Quarterly Journal of Experimental Psychology (1977) **29**, 483–497.

VOWEL RECOGNITION: INFERENCES FROM STUDIES OF FORWARD AND BACKWARD MASKING

MICHAEL F. DORMAN, DIANE KEWLEY-PORT, SUSAN BRADY AND M. T. TURVEY*

Haskins Laboratories, New Haven, Connecticut, U.S.A.

The recognition of brief vowels was studied in forward and backward masking tasks. In a series of experiments in which both target and mask parameters were systematically varied, two populations of subjects were identified. The majority (Nonmaskers) evidenced little masking at any interstimulus interval, while relatively fewer subjects (Maskers) evidenced backward masking (but not forward masking) over a 100–200 ms interval. Increasing target set size systematically increased masking for the Maskers but not for the Nonmaskers. Adding white noise to the targets increased the extent of masking for both groups. These results suggest (1) that masking does not impose a substantive constraint on vowel perception in running speech, and (2) that multiple strategies may exist for vowel recognition.

Introduction

The experiments reported here view masking in two related contexts (a) as an experimental technique by which stages in the processing of vowels may be examined and (b) as a possible constraint on the perception of vowels in running speech. The several previous studies which have used binaural masking to probe vowel recognition have reported substantially different outcomes. Thus, rather different views have been fostered of the processes underlying vowel recognition and the degree to which masking constrains the recognition of vowels in speech. Massaro (1972) presented listeners 20 ms vowels, either /i/ or /ɪ/, followed at interstimulus intervals from 0 to 500 ms by a 270 ms mask (alternating segments of /a/ and /u/). Vowel recognition was near chance at 0 ms interstimulus interval and reached asymptote by 250 ms. From these data and from studies of backward masking of nonspeech auditory signals, Massaro concluded that perceptual processing time for auditory signals, i.e. the time for formation of a preperceptual auditory image, readout of acoustic features from this image, and matching of acoustic features with a feature list in memory, can last between 120 and 250 ms. Consequently, Massaro (1972) suggested that vowel duration in normal speech

*Michael F. Dorman is also at Arizona State University; Susan Brady and M. T. Turvey at the University of Connecticut; and Diane Kewley-Port at the Graduate Center of the City University of New York. Reprint requests should be sent to M. F. Dorman, 270 Crown Street, New Haven, Connecticut 06510, U.S.A.

may be constrained by the necessity to evade backward masking by following segments of the speech signal.

A somewhat different view of the processes underlying vowel recognition can be inferred from Pisoni's (1972) studies of vowel masking. Listeners were presented computer-generated 40 ms vowels /i, ɪ, ɛ/ followed by another vowel from the same set. Performance at 0 ms interstimulus interval was 85% correct and reached asymptote by 80 ms. The minimal backward masking obtained suggests the operation of rather efficient perceptual processing routines which require minimum time and information for vowel recognition.

That Massaro's and Pisoni's studies had different outcomes is not surprising given the differences in the nature and specification of the target and mask signals. The present series of studies was designed to further examine the processes underlying vowel recognition by systematically varying target and mask parameters in forward and backward masking experiments.

Experiment I

The purpose of the first and succeeding experiments was to determine whether vowels, which were of sufficient fidelity to be identifiable in isolation, could be masked. A long series of informal listening tests suggested that only very brief vowels suffered backward masking. Therefore, to determine the vowel durations over which masking occurred vowel sets of three durations [15·5, 20 and 30 ms] were constructed and presented to listeners in both forward and backward masking tasks.

Method

Subjects

The subjects were undergraduate students from Yale University and the University of Connecticut. Yale University students received $2.00 per hour for participation; University of Connecticut students received class credit.

Preparation of stimuli

The target stimuli were the vowels /i/, /ɛ/ and /ʌ/ (the vowels in beet, bet and but respectively) spoken in isolation by a male with fundamental frequency of approximately 120 Hz. Using the Haskins Laboratories computer-controlled pulse code modulation system (Cooper and Mattingly, 1969) three sets of vowels were prepared. Segments of 15·5, 20, and 30 ms duration were excised from steady-state portions of each vowel. The mask was a computer-synthesized two-formant sound of 125 ms duration with formant frequencies at 489 and 1690 Hz. This mask was vowel-like but did not have formant frequencies similar to any English vowel. The target and mask stimuli were equated for peak-to-peak amplitude. Special care was exercised to insure that the stimuli were recorded at the best possible signal-to-noise ratio, approximately 40 dB.

Training materials

Under computer control one sequence of three repetitions of the target vowels and six 18-item sequences of target vowels (six repetitions of each vowel in each randomized sequence) were recorded on audio tape. The intertrial interval was 4 s for all sequences.

Test materials

Six test sequences were constructed. For each vowel set duration (15·5, 20 and 30 ms) both a forward and backward masking sequence were generated. (A six-item practice sequence was also generated for each test sequence). In the forward masking condition the mask preceded the target vowels at offset-to-onset intervals of 0, 25, 50, 100, 200 and 500 ms. Each vowel occurred six times at each interstimulus interval. In the backward masking condition the vowels preceded the mask at intervals of 0, 25, 50, 100, 200 and 500 ms. Each vowel occurred six times at each interstimulus interval. The sequence of vowels and interstimulus intervals were randomized in each test sequence. Each test sequence was presented twice, thus creating two blocks of 54 trials, or a test sequence of 108 items.

Apparatus

The stimuli were recorded and reproduced on an Ampex AG500 tape recorder. The tape recorder output was interfaced with a distribution amplifier which insured equal signal amplification into four sets of matched Granson-Stadler TDH39-300Z earphones. The stimuli were presented binaurally at a comfortable listening level. A calibration signal insured equal signal levels in all conditions within and between experiments.

Design

One group of subjects was trained and tested with 15·5 ms vowels, another group with 20 ms vowels, and a third group with 30 ms vowels. All subjects were tested on both the forward and backward masking sequences in counterbalanced order.

Procedure

The subjects were seated in a large sound-attenuated room and were told they would hear three very brief vowels, /i/, /ɛ/ and /ʌ/, which they were to learn to identify. First, the subjects were presented three repetitions of the three-vowel set. Next, the subjects were told they would hear six 18-item lists of the vowels in random order and were instructed to write the identity of the vowels (as ee, eh, uh) on printed response sheets. The correct responses, initially covered by a movable slider, were printed next to the space for the subjects responses. By moving the slider down the page for each succeeding trial the subjects uncovered the correct responses for the preceding trial, thus providing immediate feedback of correct responses. On the final 18-item sequences, the subjects were given no feedback of correct responses.

After a brief rest period, the subjects were told they would hear, in one sequence, the vowels followed by a mask at various intervals, and in another sequence, the mask followed by the vowels at various intervals. The subjects were instructed to write the identity of the vowels on a printed answer sheet. After six practice trials the subjects were presented a 108-item test sequence in two blocks of 54 trials. Then, after a brief rest, the subjects were given another six practice trials and the other 108-item test sequence.

Results

Since our interest was whether vowels which could be identified in isolation could be masked, we consider first only those subjects who made no errors on the final practice sequence. In the backward masking task, although the majority of subjects made few or no errors, several subjects made errors at 0 ms interstimulus interval and at intervals out to 200 ms. Averaging these two types of performance would obscure the outcome that most subjects showed no masking while others behaved in a manner indicating some masking. To reflect these different types of performance the subjects were characterized as Nonmaskers or Maskers on the basis of error patterns. One pattern was characterized by better than 80% correct responses at 0 ms interstimulus interval and very few errors at other interstimulus intervals. Subjects who performed in this manner will be referred to as Nonmaskers. Another error pattern was characterized by scores of less than 80% correct at 0 ms interstimulus interval while not reaching asymptote until 100–200 ms interstimulus interval. Subjects who performed in this manner will be referred to as Maskers. (This definition of Nonmaskers and Maskers will be used in all of the following experiments.)

The results of the backward masking task with the 15·5, 20 and 30 ms vowels are shown in Figure 1. None of the subjects tested with the 30 ms stimuli were classified as Maskers.

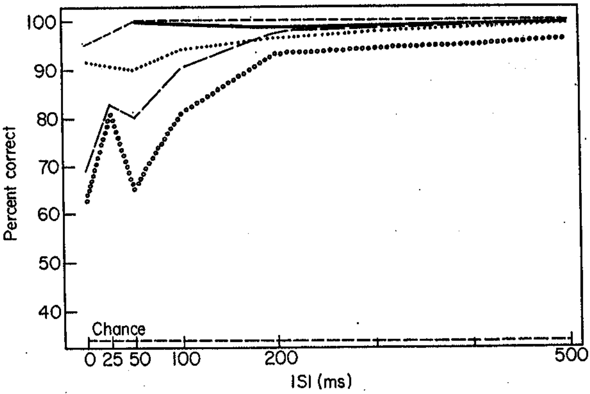

FIGURE 1. Percent correct vowel identification as a function of vowel length. Nonmaskers: vowel length 30 ms, --- ($n = 10$); 20 ms, —— ($n = 7$); 15·5 ms, ------ ($n = 6$). Maskers: vowel length 20 ms, —— ($n = 3$); 15·5 ms, ººººº ($n = 4$).

Seven of the 10 subjects in the 20 ms condition were classified as Nonmaskers and three as Maskers on the test sequence. Six of the 10 subjects in the 15·5 ms condition were classified as Nonmaskers and four as Maskers. (The depressed score for the Maskers at 50 ms interval reflected the poor performance of only one subject.)

Eight subjects who made errors (average performance = 79% correct) on the final training sequence with 15·5 ms vowels were also tested with the backward masking sequences. For these subjects performance improved slightly with an increase in interstimulus interval. At 0 ms interval target recognition was 55% correct; at 500 ms interval, 66% correct.

In the forward masking condition all of the groups achieved 97% or better correct responses at all of the interstimulus intervals.

Discussion

No forward masking was observed in any of the conditions. Since the 15·5 ms vowels were of minimum duration, it appears that perceptual interference in the recognition of vowels occurs only when a masking stimulus follows a target stimulus.

For subjects who could identify the vowels in isolation, in the backward masking task a 30 ms vowel duration allowed essentially perfect recognition of the vowel targets. In the 20 ms vowel condition, the majority of subjects showed very little or no impairment in vowel recognition. Even in the 15·5 ms vowel condition, even though the signals contained only one complete pitch period, the majority of subjects performed at better than 90% accuracy at 0 ms interval. From these data we conclude that for the majority of subjects, a stimulus duration of between 20 and 30 ms is sufficient for processing mechanisms to separate the targets from the mask and to extract the information necessary for the recognition of the vowels.

There are several possible explanations for the observed backward masking. Following Kahneman (1968) and Turvey (1973), we can initially distinguish between peripheral (sensory) and central loci of interference and between two mechanisms of interference—peripheral integration of target and mask, and central interruption of target categorization by the mask. The integration hypothesis assumes that the target and mask interact peripherally, thus presenting a "noisy" representation to the central processor for recognition. This hypothesis predicts interference in target recognition for both forward and backward masking paradigms.

The interruption hypothesis suggests that a clear target representation arrives at the central processors but categorization of the target representation is disrupted by the arrival of the mask before recognition is achieved. This hypothesis predicts severe impairment of target recognition only in backward masking paradigms. Since only backward masking was observed for the Maskers, we infer that the locus of interference was central and resulted from the mask disrupting the categorization of the target representation.

The long interstimulus interval necessary to evade masking for the Maskers in the 15·5 and 20 ms vowel conditions is in marked contrast to the essentially perfect performance of all subjects at 0 ms interstimulus interval in the 30 ms vowel condition. Although collected from different groups of subjects, these data suggest that silent processing time and stimulus duration do not have additive effects in determining vowel recognition. Allowing a 25 or 50 ms silent processing interval after the 15 and 20 ms targets was not equivalent in terms of facilitating vowel recognition to adding 15 and 10 ms stimulus duration. The recognition of brief vowels thus appears more constrained by the fidelity of the information afforded by the stimulus display than by the necessity for silent processing time.

Experiment II

Experiment I investigated the effect of several target parameters on vowel masking. Experiment II varied two mask parameters. In visual backward masking, impairment of target recognition varies as a function of the similarity of target and mask features (Schiller, 1965; Kahneman, 1968). Since the mask of Experiment I was a synthesized two-formant vowel-like stimulus, it is possible that a mask that shared more features with the target vowels would produce more interference with vowel recognition. To investigate this, in one condition of Experiment II the mask was a 125 ms vowel /o/. Turvey (1973) has shown for central backward masking in vision that an increase in mask energy beyond that of target energy does not increase the extent of backward masking. To determine whether mask energy affects vowel recognition, in another condition of Experiment II the two-formant mask of Experiment I was increased in intensity 20 dB.

Method

Subjects

The subjects were undergraduate students from Yale University and the University of Connecticut. Yale University students received $2.00 per hour for participation; University of Connecticut students received class credit.

Preparation of stimuli

The target vowels were the 20 ms vowels used in Experiment I. Two masks were constructed. One mask was the vowel /o/, spoken by the same male speaker as in Experiment I, truncated to 125 ms duration, and equated for peak-to-peak amplitude with the target vowels. A second mask was the two-formant mask of Experiment I, but made 20 dB (true root mean square) more intense than the mask of Experiment I.

Test materials

Backward and forward masking sequences were generated with both the /o/ mask and the +20 dB two-formant mask. The internal construction of the test sequences was the same as in Experiment I.

Design

One group of subjects was tested with the /o/ mask in both the forward and backward masking conditions in counterbalanced order. Another group of subjects was tested with the +20 dB mask in the forward and backward masking conditions in counterbalanced order.

The *Training materials, Apparatus,* and *Procedure* were the same as in Experiment I.

Results

The results for the Maskers in the /o/ mask and +20 dB mask conditions and the results from the Maskers in the 20 ms vowel condition of Experiment I (+0 dB mask) are shown in Figure 2. Five of the 10 subjects in the /o/ mask condition

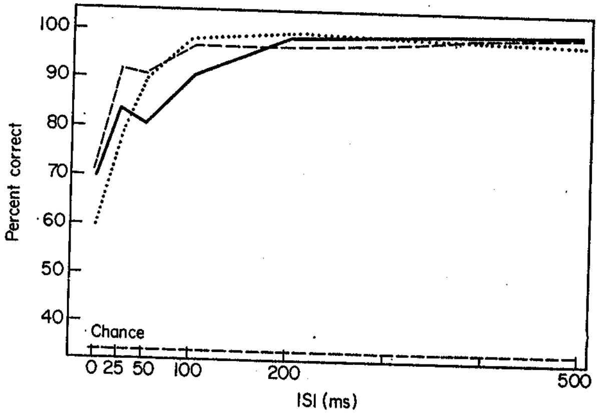

FIGURE 2. Percent correct vowel identification by Maskers as a function of mask type. 0 dB mask —— ($n = 3$); + 20 dB mask --- ($n = 2$); /o/ mask ------ ($n = 5$).

were characterized as Nonmaskers, and five as Maskers. Eight subjects in the +20 dB mask condition were characterized as Nonmaskers, and two as Maskers.

In the forward masking conditions both groups of subjects achieved 92% or better correct responses at all of the interstimulus intervals.

Discussion

The absence of forward masking in either the /o/ mask or +20 dB mask condition reinforces the impression gained from Experiment I that perceptual interference in the recognition of vowels occurs only when a mask follows a target stimulus.

In the backward masking sequences, the performance of the subjects did not differ greatly from that of the subjects in the 20 ms vowel two-formant mask condition of Experiment I. One group of subjects was characterized as Nonmaskers and a smaller group as Maskers. The difference in performance at 0 ms interval was quite marked in the /o/ mask condition. Nonmaskers averaged 98% correct while the Maskers averaged 58% correct. This difference in performance again argued for not averaging the data from the two subject groups. Increasing mask energy did not increase the number of Maskers or increase the interstimulus interval necessary to evade masking. The /o/ mask did appear to be somewhat more effective than either of the two-formant masks in terms of the number of Maskers, and in terms of percent correct vowel recognition at 0 ms interstimulus interval. However, there was no significant different in performance at 0 ms interstimulus interval as a function of the mask type. Viewed as a whole, the results from Experiments I and II, i.e. the complete absence of forward masking and the absence of an increase in backward masking with a large increase in mask energy, reinforce the conclusion that the locus of interference for the Maskers was of central rather than peripheral or sensory origin (cf. Turvey, 1973, p. 36). The data do not reveal, however, the nature of the difference in perceptual processing for the Nonmasker and Masker populations.

Experiment III

A fairly common view of stimulus recognition is that it consists of an initial operation of encoding a stimulus as a set of abstract features, and a second operation of comparing (through a matching of features or analysis-by-synthesis) the abstracted stimulus representation with a set of stored features in long-term memory. In this broad view of stimulus recognition, increasing the size of the target set in a backward masking task should increase the latency of target categorization (cf. Sternberg, 1967; Massaro, 1974) and should, therefore, increase the susceptibility of the recognition process to interference from a masking stimulus.

In order to probe possible differences in vowel recognition strategy used by the Nonmaskers and Maskers, in Experiment III vowels from sets of two, three, or four vowels were presented to listeners in a backward masking task.

Method

Preparation of stimuli

The target vowels were the 20 ms vowels of experiment with the addition of /ɪ/. The mask was the two-formant stimulus of Experiment I.

Training materials

One sequence of three repetitions of the vowel set [i ɪ ɛ ʌ] and five 24-item sequences of vowels (six repetitions of each vowel in each randomized sequence) were recorded on audio tape.

Test materials

Backward masking sequences were generated for the two-, three- and four-vowel target sets using the same interstimulus intervals as Experiment I.

Design

Each subject was tested on the two-, three- and four-vowel backward masking sequences. One group was tested in the order 2, 3, 4, another group 3, 4, 2, and a third group 4, 2, 3.

Procedure

The training procedures were similar to that used in Experiments I and II, modified for four target vowels.

Results

Of the subjects trained, 72% achieved perfect performance on the final practice sequence. The subjects were classified as Maskers and Nonmaskers on the basis of performance on the three-vowel target set. Six subjects (two from each test order) were classified as Maskers, and 12 as Nonmaskers. The averaged performance of the Nonmaskers as a function of target set size is shown in Figure 3.

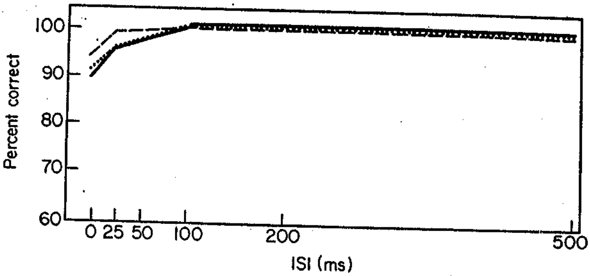

FIGURE 3. Percent correct vowel identification for Nonmaskers as a function of target set size. 2 vowels, ------; 3 vowels, ---; 4 vowels, ——. $n = 12$.

The performance of the Maskers is shown in Figure 4.

The interstimulus interval at which target recognition reached asymptote was determined by comparing performance at 500 ms interstimulus interval with performance at the shorter interstimulus intervals. For the Maskers on the two-vowel set the first point of difference from 500 ms interstimulus interval was at 25 ms interstimulus interval ($t(5) = 2{\cdot}78$, $P < 0{\cdot}05$); for the three-vowel set, 100 ms interstimulus interval ($t(5) = 4{\cdot}11$, $P < 0{\cdot}02$); and for the four-vowel set, at 100 ms interstimulus interval ($t(5) = 5{\cdot}46$, $P < 0{\cdot}01$). For the Nonmaskers, on the two-vowel set, performance differed from that at 500 ms interstimulus interval only at 0 ms interstimulus interval ($t(11) = 2{\cdot}92$, $P < 0{\cdot}02$); for the three-vowel set, at no point; and for the four-vowel set, at 0 ms interstimulus interval ($t(11) = 4{\cdot}59$, $P < 0{\cdot}01$).

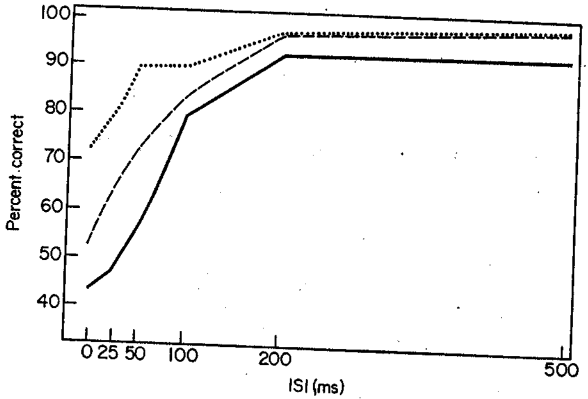

FIGURE 4. Percent correct vowel identification for Maskers as a function of target set size. 2 vowels, ------; 3 vowels, ---; 4 vowels, ——. $n = 6$.

Discussion

Two listener populations were again identified in the three-vowel condition. For the Nonmaskers increasing target set size from two to four vowels did not systematically increase backward masking (i.e. performance on the two-vowel set was similar to that on the four-vowel set). For the Maskers, however, performance was systematically affected as a function of target set size. In terms of level of performance at brief interstimulus intervals, and in the extent of masking, performance in the two-vowel condition was better than in the four-vowel condition. The absence of a difference in extent of masking for the three- and four-vowel conditions may have been a function of the few data points between 100 and 500 ms interstimulus interval.

We note that the number of vowels in the target sets was at least partially confounded by acoustic similarity between the vowels (e.g. the acoustic difference between /ɛ/–/ʌ/ is greater than that between each member of the set /i/–/ɪ/–/ɛ/). However, it is interesting that the Nonmaskers' performance was unaffected by increasing the number of items in the target set even when the task was complicated by increased acoustic similarity among the items to be recognized. For Maskers, target set size and acoustic similarity between targets could exert independent effects on vowel recognition (cf. Darwin & Baddeley, 1974; Allen and Haggard, 1974.) This remains to be determined.

Experiment IV

In Experiments I–III the Nonmaskers were apparently able to determine the identity of the target vowels before the mask disrupted the recognition process.

With this interpretation of the Nonmaskers' performance, a change in stimulus parameters that delays or retards the recognition process should make the Nonmaskers more susceptible to backward masking. Degrading the vowel stimuli with white noise could increase the amount of backward masking, as Sternberg (1967) has argued that adding noise to a visual stimulus in a character recognition task increases the latency of encoding a stimulus as a set of abstract features.

Method

Subjects

The subjects were undergraduate students at Yale University who were paid $2.00 per hour for participation.

Preparation of stimuli

For one condition (control) the target stimuli were the 20 ms vowels and 125 ms two-formant mask of Experiment I. For a second condition (noise-added) the target stimuli were constructed by adding white noise 15 dB less intense than the vowels to the 20 ms vowels of Experiment I. (Extensive pilot experiments indicated that when greater amounts of noise were added, most naive subjects were not able to identify the vowels). White noise was similarly added to the mask.

Preparation of test sequences

Backward masking sequences were generated for both the control and noise-added conditions. The internal constraints on the test sequence constructions were the same as in Experiment I. Different stimulus randomizations were used in the two test sequences.

Design

Each subject was tested in both the noise-added and control condition. Test order was counterbalanced across subjects.

Procedure

The training procedure was similar to that used in the previous experiments. The subjects were first given three practice sequences with the control vowels, then three practice sequences with the noise-added vowels. Finally, the subjects were given identification tests separately for the control and noise-added vowels.

Results

Only those listeners who made no errors on both the final control and noise-added identification tests were considered in the data analyses. Of the subjects trained, 83% achieved perfect performance on the final practice list. The listeners were classified as Nonmaskers and Maskers on the basis of performance in the control condition. In the control condition, ten listeners were classified as Nonmaskers, and two as Maskers. Of the ten Nonmaskers six made more errors in the noise-added condition (Fig. 5). For these subjects, performance in the noise-added condition differed from that in the control condition at 0 ms interstimulus interval ($t(5) = 4{\cdot}30$, $P < 0{\cdot}01$), at 25 ms interstimulus interval ($t(5) = 4{\cdot}38$, $P < 0{\cdot}01$); and at 50 ms interstimulus interval ($t(5) = 2{\cdot}71$, $P < 0{\cdot}05$). Four listeners made no errors in either condition. One of these subjects was tested further. No masking was found with −12 and −9 dB noise-added stimuli.

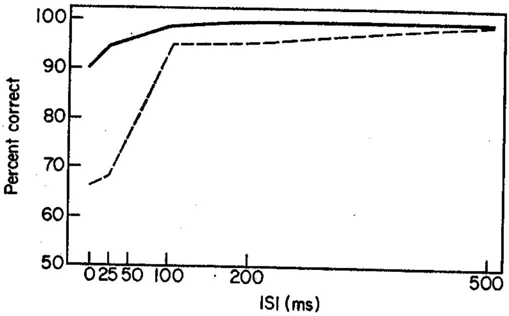

FIGURE 5. Percent correct vowel identification for Nonmaskers in the control (——) and noise-added (------) conditions. $n = 6$.

Only in a −3 dB noise-added condition was masking apparent. The two Maskers made more errors in the noise-added condition than in the control condition.

Discussion

For the majority of listeners who showed no masking of 20 ms vowels at 0 ms interstimulus interval, the addition of −15 dB white noise to the targets results in backward masking over an interval of 50–100 ms. The noise in the stimuli could be viewed as increasing the time necessary to extract and encode the vowel features. This increased processing time would, in turn, increase the vulnerability of the recognition routine to disruption by the stimulus arriving second.

The performance of some of the listeners in the noise-added condition was similar to that reported by Massaro (1972) for a two-vowel discrimination. Massaro found that performance at 0 ms interstimulus interval was near chance and reached asymptote by 250 ms interstimulus interval. He interpreted these data as providing evidence that the 150–350 ms vowel duration commonly found in running speech may be necessary to evade backward masking from following segments of the speech signal. However, if Massaro's masking functions were influenced by variables functionally similar to the noise condition of Experiment IV, then such an interpretation would be unduly pessimistic. In the following section we review other data on vowel masking and conclude that vowel perception is, in fact, relatively unconstrained by masking from following components of speech.

General discussion

It has been proposed that vowel length in running speech may be constrained by the necessity to evade masking by a following speech segment (Massaro, 1972). Of the several types of evidence offered in support of this proposal, two are particularly relevant to the present experiments (1) vowel masking experiments (2) temporal order experiments with vowel sequences.

Neither the masking experiments reported here, nor Pisoni's earlier studies, lend support to Massaro's position. Indeed, the experiments point to the difficulty of impairing vowel recognition in masking tasks. As long as vowels were identifiable in isolation, then despite reduction of signal duration, variation in mask amplitude and similarity to the signals, increases in signal set size and addition of noise to the signals, most listeners evidenced little or no backward masking. Forward masking was not observed in any condition. The absence of masking in these experiments appears to reflect the operation of very efficient perceptual processing routines which require minimal specification of features and processing time for vowel recognition. Thus, even in situations which are very much unlike speech, i.e. vowel sequences presented without the naturally occurring formant transitions, masking does not appear to impose a substantive constraint on vowel recognition.

The marked absence of backward recognition masking for vowels obtained in the present series of experiments is consistent with the absence of backward recognition masking for tones in similar experimental conditions. When stimulus certainty is maximized, e.g. when well trained observers attend to a target and mask presented to the same ear, frequency resolution for 20 ms tones is generally unimpaired by a following mask (Leshowitz and Cudahy, 1973; Loeb and Holding, 1975). This absence of masking extends, in the instance of vowels, to most relatively unpracticed observers. Moreover, even in the situations in which some unpracticed observers showed backward masking, highly practiced observers (the authors) evidenced "none". We conclude from these data that the 250 ms interval and attendant cognitive operations envisioned by Massaro as necessary for perceptual processing are certainly not requisite conditions for either tone or vowel recognition.

Stimulus brevity does constrain vowel recognition, although not by imposing a limitation on processing time. Experiments on temporal ordering of vowels illustrate this point. Several studies (Warren and Warren, 1970; Thomas, Hill, Carol and Garcia, 1970) have indicated that when a series of vowels are concatenated, i.e. placed offset-to-onset without the naturally occurring formant transitions, each vowel must be greater than 125 ms in duration for listeners to achieve accurate temporal ordering. Massaro has interpreted these data as supporting his arguments for a minimum processing time of 125–250 ms for vowel recognition. Dorman, Cutting and Raphael (1974) found, however, that difficulty in ordering concatenated vowel sequences arose primarily from auditory stream segregation, i.e. vowels with similar formant structure formed separate auditory streams. Moreover, streaming could be reduced by connecting the vowels in a manner like that found in natural speech—that is connected by formant transitions. In vowel sequences connected by formant transitions, temporal ordering was constrained by the minimum stimulus duration which afforded recognition of a vowel plus vowel sequence, e.g. /i/ followed by /a/, rather than another speech segment such as /ya/. Thus, stimulus brevity constrains temporal ordering of speech, not by limiting processing time, but by delimiting the minimum stimulus duration which specifies a given phoneme or series of phonemes *vis à vis* other speech sounds. In sum, the results of our masking and temporal ordering experiments lead us to conclude that masking does not pose a substantive constraint on the perception of vowels in running speech.

Indeed, the portions of the acoustic signal which precede and follow a vowel appear, in at least some contexts, to increase vowel intelligibility, rather than impair recognition by forward or backward masking. For example, in a recent series of experiments, Shankweiler, Strange & Verbrugge (in press) have reported a 43% error rate for the identification of vowels presented in isolation. However, when the vowels are embedded in a consonant-vowel-consonant context, the error rate fell to 17%. Formant transitions are thus seen to carry information in parallel about both vowel and consonant identity (cf. Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967). Thus, the acoustic signals which precede and follow a vowel in natural speech constrain vowel recognition, not by forward or backward masking, but by providing, in different degrees, information specifying vowel identity.

Perceptual processing of vowels

Although the majority of subjects showed no backward masking, a relatively few suffered substantial masking [LaRivere, Winitz and Harriman (1975) have also reported two subject populations, similar to Nonmaskers and Maskers, in a backward masking task with brief vowels]. It is important then that the usual procedure of averaging the scores of all the subjects in a given experimental condition was not followed in the present experiments. The outcome for such a procedure would seriously misrepresent the obtained results by indicating that the listeners, in general, evidenced some masking. This, of course, was not the outcome. Most listeners evidenced no backward masking, while only a few showed masking.

How shall we characterize the difference in perceptual processing between Nonmaskers and Maskers? One common approach would be to classify the differences simply as individual differences. On this view we would conclude that Nonmaskers and Maskers used similar strategies for vowel recognition. Maskers might then be characterized as slower information processors, or perhaps as having a poorer auditory memory.

Another possibility is that a listener could employ different processing strategies for vowel recognition depending on the fidelity of stimulus information in a particular experimental task. Subjects may differ on the auditory information necessary for the use of one or the other strategy.

The suggestion of alternative processing strategies for speech recognition is not novel. For example, Morton and Broadbent (1967) have speculated that when processing speech a phonetic representation of the signal may be derived directly from the auditory signal. However, under certain conditions, e.g. when speech sounds are presented at a "low signal to noise ratio" a different process may be used, namely, that of internally generating trial sequences of phonemes for comparison with the input signal. In short, there are possibly two quite different modes for processing the sounds of speech with their differential use dependent upon the information fidelity or adequacy of the auditory signal. In instances of good fidelity the pickup of relational information specifying the speech sound may be relatively direct (cf. Gibson, 1966); in instances of poor fidelity the perception of speech sounds may have to proceed by an indirect constructive or inferential route (cf. Neisser, 1967).

The outcomes of Experiments I–IV do not dictate a choice between the individual difference interpretation and the multiple strategies interpretation of the Nonmaskers' and Maskers' performance. However, the outcome of dichotic listening studies with vowels provides a line of evidence which bears on the different interpretations of the Nonmaskers' and Maskers' performance. Steady-state vowels usually evidence either no ear-advantage or a left-ear advantage (Shankweiler and Studdert-Kennedy, 1967). However, very brief vowels (Godfrey, 1974) and vowels in noise (Weiss and House, 1973) evidence a right-ear advantage. In the present experiments when vowel duration was reduced and when noise was added to the signals, performance changed from little or no masking, to masking extending over at least a 100 ms interval. To the extent that left- and right-ear advantages reflect different modes of processing (cf. Semmes, 1968), the results of the dichotic listening experiments support the possibility of multiple modes of vowel recognition.

This research was supported by NICHD Grant HD-10994 to Haskins Laboratories and a City University of New York Faculty Research Award to the first author.

References

Allen, J. and Haggard, M. (1974). Dichotic backward masking of acoustically similar vowels. The Queen's University of Belfast. *Speech Perception,* **2,** 35–40.

Cooper, F. and Mattingly, I. (1969). Computer-controlled PCM system for investigation of dichotic speech perception. *Journal of the Acoustical Society of America,* **46,** 115 (Abstract).

Darwin, C. J. and Baddeley, A. D. (1974). Acoustic memory and the perception of speech. *Cognitive Psychology,* **6,** 41–60.

Dorman, M., Cutting, J. and Raphael, L. (1975). Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance,* **2,** 121–9.

Gibson, J. (1966). *The Senses Considered as Perceptual Systems.* Boston: Houghton Mifflin.

Godfrey, J. J. (1974). Perceptual difficulty and the right-ear advantage for vowels. *Brain and Language,* **4,** 323–36.

Kahneman, D. (1968). Method, findings and theory in studies of visual masking. *Psychological Bulletin,* **70,** 404–25.

LaRiviere, C., Winitz, H. and Herriman, E. (1975). Vocalic transitions in the perception of voiceless initial stops. *Journal of the Acoustical Society of America,* **57,** 470–5.

Leshowitz, B. and Cudahy, E. (1973). Frequency discrimination in the presence of another tone. *Journal of the Acoustical Society of America,* **54,** 882–7.

Liberman, A., Cooper, F., Shankweiler, D. and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review,* **74,** 431–61.

Loeb, M. and Holding, D. (1975). Backward interference by tones or noise in pitch perception as a function of practice. *Perception and Psychophysics,* **18,** 205–8.

Massaro, D. (1972). Preperceptual images, processing time and perceptual units in auditory perception. *Psychological Review,* **79,** 124–5.

Massaro, D. (1974). Perceptual units in speech recognition. *Journal of Experimental Psychology,* **2,** 199–208.

Morton, J. and Broadbent, D. (1967). Passive versus active recognition models, or is your homunculus really necessary? In Wathen-Dunn, W. (Ed.), *Models for the Perception of Speech and Visual Form.* Cambridge, Mass.: M.I.T. Press.

Neisser, U. (1967). *Cognitive Psychology.* New York: Appleton-Century-Crofts.

Pisoni, D. (1972). Perceptual processing time for consonants and vowels. Haskins Laboratories, *Status Report on Speech Research, SR*-31/32, 83–9.

SEMMES, J. (1968). Hemispheric specialization: A possible clue to mechanism. *Neuropsychologia*, **6**, 11–26.

SCHILLER, P. H. (1965). Monoptic and dichoptic masking by patterns and flashes. *Journal of Experimental Psychology*, **69**, 193–99.

SHANKWEILER, D., STRANGE, W. and VERBRUGGE, R. (in press). Speech and the problem of perceptual constancy. In SHAW, R. and BRANSFORD J. (Eds), *Perceiving, Acting, and Comprehending: Toward an Ecological Psychology*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

STERNBERG, S. (1967). Two operations in character recognition: some evidence from reaction-time measurements. *Perception and Psychophysics*, **2**, 45–53.

THOMAS, I., HILL, P., CARROL, F. and GARCIA, D. (1970). Temporal order in the perception of vowels. *Journal of the Acoustical Society of America*, **48**, 1010–13.

TURVEY, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, **80**, 1–52.

WARREN, R. R. and WARREN, R. P. (1970). Auditory illusions and confusions. *Scientific American*, **233**, 30–36.

WEISS, M. and HOUSE, A. (1973). Perception of dichotically presented vowels. *Journal of the Acoustical Society of America*, **53**, 51–58.

Revised manuscript received 31 *August* 1976