The use of the gating paradigm for studying the long - short vowel opposition

Laurens W.M. Damen Louis C.W. Pols

Institute of Phonetic Sciences
University of Amsterdam
The Netherlands

ABSTRACT

A gating task was used to study the long-short vowel oppositions in Dutch. It is shown that the duration of the vowel segment is an important cue to identify the long vowels. Also, an effect of the spectral differences between the vowels of an opposition pair was found.

INTRODUCTION

One of the striking characteristics of speech is its variability; the acoustic fine structure of a given speech segment varies considerably depending on many different factors. One such factor is the rate at which speech is produced. Yet listeners appear to have little difficulty in understanding speech correctly accross a wide variety of speech rates. Several theoretical explanations have been given of this fact. The theory of invariability states that the variability in the speech wave is only a surface phenomenon, and that invariants can be found. On the other hand, the theory of normalization postulates that listeners perceive speech sounds correctly by taking into account the acoustic-phonetic context.

Several studies have shown that the interpretation of speech segments depends on the acoustic-phonetic context. For example, Johnson and Strange (1982) asked subjects to identify vowels in tVt syllables. The syllables were spoken in a neutral carrier sentence at a normal and at a fast speech rate. The subjects heard the syllables in isolation, in a rate-appropriate carrier sentence, or in a rate-inappropriate carrier sentence. Vowels spoken at a normal speech rate were always identified accurately. This was not the case for the vowels in the fast-spoken syllables. The context in which the fast-spoken syllables were presented had a significant effect on the number of errors made. Analysis of the errors revealed that most of the errors were confusions of intrinsically long vowels with their spectrally similar short counterparts.

Van Bergem and Drullman (1985) tried to replicate this effect for the Dutch language. The replication failed; no confusions of long and short vowels were found. A possible explanation for this 'null' result is that the effect of context is totally obscured by the high performance of the listeners.

Both Johnson and Strange (1982) and van Bergem and Drullman (1985) used an identification task in their experiments. This task does not give detailed information about the process of vowel identification. Therefore, we have chosen a different method. The task used in the present experiment employs an extension of the gating paradigm (Grosjean, 1980), somewhat similar to that of Salasoo and Pisoni (1985).

The purpose of our study was to investigate the suitability of the gating task for the study of perceptual confusions between spectrally similar long and short vowels.

METHOD

Subjects

Subjects were 18 students, who were paid for their services. All subjects were native speakers of Dutch, with no reported hearing loss.

Materials and Design

The Dutch language has four spectrally similar long-short vowel pairs. Twelve monosyllabic pairs of nouns were selected for each vowel pair (e.g. /tak/-/tɑk/, /bot/-/bɔt/, /pen/-/pIn/ and /pøl/-/poel/). The words were uttered as the last word of a nonsense sentence. The sentences were comparable to those used by Nakatani and Dukes (1973). In addition, 60 bisyllabic and 20 monosyllabic nouns were selected to serve as filler words. None of the filler words contained any vowel used in the experimental words. The filler words were spoken as the last word of a neutral carrier sentence.

All the sentences were read by a male speaker and recorded on audio tape. The sentences were low-pass filtered at 4.5 kHz, sampled at 10 kHz and stored on computer disk. The experimental and filler words were excised from the stored sentences with the aid of a digital waveform editor and four points were marked in each experimental word. The first point marked the end of the initial consonant(cluster), the second point marked the end of the vocalic (CV) transition, the third point the begin of the final vocalic (VC) transition and finally the last point marked the begin of the final consonant(cluster). Relative to these four markers, additional segment markers for the gates were specified. The excised parts of the word were replaced by envelope-shaped speech noise. This procedure maintained amplitude and durational cues of the word, while removing all spectral cues of the replaced speech segment.

The gate durations of the experimental words are phonetically motivated: The first gate consisted only of the initial consonant(cluster). The initial consonant(cluster) and the CV transition were presented in the second gate. Two periodes of vowel signal were added in the third gate, etc. In the penultimate gate of a word, the complete vowel, including the VC transition, was presented. In this gate only the final consonant(cluster) was replaced by noise. The last gate of a word consisted of the original speech waveform.

It is difficult for the listeners to give a word as a response to gates in which only a very short segment of speech is presented. This was one of the reasons to add filler words. The speech segment in all filler words was at least 60 ms long. It is fairly easy to give a word reponse to such gates in comparison to the many very short speech segments of the experimental words. Thus, we tried to make the task easier for the subjects and we hoped that it would motivate the subjects to give word responses to all the gates presented. The filler words had another important function: it prevented the subjects from recognizing the main purpose of the experiment.

The presentation sequence used in the present experiment differs from that used in other studies that employ the gating paradigm. Words were entered into the presentation sequence in twelve 'steps'; four experimental and four filler words at a step. The order of these eight words was randomized. At the next step eight new words were introduced in the sequence. The set of words in the presentation sequence then totals sixteen. These sixteen words were again randomized. This process was repeated until all words had been introduced. This presentation format is halfway between a successive presentation sequence (e.g. Grosjean,1980) and a single presentation format (e.g. Cotton & Grosjean, 1984).

Six stimulus tapes were prepared from the digitally stored stimuli using a 12-bit D/A converter and a Revox A77 tape recorder. Each tape contained half of the experimental words: two series of twelve words containing only long vowels (e.g. /a/ and /o/) and two series of twelve words containing short vowels (e.g. /I/ and /oe/). The filler words were the same for all six stimulus tapes. Each tape contained about 815 stimuli.

Procedure

Subjects were tested in groups of three in a quiet room. Each group heard one stimulus tape only at a comforting listening level over Sennheiser HD424 earphones. Each experimental word was recorded on three different tapes and was heard by nine subjects. Thus, the design was a between-subjects design.

Subjects were told that they would hear one word at a time and that a large part of the word was made unintelligible by noise.

The subjects were instructed to write down the word they thought they had heard after each presentation of a stimulus. They were required to guess if they were not certain of a particular word. The tape ran without interruption; the interstimulus interval was 4.5 seconds long.

Every page of the response form contained 70 numbered lines, for each response a line. Two cue tones were used to signal the subjects to turn to the next page. A short pause was held after presentation of 280 and 560 stimuli. An entire listening session took about one hour and a half.

RESULTS

Two types of dependent measures were obtained. First, we computed the 'Vowel Isolation Time' (henceforth VIT) for all experimental words. The VIT is defined as the mean amount of vowel signal in ms needed by seven out of nine subjects to guess the identity of the vowel correctly without subsequently changing that guess. Second, we divided all reponses for each word and for every gate into three categories: (a) correctly identified vowels, (b) confusions with the spectrally similar long/short vowel, and (c) all other responses. The data are collapsed over words and over subjects and are given as Cumulative Response Curves.

Vowel Isolation Times (VITs)

The mean VITs for the eight different vowels are depicted in Figure 1.

The VITs were entered in an ANOVA with Long vs Short vowels (/a/,/o/,/e/ and /ø/ vs. /ɑ/,/ɔ/, /I/ and /oe/) and Vowel as fixed effects. Words were considered as the random effect.

A main effect of Long versus Short vowels ($F(1,88) = 169.78$; $p < 0.01$), as well as a main effect of Vowel ($F(3,88) = 7.59$; $p < 0.01$) was found. The interaction of both factors was not significant ($F < 1$).

Cumulative Response Curves

The Cumulative Response Curves of the vowels /o/ and /ɔ/ are given in Figures 2a and 2b. The results of the vowels /e/ and /I/ are plotted in the Figures 3a and 3b. The graphs of the other vowels have the same general appearance but are not presented here, because of lack of space.

Looking at Figures 2a and 3a, it is clear that the long vowels are often confused with their spectrally similar short vowel counterparts, especially at the first four gates. This effect is seen for all four long vowels.

The pattern of results of the short vowels (Figures 2b and 3b) is completely different from that of the long vowels. The short vowels are identified quite accurately, even if only a very short vowel segment is presented. Another interesting point can be seen clearly in these two graphs. Short vowels hardly ever give rise to confusions with their spectrally similar long

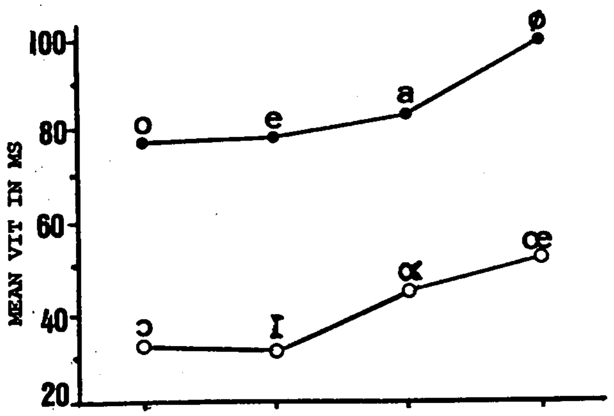

FIG 1. Mean Vowel Isolation Times of the 4 vowel oppostion pairs. Each mean is based on 12 words and 7 ss. ● long vowels; 0 short vowels

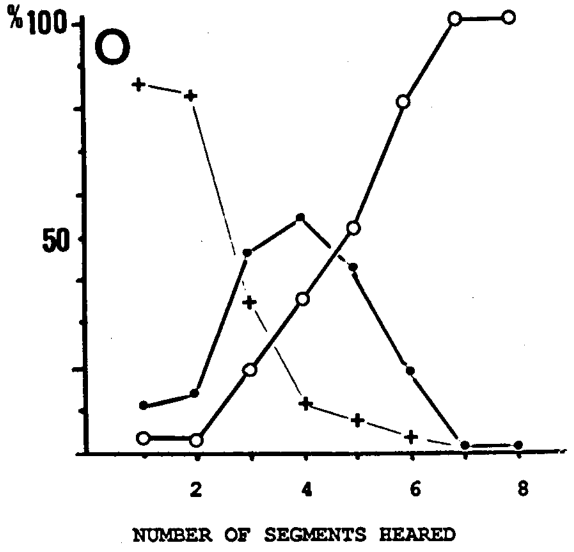

FIG. 2a. Presented vowel: /o/. Response curves: O—O long vowel responses; ●—● short vowel responses; +—+ other responses.

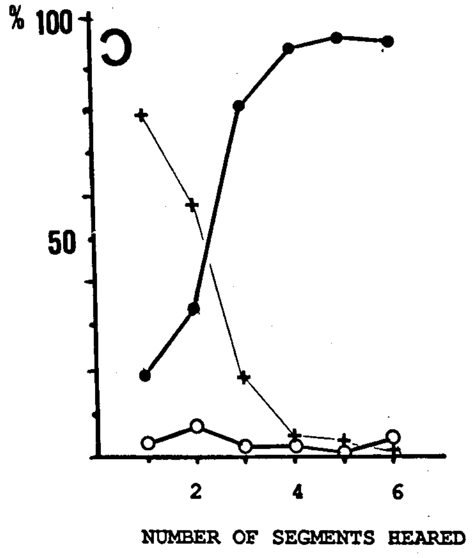

FIG. 2b. Presented vowel: /ɔ/. Response curves: O—O long vowel responses; ●—● short vowel responses; +—+ other responses.

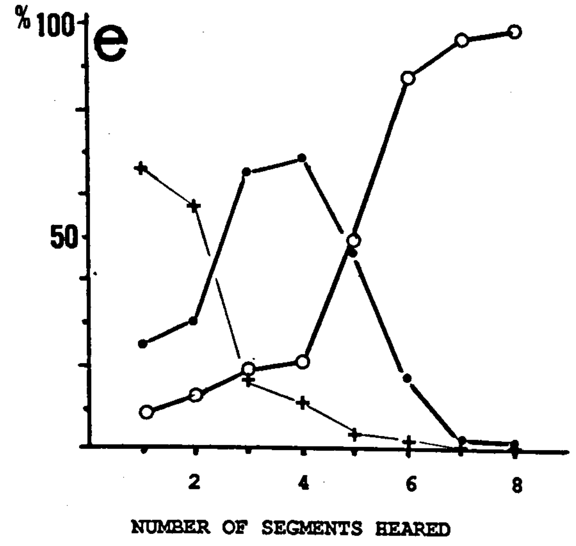

FIG. 3a. Presented vowel: /e/. Response curves: O—O long vowel responses; ●—● short vowel responses; +—+ other responses.

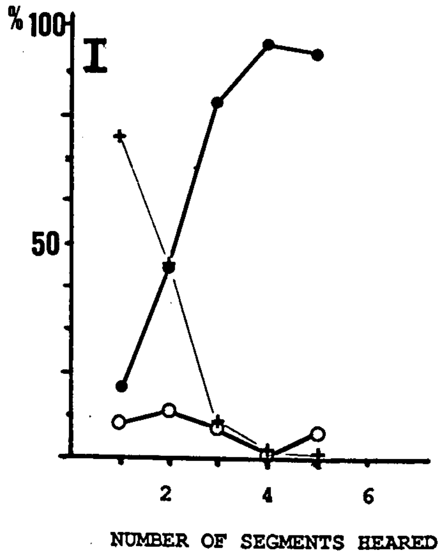

FIG. 3b. Presented vowel: /I/. Response curves: O—O long vowel responses; ●—● short vowel responses; +—+ other responses.

counterparts. This effect is observed for all four vowels, though to a lesser extent for the short vowel /oe/. This finding points strongly to spectral differences between similar short and long vowels. In an attempt to evaluate these spectral differences, we compared the percentage of incorrect short responses to long vowels (e.g. /I/ responses with the vowel /e/ presented) to the percentage of correct short responses to the short vowels (e.g. /I/ responses to the vowel /I/ presented). The data of only the first four gates were analyzed. The percentages of responses were first square-root transformed, as recommended by Winer (1971, p. 399) and subsequently entered into an ANOVA. Long vs. Short vowel and Gate were fixed effects. Vowels were the random effect.

The results are very clear. A main effect of Long vs. Short vowel ($F(1,24) = 27.40$; $p< 0.01$). As expected, there was also a main effect of Gate ($F(3,24) = 44.00$; $p < 0.01$). The interaction between both factors was not significant ($F(3,24) = 1.77$). The outcome of this analysis indicates that the listeners used the fine spectral differences between the long and the short vowel of a pair, and furthermore that they did so independently of segment duration.

Another interesting result that becomes apparent from Figures 2b and 3b is the relatively high percentage of correct vowel responses to the first gate. Only the initial consonant (cluster) was presented in this gate (see Method section). Nevertheless, the identification of the intended vowel is above chance level. The validity of this finding is dependent on the defined boundary between initial consonant(cluster) and vowel. Still, the data indicate that the consonant contains coarticulatory information about the immediately following vowel.

DISCUSSION

This experiment was set up to ask a specific empirical question: can confusions between spectrally similar vowels, differing mainly in duration, fruitfully be studied using a gating task. The results clearly show that this can be done. Both dependent measures (Vowel Isolation Times and percentage correct responses) lead towards the same conclusion, i.e. duration of the presented vowel segment is an important cue to distinguish long vowels. It has also become clear from the analysis on percentage of short vowel responses on the long versus the short vowels, that fine spectral differences between the members of a vowel pair were used by the listeners to identify the vowels. Some caution must be taken with this interpretation, because an alternative explanation can not be ruled out completely. It is likely that the gating task biased the listeners to give more short than long vowel responses. This effect of bias can be expected to be strongest for the shortest speech segments and only the data for these shortest segments were analyzed. Thus, parts of the observed differences in response percentages might be due to bias.

One aspect of the results still needs to be discussed. In the analysis of the Vowel Isolation Times a main effect of Vowel was found. This effect was mainly due to the vowels /ø/ and /oe/. This result can readily be interpreted. These two vowels have a low frequency of occurrence in Dutch and it is likely that this will lengthen the Vowel Isolation Time.

Recently, a follow-up experiment has been carried out to evaluate the presence or absence of noise in the gates. This time the gates were not filled with noise but left silent, all other things being equal. The global pattern of results of this experiment was similar to the results of the experiment reported above. This indicates that the observed effects are robust effects. Furthermore, a small effect of the presence or absence of noise was found. We take this as evidence that the gating task is a very sensitive measuring method.

Presently, an experiment is being carried out, in which acoustic-phonetic context is manipulated in a manner analoguous to the experiment of Johnson and Strange (1982), mentioned in the Introduction of this paper. Fast-spoken words are presented to subjects in a rate-appropriate or in a rate-inappropriate carrier sentence, as well as in isolation. Assuming that the gating task is a sentitive method, it is hypothized that these context manipulations will have a distinct effect on both the Vowel Isolation Times and the response curves.

The advantage of using the gating task in such an experiment is twofold. First, the task provides a detailed picture of the vowel indentification process, as has been shown by the present experiment. It is reasonable to expect that even very small context-induced differences can be found. Furthermore, using the gating task the 'dependent measure' is the moment at which the vowel is identified correctly, and not whether it is identified correctly or not. This allows for the use of high quality speech material, and this reduces the chance that the results of the experiment depend on the specific speaker and speaking situation.

REFERENCES

Bergem,D.R.van,& Drullman,R.(1985) The influence of contextual speech rate on the identification of Dutch vowels in normally and rapidly spoken tVt utterances. IFA-report nr.77 (in Dutch).

Cotton,S.,& Grosjean,F.(1984) The gating paradigm: A comparison of succcessive and individual presentation formats.
Perception and Psychophisics,35,41-48.

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

Johnson,T.L.,& Strange,W.(1982) Perceptual constancy of vowels in rapid speech. Journal of the Acoustical Society of America,72,1761-1770.

Nakatani,L.H.,& Dukes,K.D.(1973) A sensitive test of speech communication quality. Journal of the Acoustical Society of America,53,1083-1091.

Salasoo,A.,& Pisoni,D.B.(1985) Interaction of knowledge sources in spoken word identification Journal of Memory and Language,24,210-231.

Winer,B.J.(1971) Statistical principles in experimental design. New York:McGraw-Hill.