TEMPORAL COURSE OF PERCEIVED VOWEL DURATION

DOMINIC W. MASSARO *and* WENDY L. IDSON

University of Wisconsin, Madison

The current study explored the temporal course of the perception of vowel duration. A backward recognition masking paradigm was employed in which the subject was to determine which of two target vowels, differing only in duration, was presented on a given trial. The target was followed, after a variable silent intervowel interval, by a masking vowel having one of three possible durations. Subjects were instructed either to identify the long vowel as /i/ and the short vowel as /ɪ/, or were told that the targets were nonsense sounds they were to identify as long or short. For both instruction conditions, identification of the long target improved consistently with increases in the intervowel interval, whereas identification of the short target was equally accurate at the long and the short intervals. Identification of the long target vowel was most accurate when followed by a long masking vowel, whereas identification of the short target vowel was most accurate when followed by a short masking vowel. An information-processing model provided a good description of the quantitative results.

One of the critical cues for the perception of speech is the duration of the component segments of the speech signal. At both segmental and suprasegmental levels, duration provides essential information for analysis of a spoken utterance (Klatt, 1976; Lehiste, 1970). At the segmental level, for example, vowel and frication duration are cues to voicing of a postvocalic consonant (Denes, 1955; Massaro and Cohen, 1977; Raphael, 1972). In terms of the suprasegmental level, word and phrase duration influence such factors as the placement of a constituent boundary of an otherwise ambiguous phrase (Lehiste, 1973). For example, the phrase *old men and women* is more likely to be interpreted as (*old men*)(*and women*) if the duration of the phrase *men and women* is long relative to the case in which it is short.

In both linguistics and psycholinguistics, a large body of research has been directed toward delineating the factors that determine the intrinsic duration of a phonetic segment (for example, House, 1961; Klatt, 1975; Peterson and Lehiste, 1960). However, relatively little effort has been directed toward the somewhat different question of the manner in which segment duration itself is perceived. An understanding of the processing of duration should facilitate an analysis of the role of perceived duration in speech perception. For example, Klatt and Cooper (1975) have found that the just noticeable difference for the duration of the vowel /i/ within the word *deal* is smaller when the word is in utterance-final position than in any other position in a sentence, despite the fact that the absolute duration of /i/ is longest in this context. They suggest that backward masking from words following *deal* may cause the larger just noticeable differences when the word is not utterance-final. However, as Klatt and Cooper point out, there is no reason to assume that a dimension of the signal such as duration can be masked at the central level. An understanding of the dynamics of perceived duration should serve to clarify this and similar issues. Consequently, the current research was designed to explore the perception of the duration of speech stimuli.

Massaro and Idson (1976) conducted a set of experiments to determine the temporal course of the perceived duration of pure tones. A backward recognition masking paradigm was used, in which one of two target tones of differing durations was presented on each trial, followed after a variable silent interval by one of three masking tone durations. The usual finding in backward recognition masking studies, when the subject is asked to identify pitch (Hawkins, Thomas, Presson, Cozic, and Brookmire, 1974; Massaro, 1970), timbre (Massaro, 1972), loudness (Moore and Massaro, 1973), or sound lateralization (Massaro, Cohen, and Idson, 1976), is that performance improves as a negatively accelerated monotonic function of the intertone interval, asymptoting at a level comparable to that found when no masking tone is presented. In the Massaro and Idson studies, the subject was asked to identify the target tone as being long or short. Under these circumstances, performance averaged over the two targets produced results comparable to those found for other perceptual attributes. The top panel of Figure 1 presents the average percentage of correct test tone identifications as a function of intertone interval and masking tone duration. Average performance improved as a negatively accelerated monotonic function of the intertone interval. This result was obtained for all three masking tone durations. Thus, the average masking function for duration reflects processes similar to those occurring for other stimulus dimensions.

The bottom panel of Figure 1 presents the percentage of correct identifications of each of the target tones as a function of the intertone interval. Performance was initially equal on the two targets at the 5-msec interval. At intervals of 25 and 45 msec the short target was identified far more accurately than the long target, while at longer intervals performance on the short target fell to a level below that found for the long target. These results differ from those typically found in backward recognition masking studies. In masking of pitch, for example, essentially equivalent performance is found for the high and low target tones, whereas markedly different results were obtained for identification of the short and long target tones. Although performance on the long target improved with increases in the silent intertone interval, identification of the short target was better at the short than at the long intervals. Moreover, though identification of the short target was quite good on no-mask trials, identification of the long target was quite poor. Massaro and Idson (1976) extended a more general model of auditory recognition (Massaro, 1975) to account for these results. A sound is assumed to be stored initially in a very early preperceptual memory, having a capacity limit of a single sound. During the auditory recognition process, information in this store is read out continuously over the course of approximately 250 msec. If a second masking sound is presented before processing of the first sound is completed, storage of the second sound will disrupt the representation of the first sound in preperceptual memory and interfere with its processing. Thus, the improved performance found in backward recognition masking tasks with increases in the intertone interval between sounds represents the extraction of successively greater amounts of information from the target, given longer processing times prior to the onset of the mask.

To describe the perception of duration, it is also assumed that the perceived duration of the target increases with the silent processing time available after target presentation (Gol'dburt, 1961). That is, both discriminability and perceived duration are assumed to increase with increased processing time. If sufficient time is allowed for complete processing of the target, the perceived duration of the target will be overestimated, but will be directly related to the temporal extent of the tone. If less than complete processing time is available, the target will be perceived as having a duration shorter than its asymptotic value. If a masking tone is presented before the target has been completely processed, it will have the dual effect of decreasing the target's perceived

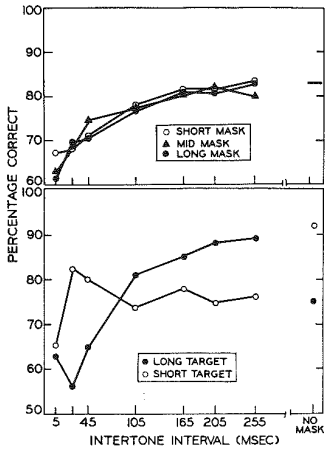

FIGURE 1. Top Panel: Percentage of correct identifications of the target tone as a function of both the duration of the masking tone and the intertone interval. Bottom Panel: Percentage of correct identifications of the target tone as a function of both the duration of the target tone and the intertone interval. (From Massaro and Idson, 1976.)

duration and decreasing its discriminability from the other target. Both discriminability and perceived duration will increase to the extent that the subject is able to process the target prior to the onset of the mask. At short intertone intervals, both long and short targets will have relatively short perceived durations, while at long intertone intervals both targets will be perceived as relatively long. Therefore, at short intertone intervals the long target will be inaccurately identified as short quite often, while identification of the short target will be relatively good. With increases in the intertone interval a successively greater proportion of targets will come to be classified as long, simultaneously increasing accuracy on the long target and decreasing accuracy on the short target. These were, in fact, the results that were obtained (Figure 1).

Five-msec intertone intervals give unique results in that both the short and long targets are poorly identified with no bias to hear the tones as long or short. In terms of this model, a process of temporal integration operates at this interval, which precludes the operations discussed above. For a further description of temporal integration, see Massaro (1975, pp. 446-453), and Massaro and Idson (1976).

A final finding in the tone-duration studies was that the duration of the masking tone had a large influence on performance. The short target followed by the short mask was identified more accurately than the short target followed by the long mask. The long target was identified best when it was followed by the long mask. Therefore, the model includes the additional assumption that the mask adds a component, proportional to its duration, to the perceived duration of the target.

This study was to determine whether the same processes are occurring during the perception of speech sounds. Previous research (Massaro, 1974; Massaro and Cohen, 1975) shows that recognition of both vowels and short consonant-vowel syllables conforms closely to the predictions of Massaro's (1975) model for auditory recognition. When vowels or consonant-vowel syllables are the targets in the backward recognition masking task, performance improves as a monotonic function of the interstimulus interval, asymptoting at an interval of approximately 200 to 250 msec. These results argue that recognition of speech sounds can be conceptualized in a manner similar to that for nonspeech sounds. Moreover, they indicate that backward recognition masking is a suitable paradigm for investigating speech stimuli, and the paradigm has subsequently been employed to investigate processes in speech perception such as consonant perception (Wolf, 1976), feature analysis (Pisoni and McNabb, 1974), and processing of consonant-vowel syllables (Repp, 1975).

A vowel recognition task is ideal for the study of processing of duration in speech. Although vowel quality is of primary importance in distinguishing between English vowels, vowel quantity can cue vowel identification, particularly in distinguishing between vowels with similar formant values (Ainsworth, 1972; Bennett, 1968). For example, the tense vowel /i/ and the lax vowel /ɪ/ have similar formant patterns, but the duration of the tense vowel averages 30% longer than the lax vowel duration (Peterson and Lehiste, 1960). Ainsworth (1972) showed that the same two-formant pattern was identified as /ɪ/ when the vowel was short and /i/ when the vowel was long. If the Massaro and Idson (1976) model is valid, a vowel with a formant structure ambiguous between /i/ and /ɪ/ should be identified progressively more often as /i/ when the duration of that vowel or the intervowel interval (IVI) is increased in a backward masking task. The current studies used this implication to evaluate the applicability of the model to the perception of vowel duration. The first experiment used durations somewhat shorter than those found in natural speech to allow for a direct comparison with the research on tonal duration. The results should still be applicable to processing of natural speech, as Suen and Beddoes (1972) have found that vowels shorter than those naturally occurring can be identified when followed by a silent interval. However, as a further check on this possibility, Experiment 2 replicated the first study using durations comparable to those found in natural speech.

EXPERIMENT 1

Method

Subjects. Subjects were 16 undergraduates who participated an hour a day for four consecutive days.

Design. On each trial, the subjects heard one of the two targets followed, after a variable silent intervowel interval (IVI), by one of the three masks. The targets assumed durations of 60 and 80 msec, while the masks assumed durations of 50, 70, and 90 msec. The intervowel intervals were 5, 25, 45, 105, 165, 205, and 255 msec. The eighth condition, the no-mask trials, was treated as an intervowel of ∞. Target duration, mask duration, and the intervowel interval were all within-subject variables. The set given the subjects for interpretation of the sounds was a between-subject variable. In the nonspeech condition, eight subjects were told that the sounds were nonsense sounds that differed in duration. They were instructed to identify the long sound as long and the short sound as short. In the speech condition, eight subjects were told that the sounds were the vowels /i/ or /ɪ/ and were instructed to identify them as such. The vowel /i/ corresponded to the long target vowel and the vowel /ɪ/ corresponded to the short target vowel, though the subjects in the speech condition were not instructed that the target vowels differed in duration.

Apparatus and Stimuli. All experimental events were controlled by a PDP-8L computer. The vowel stimuli were generated during the experiment proper by a formant series resonator speech synthesizer under digital control (Cohen and Massaro, 1976). The output of the synthesizer was amplified and presented binaurally over matched headphones. Four subjects could be tested simultaneously in separate sound-attenuated rooms.

The vowel stimulus was chosen to be ambiguous between the vowels /i/ and /ɪ/. A fundamental frequency of 194 Hz was used while the first, second, and third formants were given values of 340, 2085, and 2725 Hz, respectively. The amplitude of the vowels was 84.5-dB SPL. All vowels were initiated at the rising edge of the voicing pulse and had instantaneous onsets and offsets.

Procedure. The experiment was conducted on four consecutive days. Each day was divided into two 20-minute sessions, separated by a 10-minute rest break. Day 1 was considered to be practice. On the first session of that day, the subjects received 320 learning trials without any masking vowels. On these trials, the nonspeech subjects identified the test vowel as being either long or short. The speech subjects identified the same vowel as being either /i/ or /ɪ/. Presentation of the two test vowels was completely random and each vowel was programmed to occur with equal probability within a session. During a 1.5-sec response interval, the nonspeech subjects pressed one of two buttons labeled S and L, to indicate a short and long response, respectively. The speech subjects pressed buttons labeled I and E to indicate an /ɪ/ or an /i/ response, respectively. Following the response interval, feedback was provided by a 500-msec display of the letter associated with the target vowel presented on that trial, on a visual display of light-emitting diodes. For the nonspeech subjects, feedback was of the letters S and L, while for the speech subjects feedback was of the letters I and E.

Session 2 of Day 1, and both sessions of the three subsequent experimental days, consisted of 320 test trials. All of Session 2 of Day 1 and the first 20 trials of the six experimental sessions were eliminated from the data analysis, though the subjects were not informed of this. On 87.5% of the trials the test vowel was followed, after one of the variable silent intervowel intervals, by one of the three possible masking vowels. On 12.5% of the trials no masking vowel was presented. The responses, feedback, and intertrial interval were identical to those for the practice session. All 48 experimental conditions (two target durations × three mask durations × seven IVIs plus the no-mask condition) were completely random and were programmed to occur with equal probability within a session. The masking vowel duration was a dummy variable under the no-mask condition.

Results

Response frequencies were pooled over the last three days of the experiment. The percentage of correct judgments for each subject was computed at each target duration by mask duration by intervowel interval condition. These percentages were then submitted to two separate analyses of variance. In one analysis, the no-mask was treated as an IVI of ∞. In the second analysis, the no-mask was excluded and the IVI factor was comprised of only seven levels. All results involving the mask-duration factor are drawn from this second analysis in which the no-mask was excluded. All other results are taken from the analysis including the no-mask. Thus, it was possible to evaluate the effects of masking tone duration independently of the results on the no-mask trials.

The main effect of speech vs nonspeech set was nonsignificant ($F < 1$), as were all interactions involving set as a factor. This means that identification performance did not differ as a function of the set given the subjects, and that the effects of the three within-subject variables did not differ across the two sets. Consequently, all of the data will be discussed in terms of the average results for the two sets.

The top panel of Figure 2 presents the average percentage of correct responses as a function of both the duration of the masking vowel and the intervowel interval. For all three masking vowels, average identification performance improved as a negatively accelerated monotonic function of the intervowel interval, $F = 65{,}83$, $df = 7, 98$, $p < 0.001$. Performance asymptoted at a level only slightly below that found on no-mask trials, a difference of approximately 2%. Similar masking functions were observed for each of the three mask durations although the long masking vowel produced a slightly lower level of performance than did the other two masking vowels; $F = 4.02$, $df = 2.28$, $p < 0.05$. The mask by IVI interaction was nonsignificant.

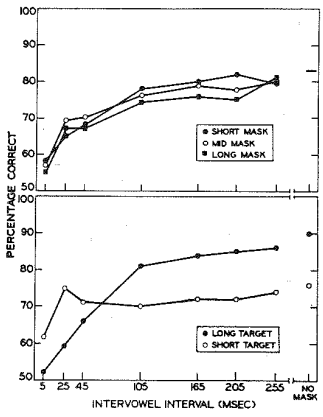

FIGURE 2. Top Panel: Percentage of correct identifications of the target vowel as a function of both the duration of the masking vowel and the intervowel interval. Bottom Panel: Percentage of correct recognitions of the target vowel as a function of both the duration of the target vowel and the intervowel interval. (Experiment 1.)

The bottom panel of Figure 2 presents the percentage of correct identifications of each of the two target vowels as a function of the intervowel interval. The two target vowels produced equivalent overall performance, both targets being identified at a level of approximately 75% correct ($F < 1$). However, the

two target vowels yielded different functions across the intervowel intervals; $F = 6.78$, $df = 7, 98$, $p < 0.001$. Performance on the long target improved monotonically with increases in the intervowel interval. In contrast, identification of the short target vowel improved 13% between intervals of 5 and 25 msec, then fell to an asymptotic level below that of the long target. In addition, a reversal in the level of performance found for the two targets at asymptote was seen on the no-mask trials. Although the long target was identified approximately 12% better at asymptote than was the short target, the short target was identified approximately 14% better on the no-mask trials.

A strong target by mask interaction was obtained in the experiment; $F = 22.42$, $df = 2, 28$, $p < 0.001$. Figure 3 presents the percentage of correct identifications of each of the two target vowels as a function of the duration of the masking vowel. The no-mask trials were not included in this analysis. The short target was identified most accurately when followed by a short mask. The opposite results were found for the long target.

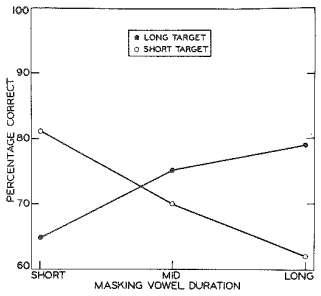

FIGURE 3. Percentage of correct recognitions of each of the target vowels as a function of the duration of the masking vowel. The no-mask trials were eliminated from this analysis. (Experiment 1.)

Figure 4 gives the percentage of correct identifications as a function of both mask duration and the intervowel interval, for the two targets separately. For identification of the short target, the differences in performance under the three masks were relatively constant across intervals, while greater variability was seen for identification on the long target; $F = 3.95$, $df = 12, 168$, $p < 0.001$.

Discussion

Performance using vowels as stimuli closely replicated the earlier results with pure tones (Massaro and Idson, 1976). This correspondence suggests that similar processes may underlie the perception of duration of speech and nonspeech stimuli. Experiment 2 was designed to replicate Experiment 1 in all respects, with the exception of the durations of the vowels, which were altered to fall in the same range as those found for the vowels /i/ and /ɪ/ in natural speech. If the same processes do in fact mediate the perception of duration for speech and nonspeech stimuli, then the results of Experiment 1 should be replicated exactly. Alternatively, if the findings from Experiment 1 are caused by the artificially short vowels employed, then the results of Experiment 2 should differ qualitatively from those of Experiment 1.

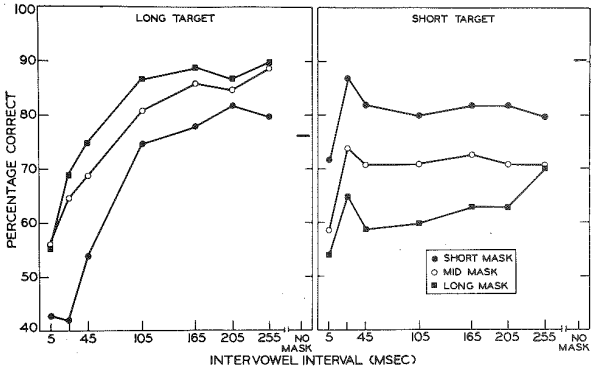

FIGURE 4. Percentage of correct identifications of the long target (left panel) and short target (right panel) vowel as a function of the duration of the masking vowel and the intervowel interval. (Experiment 1.)

EXPERIMENT 2

Method

Subjects. The subjects were eight undergraduates. The subject were tested in two groups of four subjects each.

Procedure. The procedure for Experiment 2 was identical to that for Experiment 1, with the exception that the durations of the vowels were altered. The durations of the target vowels were 180 msec (short) and 240 msec (long). The durations of the masking vowels were kept symmetrical around the target durations. The masking vowels now had durations of 150 (short), 210 (mid),

and 270 (long) msec. One additional change was made in the procedure. The experiment was conducted on five consecutive days rather than four. Day 1 was considered to be practice and these data were not analyzed.

Results

Response frequencies were pooled over the last four days of the experiment and the data analysis was equivalent to Experiment 1. As in Experiment 1, the main effect of the speech vs the nonspeech set was nonsignificant ($F < 1$), as were all interactions involving set as a factor. Consequently, all of the data will be discussed in terms of the average results for the two sets.

The top panel of Figure 5 presents the average percentage of correct responses as a function of both the duration of the masking vowel and the intervowel interval. For all three of the masking vowels, average identification performance improved as a negatively accelerated monotonic function of the intervowel interval; $F = 6.51$, $df = 7, 42$, $p < 0.001$. Performance asymptoted at a level approximately 3% lower than that found on the no-mask trials. Similar masking functions were found for all three of the durations of the masking vowel. Both the main effect of the mask duration and the mask duration by intervowel interval interaction ($F < 1$) were found to be nonsignificant in the analysis.

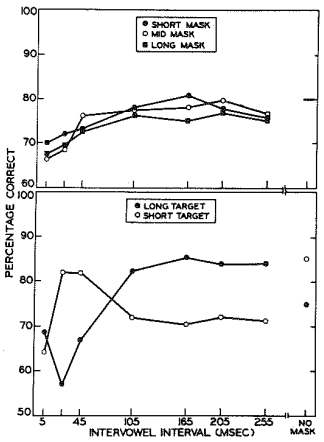

FIGURE 5. Top Panel: Percentage of correct identifications of the target vowel as a function of both the duration of the masking vowel and the intervowel interval. Bottom Panel: Percentage of correct identifications of the target vowel as a function of both the duration of the target vowel and the intervowel interval. (Experiment 2.)

The bottom panel of Figure 5 presents the percentage of correct identifications of each of the target vowels as a function of the intervowel interval. The two target vowels produced identical overall performance, both being identified at a level of 75% correct ($F < 1$). However, the two target vowels again yielded quite different functions across the intervowel intervals; $F = 8.34$, $df = 7, 42$, $p < 0.001$. Between intervals of 25 and 255 msec, performance on the long target improved monotonically with increases in the intervowel interval. In contrast, identification of the short target decreased by approximately 10% between the intervals of 25 and 105 msec, asymptoting at a level of performance below that found on the long target. While the long target was identified far better than the short target at asymptote, the opposite result was found on no-mask trials. When no masking tone was presented, the short target was identified approximately 10% more accurately than the long target. Performance did not differ for the two targets at the 5-msec interval. This result replicates previous work in recognition masking of perceived duration and has been interpreted as reflecting the operation of an integration process at short interstimulus intervals (Massaro and Idson, 1976).

Figure 6 presents the percentage of correct identifications of each of the two target vowels, as a function of the duration of the masking vowels. The no-mask trials have been eliminated from the data presented in the figure. Qualitatively similar results were found to those seen in Experiment 1. The short target was identified more accurately when it was followed by a short mask, and identification of the long target was more accurate when it was followed by a long mask. Although the effect was greater for the short target than it was for the long target, the target-by-mask interaction was smaller than that seen in Experiment 1 and was below the level required for statistical significance; $F = 2.51$, $df = 2, 12$, $p < 0.10$. Figure 7 gives the percentage of correct identifications of the two targets individually, as a function of both the duration of the mask and the intervowel interval. The three-way interaction of target, mask, and intervowel interval was not significant in the analysis.

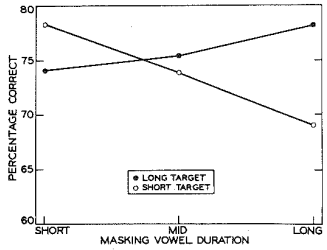

FIGURE 6. Percentage of correct recognitions of each of the target vowels as a function of the duration of the masking vowel. The no-mask trials were eliminated from this analysis. (Experiment 2.)

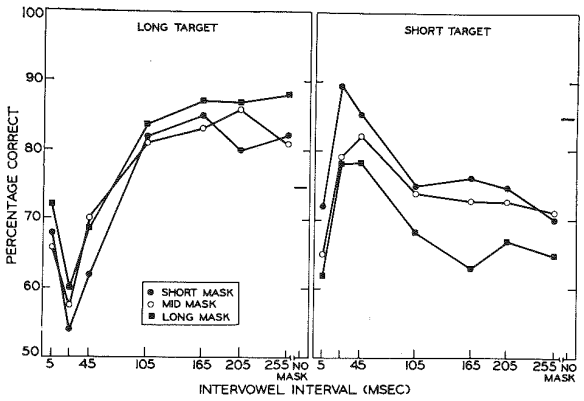

FIGURE 7. Percentage of correct identifications of the long target (left panel) and the short target (right panel) vowels as a function of the duration of the masking vowel and the intervowel interval. (Experiment 2.)

Discussion

The results of Experiment 2 closely replicated those found in Experiment 1. The similar results suggest that the same processes are operating during the perception of the duration of long and short vowels. Because the longer vowels used in Experiment 2 are in the same range as those seen in natural speech, it seems reasonable to assume that the current studies are tapping the same perceptual mechanisms for analyzing speech. The fact that the current studies replicated the earlier work with pure tones (Massaro and Idson, 1976) suggests strongly that the same processes are responsible for the perception of duration for both tone and vowel stimuli.

One additional point needs to be made with respect to the results of Experiment 2, however. Massaro's (1975) model of the auditory recognition process assumes that the time course of this process has a span on the order of 250 msec from the onset of the target. By contrast, identification of the long target improved out to a stimulus onset asynchrony between target and mask of 400 msec. The most obvious explanation for this result is that the recognition process simply had a longer time span than has generally been assumed. However, such a suggestion directly contradicts the large body of evidence that supports the quarter-second limitation on processing (see Massaro, 1975, for a review of this work). A more reasonable interpretation would suggest that the results on the long target reflect special processes operating during the perception of duration. In recognition masking of pitch or loudness, the information relevant to the discrimination between alternatives is available to the processing system virtually from target onset, as the tones differ immediately in frequency or amplitude. However, this is not the case for stimulus duration because the short target is identical to the long target in all respects except duration. Given that the time span for the short target is a subset of the time span for the long target, no useful information concerning target identity is available until the span of the short target has been exceeded. In Experiment 2, for example, for the first 180 msec following target onset the tone could be either the long or the short target. This suggests that not until 180 msec after onset was the target information being used effectively. If recognition processing primarily entails extracting information needed to discriminate between alternatives, it seems reasonable that masking of perceived duration would continue for longer stimulus onset asynchronies, as the resolution process could not work effectively until 180 msec had passed.

The results of the current experiments are consistent with the framework developed in the introduction. It is assumed that the perceived duration increases over time, leveling off at an asymptotic value that is directly related to its actual temporal extent. If a second sound is presented before processing of the first is completed, both the perceived duration of the target and its discriminability will be affected. Formalizing this assumption gives

$$d' = \alpha(1 - e^{-\theta t}) \tag{1}$$

where d' indexes stimulus discriminability, α the maximal amount of stimulus information available given infinite processing time, and θ the rate of extraction of information over time t, measured from the onset of the target to the onset of the mask. The perceived duration (PD) of the target is also assumed to increase as a negatively accelerated function of processing time. Consequently, the perceived duration (PD) of the target can be conceptualized as an exponential growth function of time,

$$PD = \alpha(1 - e^{-(\theta_D t_D + \theta t_I)}) \tag{2}$$

where α is the asymptotic value of perceived duration, θ_D is the rate of growth of PD during the time of target presentation, t_D, and θ represents the rate of growth of PD during the silent interval, t_I, following target offset. When the mask is presented it terminates further growth of the duration percept. Moreover, the mask serves to lengthen the perceived duration of the target by a constant, proportional to the actual duration of the mask. Incorporating this assumption into Equation 2 yields

$$PD = \alpha(1 - e^{-(\theta_D t_D + \theta t_I)}) + K(t_m) \tag{3}$$

where t_m is the mask duration and K is a constant of proportionality. If PDs

and PD_L represent the perceived durations of the short and the long targets, respectively, then the growth of the duration percept over time is given by

$$PD_S = \alpha_S(1 - e^{-(\theta_D t_D + \theta_S t_I)}) + K_S(t_m) \quad (4)$$

and

$$PD_L = \alpha_L(1 - e^{-(\theta_D t_D + \theta_L t_I)}) + K_L(t_m) \quad (5)$$

where θ_D represents the growth rate during target presentation, t_D, and θ_S and θ_L represent the rates of growth of the short and long targets over the silent interval, t_I, to asymptotes α_S and α_L, respectively, and K_S and K_L are constants of proportionality for the short and long targets, respectively. The discriminability of the two targets, as given by d', is then equal to the difference between the perceived durations (compare Massaro and Idson, 1976, for the derivation of d' in this model).

These two processes, the growth of the percept of target duration over time, and the lengthening effect of the mask, work together to determine the final subjective duration of the target. As a result, the combination of the target, mask, and intervowel interval will create a continuum of perceived durations. Despite the fact that only two target durations were present in the experiment, our findings indicate that a whole continuum of perceived durations existed for the subject. However, the masking paradigm does not allow the subject to report these perceived durations directly, as only two response alternatives are available. The task requires the subjects to map the continuum of subjective durations into two discrete response alternatives. The durations of the targets, the masks, and the intervowel intervals determine the range of perceived durations in the experiment. The particular subjective duration present on a given trial would simply represent some point along this continuum. The subject's response strategy would then consist of placement of a decision criterion that would partition the continuum into values of perceived duration to be categorized as long, and values to be categorized as short. The decision criterion represents the value of perceived duration, which must be exceeded for the subject to respond long. If the perceived duration on a given trial is greater than the criterion value, the target will be called long; otherwise it will be called short. The model assumes that a single and optimal criterion will be used throughout the experiment. It should be noted that a subjective continuum from /ɪ/-ness to /i/-ness can be used in the same way as the subjective duration continuum. In this case, the subject would respond /i/ if the percept exceeded the criterion value of /i/-ness, and /ɪ/ otherwise.

The model can handle the results of the current studies directly. At short intervowel intervals with relatively little processing time available, both targets will have relatively short perceived durations. As a result, the perceived durations of the targets will only rarely exceed the criterion for responding long, and the target will almost always be categorized as short. This will insure accurate identification of the short target, but quite poor identification of the long target. As the intervowel interval is lengthened, increasing the available processing time, the perceived durations of both targets will also increase. Consequently, as the criterion is fixed, a successively greater proportion of the targets will exceed criterion and will be categorized as long. This will have the effect of simultaneously increasing accuracy on the long target and decreasing accuracy on the short target. The model can thus predict the exact form of the target duration by intervowel interval interaction (see bottom panel of Figures 2 and 5). The long target is identified more accurately with increases in the intervowel interval, while the short target is identified most accurately at the short intervals.

The model can also account for the effects of the individual masking tones and the results of the no-mask trials. The masking tone is assumed to add a constant, proportional to its duration, to the perceived duration of the target. As a result, a target followed by a mask after a long intervowel interval will appear longer than that same target on no-mask trials. Thus, both the long and the short targets will be categorized as short on a disproportionately large number of no-mask trials. This predicts the obtained result of high accuracy in identifying the short target and poor accuracy in identifying the long target on no-mask trials.

To apply this model to the obtained data, the assumption is made that noise in the sensory system creates a distribution of perceived durations for each experimental condition. The noise distribution is assumed to be normal with a variance equal to one. The data were first transformed into relative standard deviation units, which provide an index of perceived duration. In the context of the current model, the probability of responding long is a measure of the perceived duration of the target. Thus, the probability of responding long given a long target, $P(L|L)$, and the probability of responding long given a short target, $P(L|S)$, were each computed for each experimental condition. A value of $P(L|L)$ or $P(L|S)$ can be used to determine the location of the mean of subjective durations at that condition with respect to the fixed criterion value. The z-distance between the mean of the relevant distribution and the criterion value can, therefore, be computed for both the long and the short targets. If $P(L|L)$ or $P(L|S)$ is greater than 0.5 this indicates that the mean of the distribution is to the right of the criterion. In this case the z-distance will be positive and will correspond to the proportion of the distribution between the mean and criterion value. If $P(L|L)$ or $P(L|S)$ is less than 0.5, the mean is to the left of the criterion and the z-distance will be negative and will correspond to the proportion of the distribution between the mean and the criterion value. In all cases the z-distance will be simply the transform of the proportion of the long or short distribution lying between the mean and the criterion (see Massaro, 1975, for a more complete discussion and a table of z-distances). To insure that all scores were positive, so that Equations 4 and 5 could be applied, the value of +3 was added to all of the z-values.

The model was then fit to these values of $z[P(L|L)]$ and $z[P(L|S)]$ by

minimizing the squared deviations between the predicted and observed values using the iterative minimization routine STEPIT (Chandler, 1969). The 5-msec intervowel interval was excluded from the fit, as a process of integration that occurs at this interval precludes direct application of the model (Massaro, 1975; Massaro et al., 1976). The value of t in Equations 4 and 5 is equal to the stimulus onset asynchrony (SOA). For the long target these values were 115, 135, 195, 255, 295, and 345 msec, while for the short target the SOAs were 75, 95, 155, 215, 255, and 305 msec.

Figure 8 presents the predicted and observed values of $z[P(L|L)]$ and

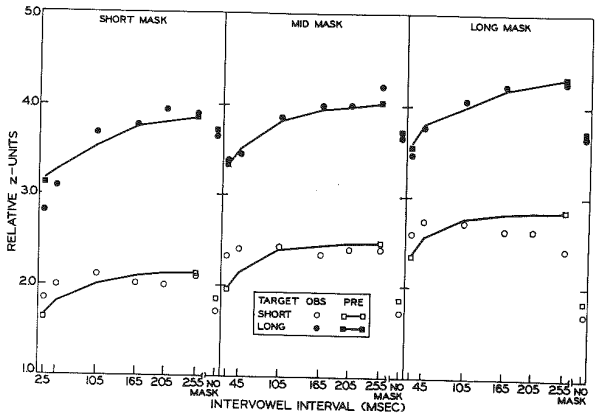

FIGURE 8. Observed and predicted values of $z[P(L|L)]$ and $z[P(L|S)]$ as a function of both the duration of the masking vowel and the intervowel interval. The left-hand panel displays the data for the short masking vowel, the middle panel for the mid masking vowel, and the right-hand panel for the long masking vowel. The no-mask values do not differ as a function of masking vowel duration because the latter is a dummy variable under the no-mask condition. (Experiment 1.)

$z[P(L|S])$ for Experiment 1. The observed points are given by the filled and unfilled circles for the long and the short targets, respectively. The predicted scores are represented by the lines connected by the filled squares for the long target and by the unfilled squares for the short target. The predicted values were computed by estimating the parameter values α_L, α_S, θ_D, θ_L, θ_S, K_L, and K_S by minimizing the squared deviations between the predicted and observed values. As can be seen from the figure, the model provides a relatively good fit to the data, considering that 38 data points (two targets × three masks × six intervowel intervals + two no-mask targets) were fit with just seven parameters. The average squared deviation between the predicted and the observed values was 0.024. The parameter estimates for α_S and α_L were 1.85 and 3.73, respectively. These parameter estimates were in accord with the model, the long target having a greater potential perceived duration than the short target. The parameter estimates for θ_D, θ_L, and θ_S were 14.23, 11.58, and 25.00, respectively. Processing of the short target, therefore, occurred at a faster rate than the long target. The parameter estimates for K_L and K_S were 5.89 and 8.99. This indicates that the duration of the mask had a greater effect on the perceived duration of the short than the long target.

The same model was then fit to the data from Experiment 2. Figure 9 shows

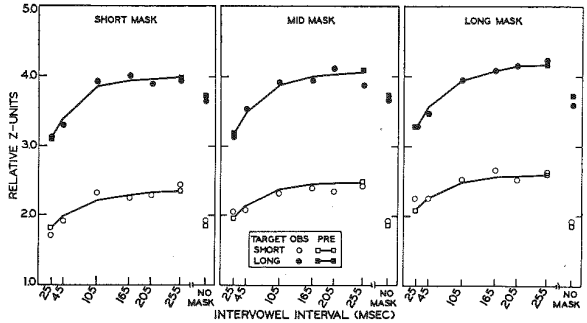

FIGURE 9. Observed and predicted values of $z[P(L|L)]$ and $z[P(L|S)]$ as a function of both the duration of the masking vowel and the intervowel interval. The left-hand panel displays the data for the short masking vowel, the middle panel for the mid masking vowel, and the right-hand panel for the long masking vowel. The no-mask values do not differ as a function of masking vowel duration because the latter is a dummy variable under the no-mask condition. (Experiment 2.)

the predicted and observed values of $z[P(L|S)]$ and $z[P(L|L)]$ for Experiment 2. As can be seen from the figure, the model provided a reasonably good fit to the observed results. Once again, the model fit 38 data points with only seven parameters. The averaged squared deviation between the predicted and observed values was 0.007. The parameter estimates for α_L and α_S were 3.76 and 1.87, respectively. The parameter estimates for θ_D, θ_L, and θ_S were 3.98, 20.34, and 22.96, respectively. The parameter estimates for K_L and K_S were 1.74 and 2.83, respectively. The relationship between the parameter estimates are similar to those found in Experiment 1.

The results of the current study suggest that the perception of duration of speech stimuli, or at least of synthesized vowels, follows essentially the same processes as the perception of duration of nonspeech stimuli. This conclusion is further supported by the finding that the set given the subjects had no effect on their performance; it did not matter whether the stimuli were treated as speech or as nonsense sounds. In addition, the quantitative model for perception of pure tones would appear to describe the manner of processing of vocalic duration as well.

ACKNOWLEDGMENT

This research was supported by Public Health Service Grant No. MH-19399 and a University of Wisconsin Graduate Traineeship. Requests for reprints should be sent to Dominic W. Massaro, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

REFERENCES

AINSWORTH, W. A., Duration as a cue in the recognition of synthetic vowels. *J. acoust. Soc. Am.*, **51**, 648-651 (1972).

BENNETT, D. C., Spectral form and duration cues in the recognition of English and German vowels. *Lang. Speech*, **11**, 65-85 (1968).

CHANDLER, J. P., Subroutine STEPIT-finds local minima of a smooth function of several parameters. *Behavl Sci.*, **14**, 81-82 (1969).

COHEN, M. M., and MASSARO, D. W., Real-time speech synthesis. *Behav. Res. Meth. Instrum.*, **8**, 189-196 (1976).

DENES, P., Effect of duration on the perception of voicing. *J. acoust. Soc. Am.*, **27**, 761-764 (1955).

GOL'DBURT, S. N., Investigation of the stability of auditory processes in micro-intervals of time (new findings in back masking). *Biophysics*, **6**, 809-817 (1961).

HAWKINS, H. L., THOMAS, G. B., PRESSON, J. C., COZIC, A., and BROOKMIRE, D., Tonal specificity and masking in auditory recognition. *J. exp. Psychol.*, **103**, 530-538 (1974).

HOUSE, A. S., On vowel duration in English. *J. acoust. Soc. Am.*, **33**, 1174-1178 (1961).

KLATT, D. H., Vowel lengthening is syntactically determined in connected discourse. *J. Phonetics*, **3**, 161-172 (1975).

KLATT, D. H., Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *J. acoust. Soc. Am.*, **59**, 1208-1221 (1976).

KLATT, D. H., and COOPER, W. E., Perception of segment duration in sentence contexts. In A. Cohen and S. G. Nooteboom (Eds.), *Structure and Process in Speech Perception.* (Vol. 7) New York: Springer-Verlag, 69-86 (1975).

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

LEHISTE, I., Phonetic disambiguation of syntactic ambiguity. *Glossa*, **7**, 107-122 (1973).

MASSARO, D. W., Preperceptual auditory images. *J. exp. Psychol.*, **85**, 411-417 (1970).

MASSARO, D. W., Stimulus information vs. processing time in auditory pattern recognition. *Percept. & Psychophys.*, **12**, 50-56 (1972).

MASSARO, D. W., Perceptual units in speech recognition. *J. exp. Psychol.*, **102**, 199-208 (1974).

MASSARO, D. W., *Experimental Psychology and Information Processing.* Chicago: Rand-McNally (1975).

MASSARO, D. W., and COHEN, M. M., Preperceptual auditory storage in speech recognition. In A. Cohen and S. G. Nooteboom (Eds.), *Structure and Process in Speech Perception.* (Vol. 7) New York: Springer-Verlag, 226-243 (1975).

MASSARO, D. W., and COHEN, M. M., Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction. *Percept. & Psychophys.*, **22**, 373-382 (1977).

MASSARO, D. W., COHEN, M. M., and IDSON, W. L., Recognition masking of auditory lateralization and pitch judgments. *J. acoust. Soc. Am.*, **59**, 434-441 (1976).

MASSARO, D. W., and IDSON, W. L., Temporal course of perceived auditory duration. *Percept. & Psychophys.*, **20**, 331-352 (1976).

MOORE, J. J., and MASSARO, D. W., Attention and processing capacity in auditory recognition. *J. exp. Psychol.*, **99**, 49-54 (1973).

PETERSON, G. E., and LEHISTE, I., Duration of syllable nuclei in English. *J. acoust. Soc. Am.*, **32**, 693-703 (1960).

PISONI, D. B., and MCNABB, S. D., Dichotic interactions and phonetic feature processing. *Brain & Lang.*, **1**, 351-362 (1974).

RAPHAEL, L. J., Preceding vowel duration as a cue to voicing characteristics of word-final consonants in English. *J. acoust. Soc. Am.*, **51**, 1296-1303 (1972).

REPP, B. H., Dichotic forward and backward "masking" between consonant-vowel syllables. *J. acoust. Soc. Am.*, **57**, 483-496 (1975).

SUEN, C. Y., and BEDDOES, M. P., Discrimination of vowel sounds of very short duration. *Percept. & Psychophys.*, **11**, 417-419 (1972).

WOLF, C. G., A recognition masking study of consonant processing. *Percept. & Psychophys.*, **19**, 35-46 (1976).

Received August 30, 1976.
Accepted July 1, 1977.