Journal of Phonetics (1980) 000–000

The role of perceived duration in the identification of vowels

Wendy L. Idson* and Dominic W. Massaro*
University of Texas at Austin

University of Wisconsin, Madison

Received 20th March 1979

Abstract

The role of perceived duration in discriminating between tense and lax vowel quality was explored. Two target vowels, having identical formant structures but different durations, were employed as stimuli in a backward recognition masking task. One of the two target vowels was presented on each trial, followed after a variable silent intervowel interval by a masking vowel. Six masking vowels were utilized, differeing in both formant structure and duration. The subjects' task was to identify the target vowel as being tense or lax. In Experiment 1, identification of the long target improved monotonically with increase in the intervowel interval, while identification of the short target decreased slightly in accuracy with increases in the intervowel interval. Identification of the long target vowel was most accurate when followed by a long masking vowel and least accurate when followed by a short masking vowel. While the opposite results were obtained for the short target vowel. In Experiment 2, subjects rated the quality of the target vowel along a continuous scale between the two vowel alternatives. The target vowel was rated as being progressively more tense with increases in the durations of the target vowel, the masking vowel, and the intervowel interval. The results were interpreted within an information processing model describing the perception of duration. Both experiments demonstrated that perceived duration rather than actual stimulus duration is critical to the tense-lax distinction.

Introduction

The role of duration in speech perception has been extensively studied. Investigations of the intrinsic durations of phonetic segments (e.g., House, 1961; Klatt, 1975; Peterson & Lehiste, 1960) and environmentally contingent modifications of intrinsic duration (e.g. Klatt, 1971) have indicated that duration contains an abundance of potential information about the speech signal. At the segmental level, speech segments have different intrinsic durations (Klatt, 1975),

*Requests for reprints may be sent to either author: Wendy L. Idson, Department of Psychology, Mezes Hall 330, University of Texas at Austin, Austin, Texas 78712; Dominic W. Massaro, Department of Psychology, 1202 West Johnson Street, University of Wisconsin, Madison, Wisconsin 53706; U.S.A.

0095–4470/80/000000 + 00$02.00/0

which can be used to discriminate among them. Duration is sufficient to distinguish, for example, between long and short vowels (Klatt, 1976; Nooteboom, 1973) or between voiced and voiceless fricatives (Cole & Cooper, 1975). Environmental modifications of duration also carry perceptually significant information. For example, a given vowel will be shorter when it precedes a voiceless rather than a voiced consonant (Delattre, 1962; House & Fairbanks, 1953), a variation which contributes to the identification of consonant (Denes, 1955, Massaro & Cohen, 1977; Raphael, 1972). At the suprasegmental level, duration can be informative with respect to the syntactic and semantic content of an utterance (Klatt, 1976; Lehiste, 1970). For example, segmental durations tend to be longer in phrase-final words (Klatt, 1975; Martin, 1970), perhaps to provide a perceptual cue for interpreting the surface structure (Klatt & Cooper, 1975). Duration can, in fact, be utilized to locate the constituent boundary in an otherwise ambiguous phrase (Lehiste, 1973. Lehiste, Olive, & Streeter, 1976). Thus, duration is central to an analysis of the speech signal on multiple levels.

Previous research, by focusing on the function of duration in speech perception, has not directly addressed the somewhat different issue of the manner in which the duration of a phonetic segment is perceived. Yet an investigation of the nature of perception may be more important for an analysis of the function of duration than for other aspects of the speech signal. A large body of research on the perception of duration of nonspeech stimuli (e.g. Allan, 1976; Cantor & Thomas, 1976; Idson & Massaro, 1977; Kristofferson, 1977; Massaro & Idson, 1976; 1978*a*; Thomas & Cantor, 1975, 1976) suggests that duration differs in important respects from other stimulus attributes (see Allan & Kristofferson, 1976; Massaro & Idson, 1976, for reviews). One of the most important of these differences concerns the close correspondence which is genreally found between the physical and psychological dimensions of a stimulus. Perceived pitch, for example, is primarily a function of frequency, subject to the influence of factors such as intensity. Yet in both audition (Efron, 1970*a, b, c*; Gol'dburt, 1961; Idson & Massaro, 1977; Massaro & Idson, 1978*b*) and vision (Cantor & Thomas, 1976; Thomas & Cantor, 1975, 1976), perceived duration is a function not only of the physical characteristics of a stimulus but also of the conditions under which it is perceived. The most important factor concern the time available to perceive the stimulus and the nature of the following stimulus. A given stimulus, having constant physical characteristics, will have a variety of perceived durations dependent upon processing time and stimulus context. These variations in perceived duration imply that physical duration may be only one of several attributes important for identification of a phonetic segment. Accordingly, the current research was designed to explore the role of variables influencing perceived duration in phonetic segment identification.

A framework for studying the perception of duration has been developed by Massaro and Idson, 1976, 1978*a* and Idson & Massaro, 1977. The employed a backward recognition masking paradigm, in which one of two alternative target tones is presented for identification on each trial. The target tone is either presented alone, or is followed after a variable silent interval by a masking tone. The usual finding in backward recognition masking studies, when the subject is asked to identify pitch (Hawkins, Thomas, Presson, Cozic, & Brookmire, 1974; Massaro, 1970) timbre (Massaro, 1972), loudness (Moore & Massaro, 1973), or sound lateralization (Massaro, Cohen & 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 duration studies, target tones of 60 and 80 ms were used, with masking tones of 50, 70, and 90 ms. The subjects were 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

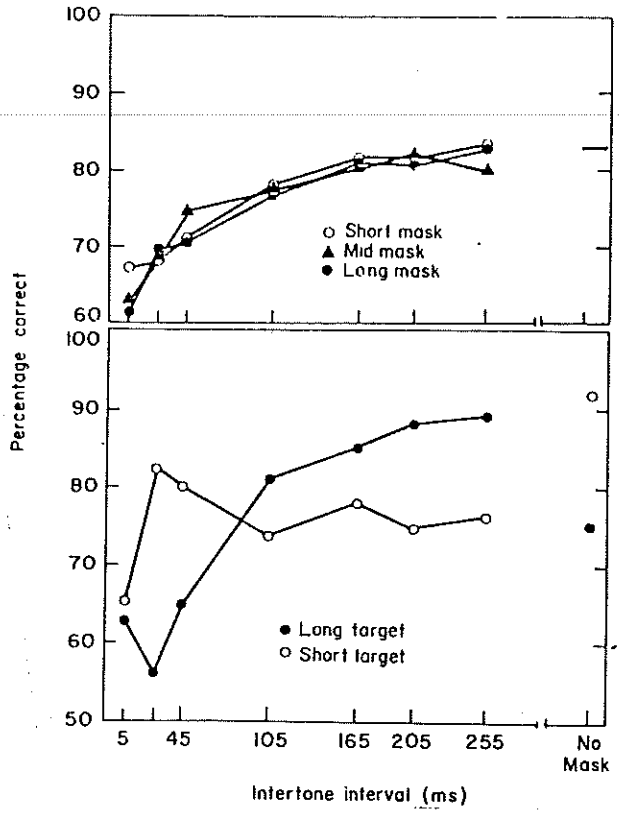

Figure 1

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

attributes. The top panel of Fig. 1 presents the average percentage of correct test tone identifications as a function of the intertone interval and masking tone duration. Average performance improved as a negatively accelerated monotonic function of the intertone interval, for all three masking tone durations, reflecting processes similar to those occurring for other stimulus dimensions.

The bottom panel of Fig. 1 presents the percentage of correct identifications of each of the target tones as a function of the intertone interval. At intervals of 25 and 45 ms the short target was identified far more accurately than the long target, while at longer intevals performance on the short target fell to a level below that found for the long target. 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, identification of the short target was quite good on no-mask trials, while identification of the long target was quite poor. These results differ markedly from those typically found in backward recognition masking studies. In recognition masking of pitch, for example, performance improves with increases in the intertone interval for both high and low frequency target tones.

These results can be interpreted within a more general model of auditory recognition

(Massaro, 1975; Massaro & Idson, 1976). A sound is assumed to be initially held in a very early preperceptual storage, having a capacity limit of a single item. During a primary recognition process, information in this store is read out continuously over the course of approximately 250 ms. The preperceptual information is matched to information in long-term memory, in order to generate a synthesized percept of the sound. The resulting percept is stored in a synthesized auditory memory, which can hold multiple items for a period of several seconds. If a second masking sound is presented before processing of the first sound is complete, 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 a backward recognition masking task with increases in the intertone interval between sounds represented the extraction of successively greater amounts of information from the target sound, given longer processing times prior to the onset of the masking sound. In the context of this model, the accuracy with which the target sound can be identified will increase with the intertone interval, but the average perceptual experience of the target sound will not change. For example, while the pitch of a tone will be resolved more accurately at longer intertone intervals, that tone will not come to sound consistently higher or lower as the interval is lengthened.

Perception of duration is assumed to occur in essentially the same manner. However, duration differs from other attributes of the auditory stimulus in that both the accuracy of discrimination and the nature of the perceptual experience are conceptualized as changing over time. Perceived duration is assumed to increase with increases in processing time to an asymptotic value directly related to physical duration. In the backward masking task, presentation of the masking tone before the target tone has been completely processed will have the dual effect of decreasing the target tone's perceived duration and decreasing its discriminability from the other target. To the extent that the subject can process the target tone prior to onset of the masking tone, both perceived duration and discriminability will increase. Accordingly, at short intertone intervals both the long and the short target tones will have relatively short perceived durations, while at long intertone intervals both target tones will have relatively long perceived durations.

Despite the fact that there are only two target tones, the subject experiences a continuum of perceived durations. To perform the task, this continuum must be mapped into the binary response choice of short or long. The subject must establish a criterion for perceived duration, above which the target tone will be categorized as short. Given this strategy, 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 both target tones will be classified as long, simultaneously increasing accuracy on the long target and decreasing accuracy on the short target.

The relatively brief intertone intervals between the target and masking tones make it unlikely that the perceived duration of the target tone could be determined prior to onset of the masking tone (see Massaro & Idson, 1976, Experiment 5). Accordingly, the perceived duration of the masking tone will influence the judged duration of the target tone. There are a number of possible mechanisms by which the masking tone could exert its influence (see Idson & Massaro, 1977), the most probable of which is that the masking tone not only disrupts processing of the duration of the target tone but also modifies the memory for the duration of that tone (Kallman & Massaro, in press; Massaro & Idson, 1978*b*). Regardless of the mechanism, the model assumes that some proportion of the perceived duration of the masking tone will be added to that of the target tone, lengthening the judged duration

of the target tone. The contribution of the masking tone entails that, with sufficient processing time, a target tone will have a longer judged duration on a masking trial than on a no-mask trial. Both target tones will be classified as short disproportionately often on no-mask trials, yielding the excellent performance on the short target tone and poor performance on the long target tone which was observed. The lengthening effect of the masking tone is also supported by the strong target tone by masking tone interaction which was found. A short target tone was identified more accurately when it was followed by a short than by a long masking tone; a long target tone was identified more accurately when it was followed by a long than by a short masking tone.

In order to provide converging evidence for the model, Idson and Massaro (1977) attempted to find a more direct index of the changes in perceived duration which occur in a backward masking task. Instead of categorizing the target tone as short or long, the subject's task was to estimate the duration of the target tone along a continuous scale of perceived duration. This procedure allowed a direct report of the perceived duration of the target tone. In accord with the model, the rated duration of the target tone increased monotonically with increases in the intertone interval. The top panel of Fig. 2 presents the rated

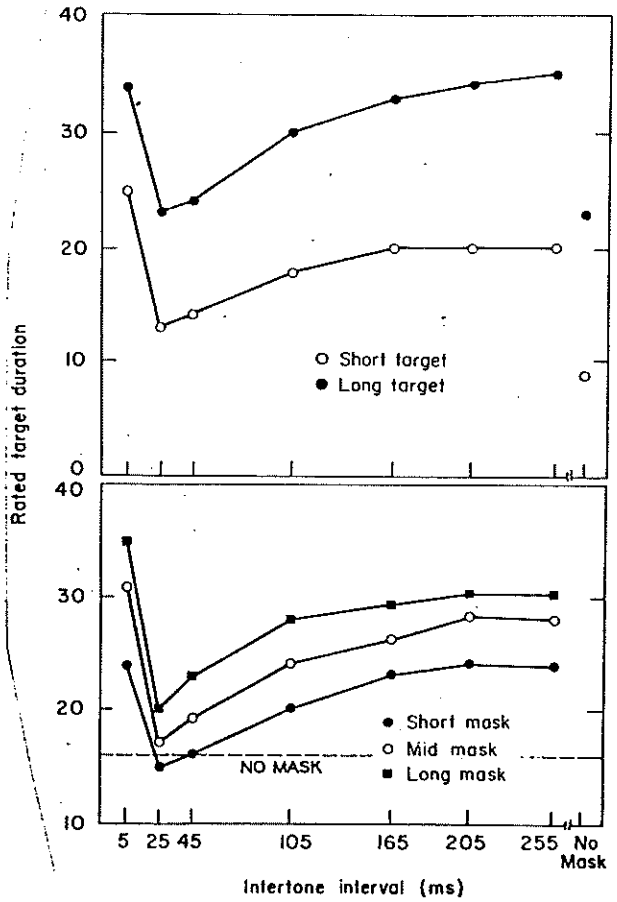

Figure 2

Top Panel: Rated durations of each of the two target tones, as a function of the intertone interval. Bottom Panel: Rated target durations, as a function of both the intertone interval and the duration of the masking tone (from Idson & Massaro, 1977).

duration of each target tone as a function of the intertone interval. As can be seen from the figure, the rated duration of the target tone increased as the intertone interval was lengthened from 25 to 255 ms. The longer rated durations at the 5 ms interval reflected an integration of the target and masking tones into a single composite sound (Massaro, 1975), which yields a longer perceived duration at this than at any other interval (Idson & Massaro, 1977). This result can also be seen in the identification task, where performance at the 5 ms interval also reflects an integration of the target and masking tones (see Fig. 1).

Consonant with the assumption that the physical duration of the target tone contributes to its perceived duration, the long target tone was always rated as having a longer perceived duration than the short target tone, at the same intervals. For both target tones, however, rated duration increases monotonically with the intertone interval, indicating that perceived duration increased with increasing processing time. Moreover, the target tones were both rated as shorter on no-mask trials than on masking-tone trials, supporting the assumption that the masking tone also contributes to the perceived duration of the target tone. Additional support for this assumption came from the findings, shown in the bottom panel of Fig. 2, that the rated duration of both target tones increased with the duration of the masking tone. The target tones were rated as shortest when followed by a short masking tone and as longest when followed by a long masking tone.

The model has been shown to be applicable to the perception of duration of speech stimuli as well. When vowels or consonant-vowel syllables are used as targets in a backward masking task (e.g. Massaro, 1974; Massaro & Cohen, 1975), the results are comparable to those obtained with pure tones, performance improving as a monotonic function of the intertone interval. Massaro and Idson (1978) utilized the backward masking paradigm to explore duration perception with speech stimuli. A vowel having formant frequencies ambiguous between /i/ and /I/ was employed as a stimulus in the binary choice task. The target vowel had either a long or a short duration, with the durations of the masking vowel being symmetrical around those of the target vowel. Figure 3 presents the results of a study utilizing durations equivalent to those employed in the pure tone experiments (target = 60/80 ms, mask = 50/70/90 ms). A second study used durations comparable to those found for vowels in natural speech (target = 180/240 ms, mask = 150/210/270 ms). In both cases, the results were parallel to those obtained for pure tones. The comparability of the results for vowels and tones suggests that the model is tapping fundamental processes in duration perception.

The backward masking paridigm can be considered to represent almost a prototype of the stimulus situation contained in the speech signal. The speech signal consists of qualitatively different phonetic segments of quite short intrinsic durations (Peterson & Lehiste, 1960), with successive segments following one another rapidly in time. Accordingly, there will often be insufficient time to completely process one phonetic segment prior to onset of a subsequent segment. If perceived duration is a function of both available processing time and the duration of a following stimulus then the perceived duration of a phonetic segment may not be directly determined by its actual duration. Thus, perceived duration may provide functional information for the identification of a phonetic segment.

The current research was designed to provide a case in point, by considering the role of percieved duration in the tense-lax distinction for vowels. Within a pair of vowels having highly similar formant frequencies, the lax vowel tends to be shorter than the tense vowel (Ainsworth, 1972; Bennet, 1968), a difference which can be used to discriminate between members of the pair. If only the actual duration of the vowel is critical to the tense-lax distinction, then the vowel of a given duration should always be perceived as either tense or

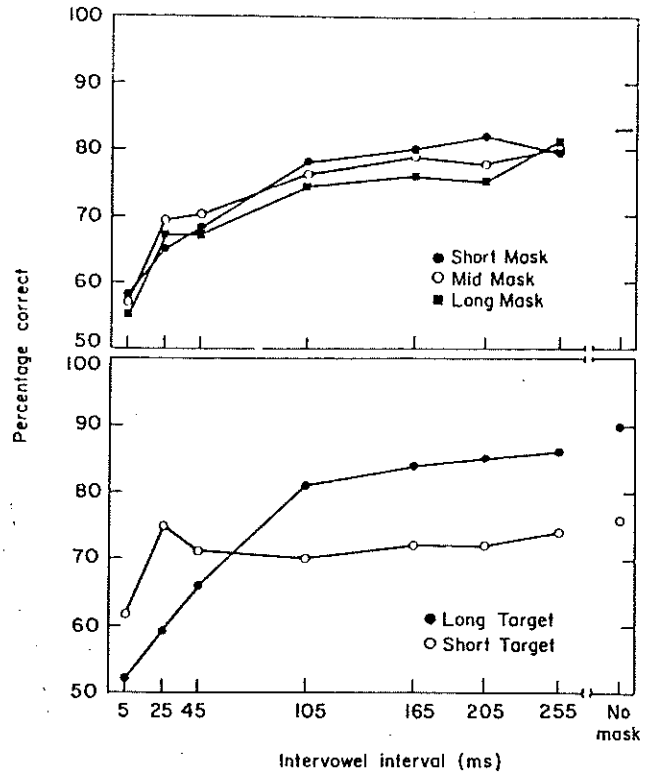

Figure 3

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 durations of the target vowl and the intervowel interval (Ffrom Massaro & Idson, 1978*a*).

lax. Alternatively, if perceived duration is important, then categorization of a vowel as tense or lax may also be dependent upon the available processing time and the nature of the following segment.

Experiment 1 was designed to distinguish between these alternatives, using the backward masking paradigm. Two vowels were employed as stimuli, a front vowel having formant frequencies ambiguous between the tense vowel /i/ and the lax vowel /I/ and a back vowel having formant frequencies ambiguous between the tense vowel /u/ and the lax vowel /U/. Separate groups of subjects received long and short versions of either the front of the back vowel as targets. Each group received as masking vowels three durations of both the front and the back vowel. The subject's task was to identify the front vowel as being /i/ or /I/, or to identify the back vowel as being /u/ or /U/. While no explicit judgement of duration was required, the target vowel could be discriminated only in terms of duration. If identification of tense-lax vowel quality is dependent solely upon the actual duration of the target vowel, than neither the intervowel nor the duration of the masking vowel should affect identification of the target vowel. Alternatively, identification of the target vowel might be influenced by both of these variables, as they both affect perceived duration. Under this conceptualization, both target vowels should be categorized as lax at short intervowel intervals and as tense as long intervowel intervals. Since the presence of a masking vowel will

increase the perceived duration of the target vowel, both target vowels should be categorized as lax more often on no-mask trials than on trials on which a masking vowel is presented. Moreover, at any given interval, both target vowels should be categorized as lax more often when followed by a short masking vowel and as tense more often when followed by a long masking vowel. The effect of masking vowel duration might be lessened when the target and masking vowels are dissimilar in formant structure, since this dissimilarity could make it easier for the subject to exclude the masking vowel from judgement of the target vowel.

Experiment 1

Method

Design

The experiment design was a 2 x 2 x 6 x 4 factorial, with the quality of the target vowel, the duration of the target vowel, the masking vowel, and the intervowel interval (IVI) as factors. The vowel quality factor was between-subjects variable; all other factors were within-subject variables. Two stimulus vowels were employed. The first vowel had formant frequencies ambiguous between the tense vowel /i/ and the lax vowel /I/, and will be referred to as the front vowel. The second vowel had formant frequencies ambiguous between the tense vowel /u/ and the lax vowel /U/, and will be referred to as the back vowel. Half of the subjects received the front vowel as the target vowel, with the remaining subjects receiving the back vowel as the target vowel. The two target vowel durations were 50 ms (short) and 90 ms (long). The six masking vowels resulted from a combination of vowel quality and duration. The front and back vowels were each presented at each of three durations, 30 ms (short), 70 ms (mid), and 110 ms (long). The four IVIs corresponded to the three temporal intervals between the target and masking vowels of 25, 85 and 165 ms, plus a condition in which a masking vowel was not presented (no-mask condition).

Apparatus and stimuli

All experimental events were controlled by a PDP-8L computer. The vowel stimuli were generated on-line by a formant series resonator speech synthesizer (Fonema OVE-IIId), under digital control (see Cohen & Massaro, 1976). The output of the synthesizer was gated by two computer controlled audio switches (Iconic Model #0137) to separate amplifiers (McIntosh Model MC-50) for the two ears, and then presented over headphones (Grason-Stadler TDH-49). The visual feedback was given over a display of light emitting diodes (Monsanto Model MDA-III). Subjects were tested simultaneously in four individual sound insulated rooms.

A fundamental frequency of 194 Hz was utilized in creating both stimulus vowels. The front vowel had first, second, and third formant frequencies of 330, 2190 and 2780 Hz, respectively. The back vowel had first, second, and third formant frequencies of 370, 945, and 2240 Hz, respectively. All vowels were initiated at the rising edge of the voicing pulse, and had instantaneous onsets and offsets.

Procedure

The experiment was conducted on five consecutive days. Each day was divided into two 20-min sessions, separated by a 10-min rest break. Each session consisted of 320 trials, the first 20 being unscored practice trials. Both sessions of day 1 were learning sessions. The experiment proper was conducted on days 2–5, with day 2 considered as practice. The subjects were not informed that any portion of the experiment was to be treated as practice.

In the learning sessions, the subjects were taught to identify the target vowels, when the masking vowels were not presented. The group receiving the front vowel as a target vowel

were instructed to call the short vowel /I/ and the long vowel /i/. The short vowel was identified by the letter "I" and the long vowel by the letter "E". The group receiving the back vowel as a target were instructed to call the short vowel /U/ and the long vowel /u/. The short vowel was identified by the letter "O" and the long vowel by the letter "U". The subjects were *not* told that the alternative target vowels differed in duration, but rather were instructed to idendify them in terms of quality. On each of the trials, one of the two target vowels was presented. During a 1.5 s response interval, timed from the offset of the target vowel, the subject determined which vowel had been presented. The response was indicated by pressing one of two buttons labeled with either "I" and "E" or "O" and "U", depending upon the level of the between-subjects factor. Following the response interval, visual feedback was given by a 500 ms presentation of the letter associated with the target vowel actually presented on that trial. A 1 s interval separated successive trials.

In the experimental sessions, the subjects identified the target vowels in the manner mastered during the learning sessions. On 3/4 of the trials, the target vowel was folled after a variable silent IVI by one of the six possible masking vowels. On the remaining 1/4 of the trials, no masking vowel was presented. The response, feedback, and intertrial interval were identical to those employed in the practice sessions. All 48 within-subject experimental conditions (2 target vowel durations x 6 masking vowels x 4 IVIs) were completely random and were programmed to occur equally often within a session.

Subjects

The subjects were six University of Wisconsin undergraduates, who volunteered their participation as part of an Introductory Psychology course.

Results and discussion

The dependent variable is the percentage of correct responses. The percentage correct was computed for each subject, on each day, at each target identity by target duration by masking vowel by IVI condition. These percentages were submitted to an analysis of variance, which revealed the main effect of days ($F < 1$) and all interactions involving days as a factor to be nonsignificant. These results indicate that performance was not changing with practice. Accordingly, in order to increase the reliability of the individual scores, the data were pooled over the three experimental days and reanalyzed. Two separate analyses were conducted. In the first analysis, the no-mask condition was treated as an IVI of ∞ (see Massaro, 1975, for a theoretical justification of this procedure) and entered the analysis as a fourth level of the IVI factor. In the second analysis, the no-mask condition was excluded and the IVI factor had only three levels. This latter analysis was conducted so as to allow a direct evaluation of the effects of the masking vowel, unconfounded by those trials on which a masking vowel was not presented. All of the results given below which involve the masking vowel duration as a factor were drawn from the analysis excluding the no-mask condition.

The between-subjects factor of vowel quality had little influence upon performance. Although the back target vowel was identified 5% more accurately than the front target vowel, the result was not significant, $F(1, 4) = 1.14$, $P > 0.26$. With the exception of the target identity by target duration by masking vowel interaction, $F(5, 20) = 3.69$, $F < 0.025$, all interactions involving the identity of the target vowel as a factor were statistically nonsignificant ($F < 1$ in all cases). Accordingly, all of the data will be discussed in terms of the average results for the two target vowel identities.

Figure 4 presents the percentage of correct identifications of the durations of the individual target vowels, as a function of the IVI. Quite different results were found for the two target vowel durations. Performance on the long target vowel improved monotonically across

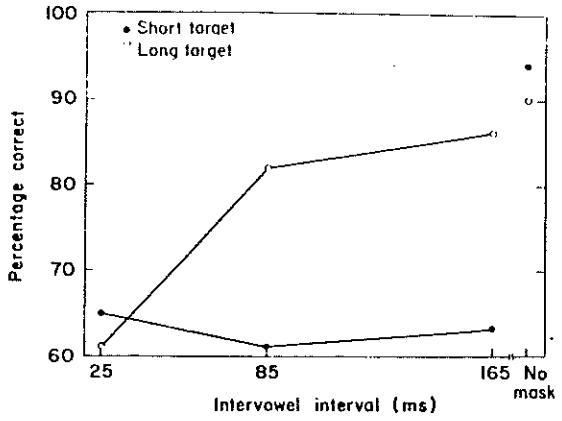

Figure 4 Percentage of correct identifications of the target vowels, as a function of the intervowel interval and the duration of the target

the IVI without ever reaching an symptotic level of performance. The long target vowel was recognized 4% more accurately in the no-mask condition than at the longest IVI of 165 ms. Performance on the short target vowel decreased between IVI's of 25 and 85 ms, then improved slightly at 165 ms. At the longest IVI of 165 ms a 31% performance decrement was observed, relative to the no-mask condition. While the main effect of target vowel duration was statistically non-significant, $F(1, 4) = 6.21$, $P > 0.1$, the target duration by IVI interaction was highly significant, $F(3, 2) = 10.51$, $P < 0.005$.

The effects of the IVI indicate that identification of the tense-lax distinction is not solely a function of target duration. For both the front and back target vowels, identification varied with processing time. At short IVIs both long and short target vowels would have short perceived durations, while with increases in the IVI both target vowels would come

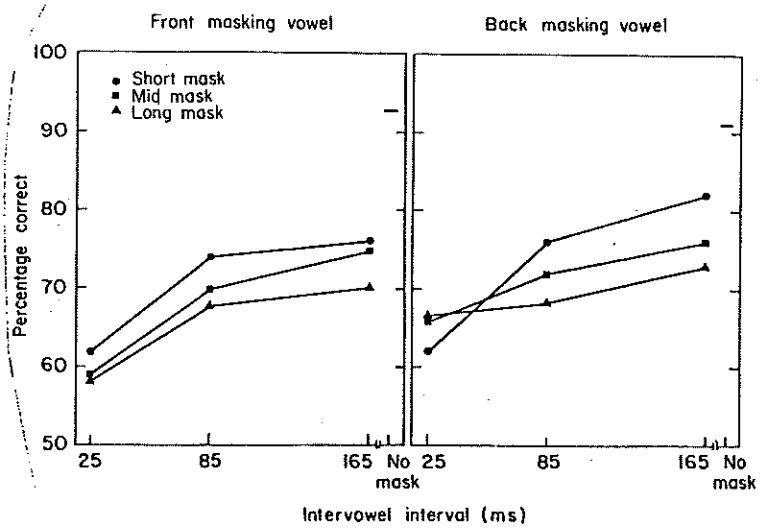

Figure 5 Percentage of correct identifications of the target vowel, as a function of the intervowel interval and the masking vowel. For clarity of presentation, the results for the two masking vowel identities are presented in separate panels (Experiment 1).

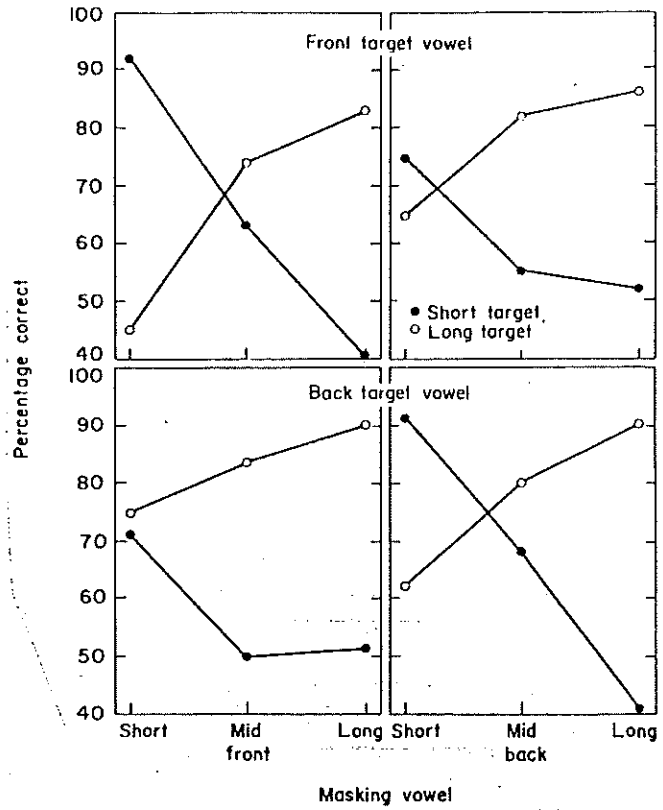

Figure 6 Percentage of correct identifications of each of the target vowel durations, as a function of both the quality of the target vowel and the duration of the masking vowel (Experiment 1).

to have increasingly longer perceived durations. Accordingly, both target vowels were classified as lax disproportionately often at short IVIs and as tense disproportionately often at long IVIs.

Figure 5 presents the percentage of correct identities of the duration of the target vowel, as a function of both the IVI and the masking vowel. The results for the two masking vowel qualities are presented in separate panels. The six masking vowels produced somewhat different results. Overall identification of the target vowel was 3% better for the back masking vowels than for the front masking vowels. For both masking vowel qualities, recognition of the duration of the target vowels was most accurate with the short masking vowels at least accurate with the long masking vowels. All six masking vowels yielded performance that improved monotonically with increases in the IVI, without ever reaching the level of performance observed in the no-mask condition. For the front masking vowels, similar performance was found across the IVI for all three masking vowel durations. For the back masking vowels, a much greater improvement was found for the short masking vowel at the short IVI's. Both the main effect of the IVI, $F(3, 12) = 31.54, P < 0.005$, and the masking vowel by IVI interaction, $F(10, 40) = 3.16, P < 0.025$, were significant. the main effect of masking vowel duration did not reach statistical significance, $F(5, 20) = 2.15, P > 0.2$.

Figure 6 presents the percentage of correct identifications of the duration of the individual target vowels, as a function of both the quality and duration of the target and masking

vowels. The accuracy with which the duration of the target vowel could be recognized was strongly contingent upon the duration of the masking vowel. The long masking vowel produced excellent performance on the long target vowel and quite poor performance on the short target vowel, while the short masking vowel produced excellent performance on the short target vowel and quite poor performance on the long target vowel. The magnitude of this effect was greater, however, when the target and masking vowels had the same formant structure. Both the target duration by masking vowel interaction $F(5, 20) = 12.85$, $P < 0.001$, and the target identity by target duration by masking vowel interaction, $F(5, 20) = 3.18$, $P < 0.05$, were significant.

The effects of the duration of the masking vowel also argue for the importance of the characteristics of a following speech segment in preception of the tense-lax distinction. The judged duration of the target vowel was longer for long than for short masking vowels. These variations in perceived duration caused both the front and back target vowels to be categorized more often as tense before a long masking vowel and as lax before a short masking vowel. The attenuated influence of masking vowel duration for masking vowels dissimilar to the target vowels in formant structure suggests that a quality difference along an irrelevant dimension allows a partial exclusion of the masking vowel from the judgmental process.

The results of Experiment 1 argue for the utility of perceived duration in identifying tense and lax vowels. The question is thus generated as to how duration can be used to discriminate between the tense and lax members of a similar pair. One possibility is that vowel quality is categorized with respect to some criterion. Vowels with durations shorter than this criterion would be considered lax; vowels with durations longer than this criterion would be considered tense. Under this conceptualization, being lax or tense would be a binary attribute of a vowel. An alternative possibility suggests that instead of a binary opposition, a continuum exists from lax to tense (e.g. Oden & Massaro, 1978). With increases in the perceived duration of a vowel, that vowel would grow progressively less lax and more tense, with categorization changing at some point along the continuum. Thus, in the former view both perception and identification are categorical, while in the latter view perception is continuous and only identification categorical.

The results of Experiment 1 cannot be used to distinguish between these alternative conceptualizations. Since the binary choice task requires categorization of the vowel as lax or tense, the quality of the percept will not necessarily be reflected in the response. This problem can be obviated, however, by use of a rating task (Idson & Massaro, 1977), which allows the subject to provide a direct index of his or her perceptual experience. If perception of the tense-lax quality of vowels is categorical, then no information should be provided by the rating response which is not contained in the categorization response. Alternatively, if perception is continuous, then the rating responses should yield changes in vowel quality within the two generic categories of lax and tense.

Experiment 2 replicated Experiment 1 in all respects except for the response mode. The subjects were now asked to rate the quality of the target vowel on a continuous scale from lax (/I/ or /U/) to tense (/i/ or /U/). If perception of the lax-tense distinction is categorized, then one of two patterns of results should be obtained. First, the curve over the IVI could be discontinuous. At short IVI's, the perceived durations of a target vowel should be less than the criterion duration and it should be categorized as lax. At some critical IVI the perceived durations of the target vowel should exceed criterion, after which it should be rated as tense. Second, the curves for the two target vowels could be continuous and divergent across the IVI. With increased processing time resolution of both target vowels will improve; the short vowel will come to be rated as increasingly lax and the long vowel as increasingly

tense. While these alternative patterns of results are quite different, they both reflect an underlying binary discrimination between tense and lax. The distinction lies in that in the former case accuracy of resolution is not dependent upon processing time, while in the latter case it is so dependent. In contrast, if a continuum exists from lax to tense at the perceptual level, with only identification being categorical, then the rating task should produce the same continuous function across the IVI for the two target vowels. The rating task does not require that the continuum of vowel qualities produced by changes in vowel duration be mapped into a binary categorization of lax and tense. Instead, the perceived quality of the target vowels can be responded to directly. Since perceived duration will increase continuously with increases in the IVI, the quality of both target vowels should become progressively less lax and more tense as the IVI is lengthened. Thus, the degree to which both target vowels are rated as tense should increase gradually with IVI.

Experiment 2

Method

Design

The experimental design was identical to that of Experiment 1.

Apparatus and stimuli

The same apparatus used in Experiment 1 was employed to generate the simuli and prövide the feedback. The responses were recorded on a continuous response scale, consisting of a pointer connected to a potentiometer and a push button. The subject set the pointer to the desired location along a 5.5 cm scale, and then pressed the pushbutton to indicate that a response had been made. The placement of the pointer was indicated by the voltage passed by the potentiometer to an analog-to-digital converter. The voltage given by the linear potentiometer allows the 5.5 cm scale to be mapped into 50 intervals from left to right, which were assigned the integer values from 0–49.

The stimuli had the same characteristics as those of Experiment 1.

Procedure

The procedure was identical to that of Experiment 1, with the exception of the response required. On each trial, the subject's task was to locate the target vowel along the continuum from /I/ to /i/ or /U/ to /u/, ignoring the masking tones. The response was made by placing the pointer at a location along the response scale. The left end of the scale was labeled "I" or "O" and the right end of the scale was labeled "E" or "U", depending upon the level of the between-subjects factor. Though only one target vowel quality, having two durations, occurred in the experiment, the subjects were instructed that they would hear a variety of target vowels, spanning the range between /I/ and /i/ or /U/ and /u/. They were asked to map this range of vowel qualities onto the response scale, utilizing the entire length of the scale. Subjects were instructed that the most /i/-like or /u/-like vowel would represent the rightmost endpoint of the scale, and the most /I/-like or /U/-like vowel the leftmost endpoint of the scale. Subjects were also instructed to attempt to have equal changes in vowel quality represented by equal distances along the scale. In a post-experiment questionnaire, no subject reported difficulty in using the scale in the required manner. A 3 s response interval was employed, timed from target onset. Though the subjects could begin responding as soon as the target vowel was presented, the reaction time always exceeded the longest IVI. Following the response interval, a 250 ms visual display of an asterick was presented, to indicate that the response interval was over and a new trial about to begin.

Subjects

The subjects were eight University of Wisconsin undergraduates, who participated as part of an introductory psychology course.

Results and discussion

The ratings of the target vowel, indicated by placement of the pointer along the response scale, were recorded as integers between 0 and 49. The value of 0 represents the most /I/-line or /U/-like vowels and the value of 49 represents the most /i/-like or /u/-like vowels. The mean ratings of vowel quality were computed for each subject, on each day, at each target-quality by target duration by masking vowel by IVI condition. These ratings were submitted an analysis of variance, which revealed the main effect of days, and all interactions involving days as a factor, to be statistically non-significant. Since performance was not changing as a function of practice, the ratings were pooled over the three experimental days. The pooled ratings were submitted to two separate analyses of variance, as in Experiment 1.

The between-subjects factor of target vowel quality had virtually no effect upon performance. Overall, the back target vowel was identified approximately 3% more poorly than the front target vowel. The main effect of target vowel quality, and all interactions involving target vowel quality as a factor, were statistically non-significant. These results indicate that performance was not changing as a function of the target vowel quality, nor were any of the three within-subject variables affected by this factor. Accordingly, all of the data will be discussed in terms of the average results for the two target vowel qualities. Higher ratings will be referred to as increasingly tense, to indicate that the vowels were judged as increasingly /i/-like or /u/-like. Lower ratings will be referred to as increasingly lax to indicate that the vowels were judged as increasingly /I/-like or /U/-like.

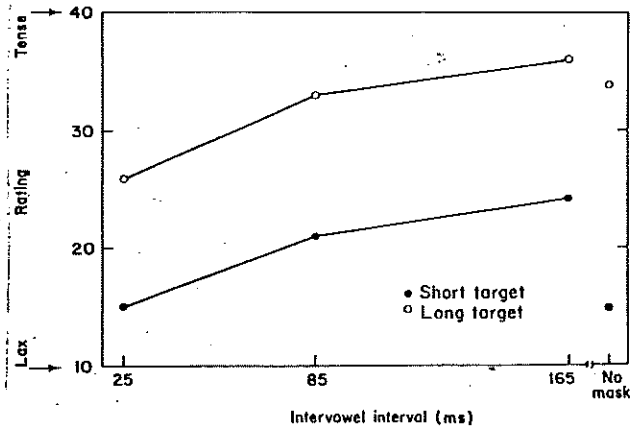

Figure 7 Ratings of each of the target vowel durations, as a function of the IVI (Experiment 2).

Figure 7 presents the ratings of the target vowels, as a function of the duration of the target vowel and the IVI. The long target vowel was always rated as being more tense, than the short target vowel. Both target vowels came to be rated as progressively more tense with increases in the IVI. The long and short target vowels were both rated as being more tense at an IVI of 165 ms than in the no-mask condition. Both the main effect of target vowel duration, $F(1, 6) = 10.81, P < 0.025$, and the interaction of this factor with the IVI $F(3, 18) = 3.38, P < 0.05$ were statistically significant. The influence of the IVI on the ratings suggests

that a continuum exists from lax to tense. As perceived duration increased with increases in the IVI, both the long and short target vowels came to be rated as progressively more tense.

Figure 8 presents the ratings of the target vowel, as a function of the IVI and the masking vowel. The results for the two masking vowel qualities are presented in separate panels. The

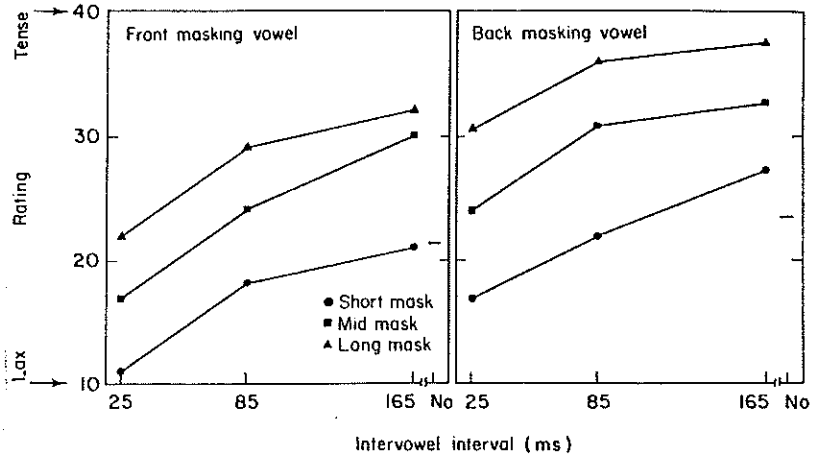

Figure 8 Ratings as a function of both masking vowel duration and the IVI (experiment 2).

overall ratings were approximately 3% higher on the back than on the front masking vowels. For both masking vowel durations, the target vowel was rated as increasingly more tense with increases in the duration of the masking vowel. The target vowel was also identified as more tense with increases in the IVI. Higher ratings were observed at the longest IVI of 165 ms than in the no-mask condition. Both the main effects of the masking vowel, $F(5, 30) = 3.95$, $P < 0.01$, and of the IVI, $F(3, 18) = 7.21$, $P < 0.005$, were significant. The interaction of these factors was not statistically significant.

Figure 9 presents the ratings of the target vowel, as a function of the duration of the target vowel, the quality of the target vowel, and the masking vowel. In contrast to Experiment 1, the quality of the target vowel had no effect upon the target duration by masking vowel duration interaction. The long target was always rated as more tense than the short target. Both target vowels came to be rated as increasingly more tense with increases in the duration of the masking vowels. Both the target duration by masking vowel interaction ($F < 1$) and the target duration by target identity by masking vowel interaction ($F < 1$) were statistically non-significant.

The results found for the masking vowels confirm the inferences drawn from Experiment 1. A masking vowel will increase the judged duration of a target vowel, causing both the long and short target vowels to be rated as more tense in the presence of a masking vowel than in the no-mask condition. The degree to which the masking vowel will lengthen the judged duration of the target vowel will be proportional to the actual duration of the masking vowel. Accordingly, the degree to which both the long and short target vowels were rated as tense increased directly with increases in the duration of the masking vowel.

In contrast to Experiment 1, the magnitude of the effect of masking vowel duration was not a function of the similarity between the formant structures of the target and masking vowels (replicating the absence of a similarity effect in a rating task in Massaro & Idson, 1978*a*). There is no obvious explanation for this disparity. One possibility, however, would

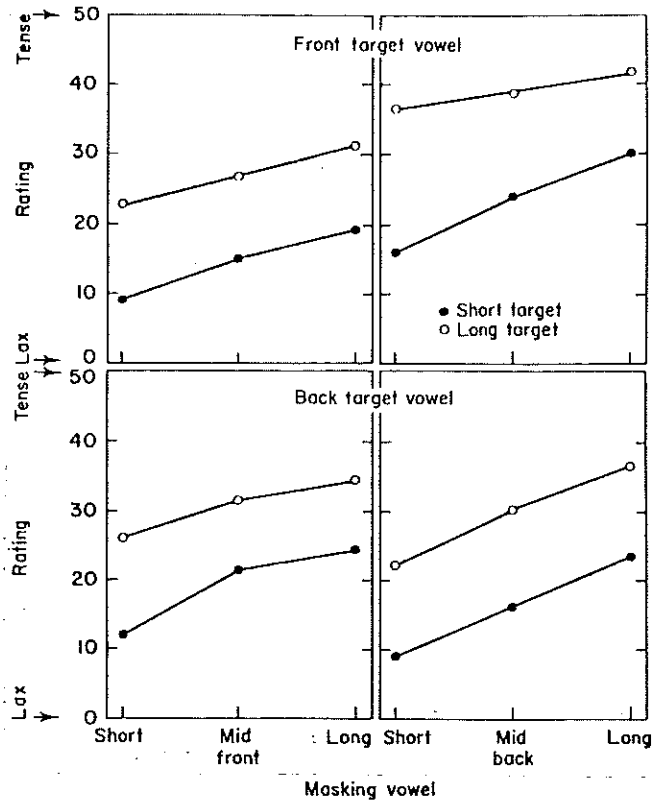

Figure 9 Ratings for each of the target vowel durations, as a function of both the duration of the masking vowel and the quality of the target vowel (Experiment 2).

reflect the differential demands of the rating and binary choice tasks. In the binary-choice task, the subject is attempting to make an accurate decision as to whether the target vowel is lax or tense. Accuracy would be facilitated by disregarding, in so far as is possible, the duration of the masking vowel. Dissimilarity in the format structure of the target and masking vowels would provide grounds for excluding the masking vowel from the judgemental process, accounting for the smaller influence of dissimilar masking vowels. In contrast, the rating task, by requiring a report of perceptual experience, places a premium not on accuracy but on reflecting all aspects of the perceptual experience. Accordingly, there would be no reason to exclude the masking vowel from the judgemental process and no effects of similarity would be expected.

General discussion

The present resaarch provides further evidence that duration offers a potent cue for discriminating between tense and lax vowels of a similar format structure (see also, Ainsworth, 1972). The target vowels in the current studies differed only in duration, and yet they could be discriminated in terms of the lax-tense distinction. Of greater interest, perceived duration appears to be primary for this distinction. Identification of the target vowels as lax or tense was shown to be a function of the same variables which influence perceived duration: the actual duration of the target vowels, the durations of the masking vowels, and

the intervowel intervals. Moreover, the nature of the influence exerted by these variables on the tense-lax distinction was quite comparable to the nature of the influence which they exert upon perceived duration. Perceived duration increases with increases in the IVI and the duration of the masking vowel. Similarly, a target vowel comes to be perceived as increasingly tense with longer IVIs and longer duration masking vowels. Thus, it appears that the tense-lax vowel quality distinction can be mediated by the perceived duration of a vowel.

A further implication of this research is that the lax-tense quality distinction represents a continuum rather than a binary opposition. Vowel quality changes gradually from lax to tense with the gradual changes in perceived duration engendered by increases in target duration, processing time, or masking vowel duration. Rather than being either lax or tense, a given vowel is both lax and tense to differeing degrees. It appears that vowels are identified in terms of two opposing classes only if an explicit categorization response is required.

The demonstrated importance of perceived duration for discriminating tense and lax vowels, may have interesting implications for the role of duration in other aspects of speech perception. Since perceived duration is inherently context dependent, and durational contrast would seem to be relative. If invariant physical duration were central, then a vowel could be identified in any context by comparing its duration to some absolute reference value. Given the primacy of perceived duration, identification of a vowel might proceed differently, by comparing the relative durations of the alternative tense and lax vowels in that similar context.

Somewhat more speculative inferences from this research are also possible. To the extent that vowel duration is to be used as a cue to vowel identity, a lax vowel must have a shorter duration than the associated tense vowel. Since perceived duration varies with available processing time, this perceptual demand imposes an important constraint on the structure of the speech signal: a vowel must occur in a context that yields an appropriate perceived duration. As a possible example, a stressed lax vowel can appear only in a closed syllable. One consequence of this restriction is that processing time for the lax vowel will always be limited by the occurrence of the following consonant, insuring that the perceived duration of that vowel will be short. Such examples at least suggest that certain duration rules may be motivated by the mechanisms involved in perception as well as those involved in production. All of these inferences are highly conjectural, and clearly require explicit investigation. Yet the fact that an analysis of duration perception yielded intriguing hypotheses suggests that it may be worthwhile to consider the nature of perception in evaluating the utility of duration in speech perception.

This research was supported by U.S. Public Health Service Grant MH-19399 to D. W. Massaro, and by both a University of Wisconsin Graduate Fellowship and NIMH small Research Grant MH-30983 to W. L. Idson. Preparation of the paper was facilitated by a University of Texas Research Institute Summer Fellowship to W. L. Idson. This work represents a collaborative effort; order of authorship is arbitrary. We would like to thank Jim Bryant for assistance with the computer programming.

References

Ainsworth, W. (1972). Duration as a cue in the recognition of synthetic vowels. *Journal of the Acoustical Society of America*, 51, 648–651.

Allan, L. G. (1976). Is there a constant minimum perceptual duration? *Quarterly Journal of Experimental Psychology*, 28, 71–76.

Allan, L. G. & Kristofferson, A. B. (1974). Psychophysical theories of duration discrimination. *Perception and Psychophysics*, 16, 26–34.

Allan, L. G. & Rousseau, R. (1977). Backward masking in judgments of duration. *Perception and Psychophysics*, 31, 482–486.

Bennett, D. C. (1978) Spectral form and duration cues in the recognition of English and German vowels. *Language and Speech*, 11, 65–85.
Cantor, N. E. & Thomas, E. A. C. (1976). Visual masking effects on duration, size and form discrimination. *Perception and Psychophysics*, 19, 321–327.
Cohen, M. M. & Massaro, D. W. (1976). Real-time speech synthesis. *Behavioral Research Methods and Instrumentation*, 8, 189–196.
Cole, R. A. & Cooper, W. E. (1975). The perception of voicing in English affricates and fricatives. *Journal of the Acoustical Society of America*, 58, 1280–1287.
Delattre, P. (1962). Some factors of vowel identity and their cross-linguistic validity. *Journal of the Acoustical Society of America*, 34, 1141–1143.
Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, 27, 761–764.
Efron, R. (1970*a*). Effects of stimulus duration on perceptual onset and offset latencies. *Perception and Psychophysics*, 8, 231–234.
Efron, R. (1970*b*). The measurement of perceptual duration. *Studium Generale*, 23, 558–561.
Efron, R. (1970*c*). The minimum duration of perception. *Neuropsychologia*, 8 57–63.
Gol'dburt, S. N. (1961). Investigation of the stability of auditory processes in micro-intervals of time (new findings in back masking). *Biophysics*, 6, 809–817.
Hawkins, H. L., Thomas, G. B., Presson, J. C., Cozic, A. & Brookmire, D. (1974). Tonal specificity and masking in auditory recognition. *Journal of Experimental Psychology*, 103, 530–538.
House, A. S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America*, 33, 1174–1178.
House, A. S. & Fairbanks, G. (1953). The influence of consonantal environment upon the secondary acoustic characteristics of vowels. *Journal of the Acoustical Society of America*, 25, 105–113.
Idson, W. L. & Massaro, D. W. (1977). Perceptual processing and experience of auditory duration. *Sensory Processes*, 1, 316–337.
Kallman, H. J. & Massaro, D. W. (in press). Similarity effects in backward recognition masking. *Journal of Experimental Psychology: Human Perception and Performance.*
Klatt, D. H. (1971). Generative theory of segmental duration in English. *Journal of the Acoustical Society of America*, 51, 101 (A).
Klatt, D. H. (1975). Vowel lenghening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3, 129–140.
Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208–1221.
Klatt, D. H. & Cooper, W. I. (1975). Perception of segment duration in sentence contexts. In *Structure and process in speech perception.* (Cohen, A. & Nooteboom, S. eds). New York: Springer-Verlag.
Kristofferson, A. B. (1977). A real-time criterion theory of duration discrimination. *Perception and Psychophysics*, 21 105–117.
Lehiste, I. (1970). *Suprasegmentals.* Cambridge: M.I.T. Press.
Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, 7, 107–122.
Lehiste, I., Olive, J. P. & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, 60, 1199–1203.
Martin, J. G. (1970). On judging pauses in spontaneous speech. *Journal of Verbal Learning and Verbal Behavior* 9, 75–78.
Massaro, D. W. (1970). Preperceptual auditory images. *Journal of Experimental Psychology*, 102, 199–208.
Massaro, D. W. (1974). Perceptual units in speech recognition. *Journal of Experimental Psychology*, 102, 199–208.
Massaro, D. W. (1975). *Experimental psychology and information processing.* Chicago: Rand-McNally.
Massaro, D. W. & Cohen, M. M. (1975). Preperceptual auditory storage in speech recognition. In *Structure and process in speech perception.* (Cohen, A. & Nooteboon, S. G. eds.). New York: Springer-Verlag.
Massaro, D. W. & Cohen, M. M. (1977). The contribution of voice-onset time and fundamental frequency as cues to the /zi/–/si/ distinction. *Perception and Psychophysics*, 22 373–382.
Massaro, D. W., Cohen, M. M. & Idson, W. L. (1976). Recognition masking of auditory lateralization and pitch judgments. *Journal of the Acoustical Society of America*, 59, 434–441.
Massaro, D. W. & Idson, W. L. (1976). Temporal course of perceived auditory duration. *Perception and Psychophysics*, 20, 331–352.
Massaro, D. W. & Idson, W. L. (1978*a*). Temporal course of received vowel duration. *Journal of Speech and Hearing Research*, 21, 37–55.
Massaro, D. W. & Idson, W. L. (1978*b*). Target-mask similarity in backward recognition masking of perceived tone duration. *Perception and Psychophysics*, 24, 225–230.
Moore, J. J. & Massaro, D. W. (1973). Attention and processing capacity in auditory recognition. *Journal of Experimental Psychology*, 99, 49–54.
Nooteboon, S. G. (1973). The perceptual reality of some prosodic duration. *Journal of Phonetics*, 1, 25–45.

Oden, G. C. & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, 85, 172–191.

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

Raphael, L. J. (1972). Preceding vowel duration as a cue to the voicing characteristics of word-final consonants in English. *Journal of the Acoustical Society of America*, **51**, 1296–1303.

Thomas, E. A. C. & Cantor, N. E. (1975) On the duality of simultaneous time and size perception. *Perception and Psychophysics*, 00, 44–48.

Thomas, E. A. C. & Cantor, N. E. (1976) Simultaneous time and size perception. *Perception and Psychophysics*, **19**, 353–360.