation between correct identification of vowels and consonants in CV-type tokens (figures 5 and 6).

In conclusion, our results show that our listeners used the transition parts between vowels and consonants to identify both vowel and consonant realizations. If present, speech beyond these transitions was used too. In all experiments it could be shown that speech added in front of the target phoneme improved identification more than speech added at the back of the target phoneme. This was found even when the added speech originated from another phoneme (e.g., from the vowel when identification of post-vocal consonants was at stake, see figure 4). Asymmetries of this kind have been reported before but there seems to be no consensus about an explanation (Ohde and Sharf, 1977; Pols, 1979, but see also: Di Benedetto, 1989; Mann and Soli, 1991; Pols and Van Son, 1993; Van Son, 1993a; Van Son and Pols, 1993; Van Wieringen and Pols, 1991, 1995; Van Wieringen, 1995). The report of Mann and Soli (1991) is interesting in this respect because it states that the asymmetry is reversed, it is the vowel following /fʃ/ that contributes the strongest cues to fricative identification. Our data support the explanation proposed by Ohde and Sharf (1977) and Mann and Soli (1991) that this asymmetry can be attributed to the order in which context and target phoneme are presented. In our experiments, speech preceding the target phoneme always is a stronger cue to its identity than speech following the target phoneme. The CV-VC asymmetry was also found when the correlation between correct identification of vowels and consonants was investigated. Correct identification of vowels and consonants was strongly correlated in CV-type tokens. There was no correlation whatsoever between vowel and consonant identification error-rates in VC-type tokens.

An important question remains. Is it actually the original sound that is important for the listener, or would any speech sound do? The latter possibility could be expected when listeners use the preceding sound to "normalize" for the speaker or to give time for the ear to adapt to speech. Our own results do not differentiate between the two possibilities. But earlier work showed that adding just synthetic consonants to synthetic vowels had very little effect on the identification of the vowels (Pols and Van Son, 1993; Van Son, 1993a; Van·Son and Pols, 1993). From the present study with natural speech, and from our earlier study with synthetic speech, it can be concluded that segment identification benefits from the presence of context if this context is appropriate for the segment.

5. Conclusions

Listeners use all speech available to identify vowels and consonants, even when this speech is beyond the transitions to and from a neighbouring phoneme. The presence of speech preceding the target segment benefits identification more than that of speech following the target segment.

6. Acknowledgements

The authors wish to thank dr. Astrid van Wieringen for her advice and help in all stages of the experiments. This research was made possible by grant 300-173-029 of the Dutch Organization of Research (NWO).

7. References

Andruski, J.E. & Nearey, T.M. (1992): 'On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables', Journal of the Acoustical Society of America 91, 390-410.

Benguerel, A.-P. & McFadden, T.U. (**1989**): 'The effect of coarticulation on the role of transitions in vowel perception', *Phonetica* **46**, 80-96.

Cassidy, S. & Harrington, J. (**1995**). 'The place of articulation distinction in voiced oral stops: evidence from burst spectra and formant transitions', *Phonetica* **52**, 263-284.

Cooper, F.S., Delattre, P.C., Liberman, A.M., Borst, J.M. & Gerstman, L.J. (**1952**): 'Some experiments on the perception of synthetic speech sounds', *Journal of the Acoustical Society of America* **24**, 597-606.

Delattre, P.C., Liberman, A.M. & Cooper, F.S. (**1955**): 'Acoustic loci and transitional cues for consonants', *Journal of the Acoustical Society of America* **27**, 769-773.

Di Benedetto, M.G. (**1989**): 'Frequency and time variations of the first formant: Properties relevant to the perception of vowel height', *Journal of the Acoustical Society of America* **86**, 67-77.

Diehl, R.L. & Walsh, M.A. (**1989**): 'An auditory basis for the stimulus-length effect in the perception of stops and glides', *Journal of the Acoustical Society of America* **85**, 2154-2164.

Fox, R.A. (**1989**): 'Dynamic information in the identification and discrimination of vowels', *Phonetica* **46**, 97-116.

Gottfried, T.L. & Strange, W. (**1980**): 'Identification of coarticulated vowels', *Journal of the Acoustical Society of America* **68**, 1626-1635.

Harrington, J. & Cassidy, S. (**1994**). 'Dynamic and target theories of vowel classification: evidence from monophthongs and diphthongs in Australian English', *Language and Speech* **37**, 357-373.

Huang, C.B. (**1991**): 'An acoustic and perceptual study of vowel formant trajectories in American English', Ph.D. Thesis, Massachusetts Institute of Technology, USA (Research Laboratories of Electronics, Technical report no. 563, Cambridge, MA), 203 pp.

Huang, C.B. (**1992**): 'Modelling human vowel identification using aspects of formant trajectory and context' in *Speech perception, production and linguistic structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (Ohmsha, Tokyo; IOS Press, Amsterdam), 43-61.

Jenkins, J.J., Strange, W. & Miranda, S. (**1994**): 'Vowel identification in mixed-speaker silent-center syllables', *Journal of the Acoustical Society of America* **95**, 1030-1043.

Keating, P.A., Lindblom, B., Lubker, J. & Kreiman, J. (**1994**): 'Variability in jaw height for segments in English and Swedish VCVs', *Journal of Phonetics* **22**, 407-422.

Klaassen-Don, L.E.O. (**1983**): 'The influence of vowels on the perception of consonants' Ph.D. Thesis, University of Leiden, The Netherlands.

Kuwabara, H. (**1983**): 'Vowel identification and dichotic fusion of time-varying synthetic speech sounds', *Acustica* **53**, 143-151.

Kuwabara, H. (**1985**): 'An approach to normalization of coarticulation effects for vowels in connected speech', *Journal of the Acoustical Society of America* **77**, 686-694.

Kuwabara, H. (**1993**): 'Temporal effect on the perception of continuous speech and a possible mechanism in the human auditory system', *Proceedings of Eurospeech '93*, Berlin, Germany, 713-716.

Lehiste, I. & Peterson, G.E. (**1961**): 'Transitions, glides, and diphthongs'. *Journal of the Acoustical Society of America* **33**, 268-277.

Lindblom, B. (**1990**): 'Explaining phonetic variation: A sketch of the H&H theory', in: Hardcastle, W.J. & Marshal, A. (eds.), *Speech production and speech modelling*, Kluwer Academic Publishers, Dordrecht, 403-439.

Lindblom, B. & Studdert-Kennedy, M. (**1967**): 'On the role of formant transitions in vowel recognition', *Journal of the Acoustical Society of America* **42**, 830-843.

Lisker, L. (**1986**): 'Voicing in English: A catalogue of acoustic features signalling /b/ versus /p/ in trochees' Haskins Laboratories: Status Report on Speech Research SR-86/87, 45-53.

Macchi, M.J. (**1980**): 'Identification of vowels spoken in isolation versus vowels spoken in consonantal context', *Journal of the Acoustical Society of America* **68**, 1636-1642.

Mack, M. & Blumstein, S.E. (**1983**): 'Further evidence of acoustic invariance in speech production: The stop-glide contrast', *Journal of the Acoustical Society of America* **73**, 1739-1750.

Mann, V. & Soli, S.D. (**1991**): 'Perceptual order and the effect of vocalic context on fricative perception'. *Perception and Psychophysics* **49**, 399-411.

Manuel, S.Y. (**1995**). 'Speakers nasalize /ð/ after /n/, but listeners still hear /ð/', *Journal of Phonetics* **23**, 453-476.

Miller, J.L. & Baer, T. (**1983**): 'Some effects of speaking rate on the production of /b/ and /w/', *Journal of the Acoustical Society of America* **73**, 1751-1755.

Moon, S.Y. & Lindblom, B (**1994**): 'Interaction between duration, context, and speaking style in English, stressed vowels', *Journal of the Acoustical Society of America* **96**, 40-55.

Nearey, T.M. (**1989**): 'Static, dynamic, and relational properties in vowel perception', *Journal of the Acoustical Society of America* **85**, 2088-2113.

Nearey, T.M. & Assmann, P.F. (**1986**): 'Modelling the role of inherent spectral change in vowel identification', *Journal of the Acoustical Society of America* **80**, 1297-1308.

Nossair, Z.B. & Zahorian, S.A. (1991). 'Dynamic spectral shape features as acoustic correlates for initial stop consonants', *Journal of the Acoustical Society of America* **89**, 2978-2991.

Ohde, R. (**1994**): 'The development of cues to the [m]-[n] distinction in CV-syllables', *Journal of the Acoustical Society of America* **96**, 675-686.

Ohde, R.N. & Sharf, D.J. (**1977**). 'Order effect of acoustic segments of VC and CV syllables on stop and vowel identification', *Journal of Speech and Hearing Research* **20**, 543-554.

Öhman, S.E.G. (**1966**): 'Coarticulation in VCV utterances: Spectrographic measurements', *Journal of the Acoustical Society of America* **39**, 151-168.

Öhman, S.E.G. (**1967**): 'Numerical model of coarticulation', *Journal of the Acoustical Society of America* **41**, 310-.

Ohta, F., Yanagihara, N. & Hosoda, I. (**1962**): 'The intelligibility of short Japanese vowels as a function of duration', *Studia Phonologica* **II**, 61-70.

Peeters, W.J.M. (**1991**): 'Diphthong dynamics', Ph.D. Thesis, State University of Utrecht, The Netherlands, 356 pp.

Polka, L. & Strange, W. (**1985**): 'Perceptual equivalence of acoustic cues that differentiate /r/ and /l/', *Journal of the Acoustical Society of America* **78**, 1187-1197.

Pols, L.C.W. (1979). 'Coarticulation and the identification of initial and final plosives', *ASA*50 Speech Communication Papers*, J.J. Wolf and D.H. Klatt (eds.), 459-462.

Pols, L.C.W., Boxelaar, G.W. & Koopmans-van Beinum, F.J. (**1984**): 'Study on the role of formant transitions in vowel recognition using the matching paradigm', *Proceedings of the Institute of Acoustics* **6** (4), 371-379.

Pols, L.C.W. & Schouten, M.E.H. (**1978**). 'Identification of deleted consonants', *Journal of the Acoustical Society of America* **64**, 1333-1337.

Pols, L.C.W. & Van Son, R.J.J.H. (**1993**): 'Acoustics and perception of dynamic vowel segments', *Speech Communication*. **13**, 135-147.

Schouten, M.E.H. & Pols, L.C.W. (**1979**). 'Vowel segments in consonantal contexts: a spectral study of coarticulation-Part I', *Journal of Phonetics* **7**, 1-23.

Schouten, M.E.H. & Pols, L.C.W. (**1979**). 'CV- and VC- transitions: a spectral study of coarticulation-Part II', *Journal of Phonetics* **7**, 205-224.

Schouten, M.E.H. & Pols, L.C.W. (**1981**). 'Consonant loci: a spectral study of coarticulation- Part **III**', *Journal of Phonetics* **9**, 225-231.

Strange, W. (**1989a**): 'Evolving theories of vowel perception', *Journal of the Acoustical Society of America* **85**, 2081-2087.

Strange, W. (**1989b**): 'Dynamic specification of coarticulated vowels spoken in sentence context', *Journal of the Acoustical Society of America* **85**, 2135-2153.

Strange, W. & Gottfried, T.L. (**1980**): 'Task variables in the study of vowel perception', *Journal of the Acoustical Society of America* **68**, 1622-1625.

Strange, W., Jenkins, J.J. & Johnson, T.L. (**1983**): 'Dynamic specification of coarticulated vowels', *Journal of the Acoustical Society of America* **74**, 695-705.

Strange, W., Verbrugge, R.R., Schankweiler, D.P. & Edman, T.R. (**1976**): 'Consonant environment specifies vowel identity', *Journal of the Acoustical Society of America* **60**, 213-224.

Sussman, H.M., McCaffrey, H.A.L. & Matthews, S.A. (**1991**): 'An investigation of locus equations as a source of relational invariance for stop place categorization', *Journal of the Acoustical Society of America* **90**, 1309-1325.

Sussman, H.M., Hoemeke, K.A. & Ahmed, F.S. (**1993**): 'A cross-linguistic investigation of locus equations as a phonetic descriptor of articulation', *Journal of the Acoustical Society of America* **94**, 1256-1268.

Sussman, H.M., Fruchter, D. & Cable, A.. (**1995**): 'Locus equations derived from compensatory articulation', *Journal of the Acoustical Society of America* **97**, 3112-3124.

Tarter, V.C., Kat, D., Samuel, A.G., & Repp, B.H. (**1983**): 'Perception of intervocalic stop consonants: The contribution of closure duration and formant transitions', *Journal of the Acoustical Society of America* **74**, 715-725.

Van der Kamp, L.J.Th. & Pols, L.C.W. (**1971**): 'Perceptual analysis from confusions between vowels', *Acta Psychologica* **35**, 64-77.

Verbrugge, R.R. & Rakerd, B. (**1986**): 'Evidence of talker-independent information for vowels', *Language and Speech* **29**, 39-57.

Van Santen, J.P.H. (**1992**). 'Contextual effects on vowel duration', *Speech Communication* **11**, 513-546.

Van Son, R.J.J.H. (**1993a**): 'Spectro-temporal features of vowel segments', in *Studies in Language and Language use 3*. Ph.D. Thesis, University of Amsterdam, 195 pp.

Van Son, R.J.J.H. (**1993b**): 'Vowel perception: A closer look at the literature'. Proceedings of the Institute of Phonetic Sciences, University of Amsterdam, **17**, 33-64.

Van Son, R.J.J.H. & Pols, L.C.W. (**1993**): 'Vowel identification as influenced by vowel duration and formant track shape', *Proceedings of Eurospeech '93*, Berlin, Germany, 285-288.

Van Son, R.J.J.H. & Pols, L.C.W. (**1995**). "The influence of local context on the identification of vowels and consonants", *Proceedings of Eurospeech 95*, 967-970

Van Wieringen, A. & Pols, L.C.W. (**1991**): 'Transition rate as a cue in the perception of one-formant speech-like synthetic stimuli', *Proceedings of the XIIth International Congress of Phonetic Sciences*, Aix-en-Provence, France, 446-449.

Van Wieringen, A. (1995). 'Perceiving dynamic speechlike sounds, psycho acoustics and speech perception', Ph.D. Thesis, University of Amsterdam, 257 pp.

Van Wieringen, A. & Pols, L.C.W. (**1995**): 'Discrimination of single and complex consonant-vowel- and vowel-consonant-like formant transitions', *Journal of the Acoustical Society of America* **98**, 1304-1312.