Time's Role for Information, Processing, and Normalization[a]

DOMINIC W. MASSARO

Program in Experimental Psychology
University of California, Santa Cruz
Santa Cruz, California 95064

> *. . . What then is time? If someone asks, I know.*
> *If I wish to explain it to someone who asks,*
> *I know not . . .*
> —ST. AUGUSTINE

> *As a rule sensations outlast for some little time the*
> *objective stimulus which occasioned them.*
> —WILLIAM JAMES[1]

INTRODUCTION

Time serves multiple roles in our interaction with our environment. The most obvious role is in terms of information; the duration of an event provides a cue to the identity of the event. We will consider examples of duration information in the domain of speech perception. Our concern will be with how duration cues in speech are evaluated and integrated in the classification of speech patterns. Second, time is necessary for perceptual processing and restricts memory of auditory events. We will briefly review the study of perceptual processing time and the time course of preperceptual auditory memory. Finally, we will address the issue of normalizing the information available to the perceptual process. The important question will be how duration information of a speech segment is evaluated relative to the duration of the context.

DURATION INFORMATION IN SPEECH

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 identifying information for a spoken utterance. At the segmental level, for example, duration can distinguish long and short vowels as in "bit" and "beet"[2]; and vowel and consonant duration are cues to voicing of a postvocalic consonant.[3-5] Duration information also serves as a supplementary cue to lexical tone in Mandarin Chinese.[6] At the suprasegmental level, word and phrase duration influence such factors as the placement of a constituent boundary of an otherwise ambiguous phrase.[7] For example, the ambiguous 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.[8]

[a] This work was supported in part by Grant MH-35334 from the National Institute of Mental Health.

Vowel Identification

In English, about half of the vowels are relatively short in duration and the other vowels are relatively long. It is only natural that listeners might utilize duration as a feature for vowel identification. The importance of vowel duration as an acoustic feature has been demonstrated by Ainsworth.[9] Listeners identified synthetic vowels differing in the first and second formant and duration. The number of correct identifications of the normally shorter vowels decreased as the vowel durations were increased. The converse held true for identification of the normally longer vowels.

An important question is how duration information is combined with other information about vowel identity. Duration can function as a supplemental feature to vowel identification, particularly in distinguishing between vowels with similar formant patterns. Bennett[10] independently varied the duration and formant levels of vowels in words using synthetic speech. In one set, the formant patterns were changed in four steps from the word "shut" to the word "shirt." The formant patterns of these two vowels are fairly similar. For each of the four formant patterns, the word was synthesized with four different vowel durations. Listeners identified each of the 16 stimuli as either "shut" or "shirt." The left panel of FIGURE 1 gives the proportion of "shirt" identifications as a function of the formant structure of the vowel; duration is the curve parameter. Changes in the formant structure changed the identification from one alternative to the other. Duration had a significant, but smaller effect; its effect was most noticeable at the second level of the formant structure. We might assume that the second level of formant structure was relatively ambiguous and, therefore, the influence of duration made more apparent.

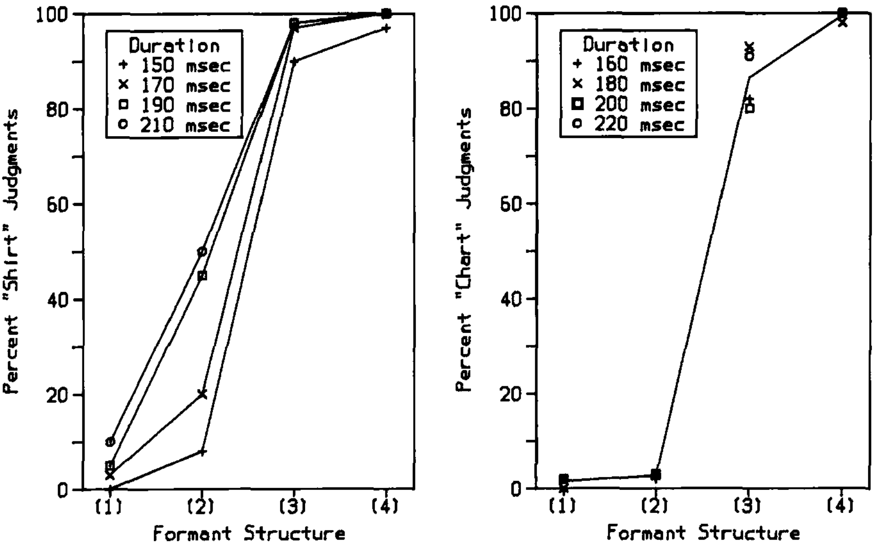

FIGURE 1. Predicted and observed percentage of vowel identification as a function of the formant values; duration of the vowel is the curve parameter. (Data from Bennett.[10])

Bennett[10] replicated this experiment, using the words "chat" and "chart," respectively. The vowels in these two words differ in duration in the same way as the vowels in "shut" and "shirt," but normally have a greater difference between their formant patterns. Therefore, a four-stimulus continuum between "chat and chart" is unlikely to create an ambiguous level of formant structure. As can be seen in the right panel of FIGURE 1, vowel duration had no effect. The judgments were not ambiguous at any of the four levels of formant structure. We expect that a duration effect might have been observed if additional formant levels were created to produce more ambiguous judgments. Taken together, the results of the two experiments show that duration influences vowel identification primarily when the formant structure is ambiguous. Vowel duration has the potential of being a relevant feature in vowel recognition when it is necessary to identify vowels with relatively ambiguous patterns. The ambiguity could be either in the speech signal itself or could result from some limitation in the processing of the formant patterns.

The results of these two experiments can be described by a fuzzy logical model of perceptual recognition that will be described in the next section. The formant structure and duration are treated as independent acoustic features defining the vowel alternatives. The predictions of the results given by the model are given by the lines in FIGURE 1. The model gives a good description of the results and quantitatively describes the larger contribution of duration at the more ambiguous level of formant structure.

Integrating Multiple Durations in Speech Perception

Both vowel duration and consonant duration provide information about consonant voicing in a vowel–consonant syllable.[3-5] According to the fuzzy logical model,[28] speech perception involves feature evaluation, prototype matching, and pattern classification operations. Vowel duration and consonant duration are assumed to be independent cues at feature evaluation. The information available at feature evaluation is assumed to be continuous rather than categorical, and is represented by fuzzy-truth values between zero and one. During prototype matching, the cue values are inserted into prototypes in long-term memory and integrated (conjoined) together to arrive at a goodness-of-match of the stimulus with each prototype. For the present speech contrast, prototypes for the voiced and voiceless alternatives are defined as the conjunction (&) of vowel duration (V) and consonant duration (C) information. In addition, negation is realized by the unary complement.

voiced: long vowel and NOT (long consonant)

$$= V \,\&\, (1 - C) \tag{1}$$

voiceless: NOT (long vowel) and long consonant

$$= (1 - V) \,\&\, C \tag{2}$$

Subjects evaluate the degree to which the vowel and consonant durations are long and enter these values into the prototype definitions. The cues are treated independently and given fuzzy-truth values between zero and one. The conjunction operation is defined as multiplication.[11] To arrive at a particular judgment, the relative goodness-of-match is taken. In this case, the probability of a voiced judgment. P(Voiced), is equal to

$$P(\text{Voiced}) = \frac{V \times (1 - C)}{[V \times (1 - C)] + [(1 - V) \times C]}. \tag{3}$$

The idea of independent consonant and vowel duration cues should not be interpreted as a statistical independence (noninteraction) of the effects of these two variables.[29] Multiplying the cues together at the prototype matching operation does not violate the principle of independent cues at the feature evaluation operation. The feature value of one cue remains independent of the feature value of another cue, even though the conjunction of two or more cues is, by definition, an interactive process. Since the outcome of the conjunction depends on both values, the cues can interact statistically in their effects on performance.

In contrast to the description of the fuzzy logical model, other studies have been taken as evidence for the ratio of the consonant duration to the vowel duration as the critical cue to consonant voicing.[3] Port and Dalby[12] (experiment 2) asked subjects to identify synthetic speech stimuli as "digger" or "dicker." The initial vowel duration and the medial consonant silent closure duration were orthogonally varied in a factorial design. If the consonant/vowel (C/V) ratio is the critical stimulus parameter for voicing of the medial stop, then the likelihood of a voiced response should vary *only* as a function of this ratio; the actual durations of the vowel and consonant closure should not matter. Port and Dalby[12] concluded that the judgments were determined primarily by the C/V ratio, but did not contrast this conclusion against other alternatives.

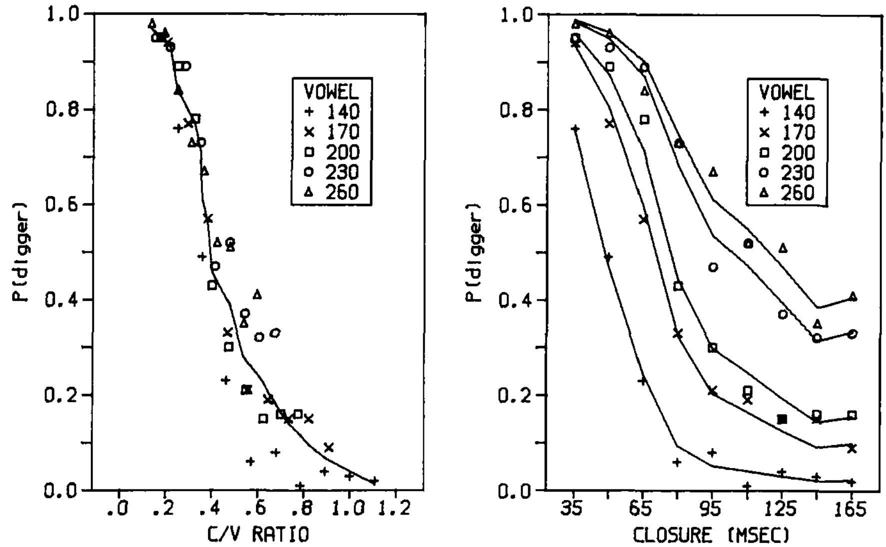

FIGURE 2. *Left panel:* Proportion of "digger" judgments as a function of the closure/vowel ratio. *Right panel:* Proportion of "digger" judgments as a function of closure duration and vowel duration. (Data from Port and Dalby.[12]).

In addition to testing the fuzzy logical model, Massaro and Cohen[13] quantified the C/V ratio model and tested it against Port and Dalby's results. Both models were quantified with the same number of free parameters, which were estimated by finding those values that minimized the squared deviations between the predicted and observed results for each subject using the minimization routine STEPIT.[14] The measure of goodness-of-fit used to compare the two models was the root mean squared deviation (RMSD) between predicted and observed values. As can be seen in FIGURE 2, independent vowel duration and consonant duration cues gave a much better description of the results than did a C/V ratio (RMSD of .030 versus .078).

Massaro and Cohen[13] also demonstrated that the fuzzy logical model gave a better description than did the C/V ratio model for the individual subject identification and rating results for the /juz/ – /jus/ distinction.[4] Thus, vowel duration and consonant duration appear to provide independent cues to voicing of the consonant and the integration of the cues is adequately described by the fuzzy logical model.

Actual Duration, Perceived Duration, and Feature Values

It is reasonable to assume that perceived duration, P, is a linear function of physical duration, D.

$$P = KD \qquad (4)$$

The slope, K, of the linear function might depend on a number of factors; for example, Derr and Massaro[4] found that the slope varied with the fundamental frequency pattern of the vowel. Rising or falling patterns increased the perceived duration relative to steady-state patterns (see also Lehiste[15]).

We might also expect that the slope of the function is dependent on surrounding context. Perceived duration of a given segment of sound might be longer to the extent that the surrounding segments are short. A contrast effect of this type seems to hold for both speech and nonspeech stimuli.[16,17]

Perceived duration cannot be identical to the feature values of duration cues in speech. A given perceived duration will have different consequences, depending on the segment of sound being perceived and the distinction being cued. Consider how

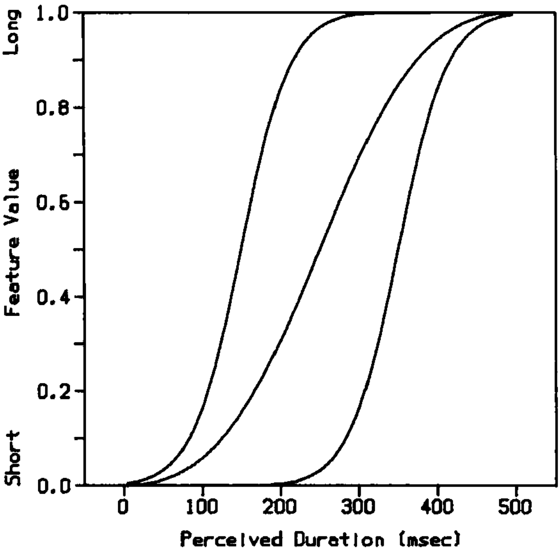

FIGURE 3. Three functions describing changes in the feature value, F, with changes in perceived duration, P.

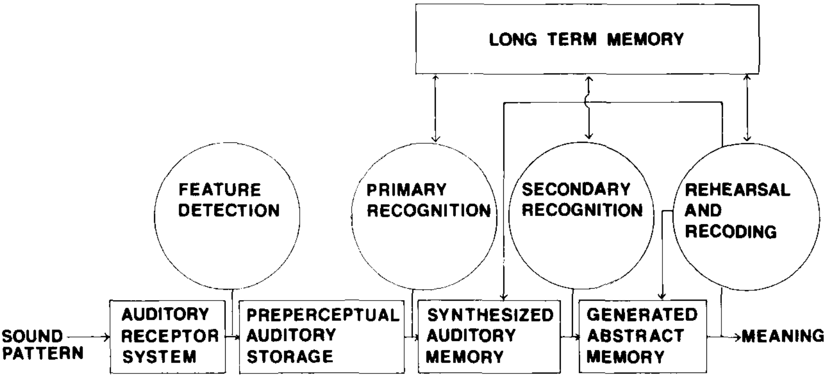

FIGURE 4. An auditory information-processing model. (After Massaro.[19])

changes in perceived vowel duration might cue the voicing of a following stop in a *VC* syllable. For very short or very long durations, a given change might not produce much of a change in the feature value. For intermediate durations, however, the same amount of change might produce a larger change in cue value. This relationship between perceived duration and feature value can be described by an ogival function relating cue value *F* and perceived duration *P* given by Equation 4

$$F = \frac{X^y}{X^y + (1 - X)^y}, \tag{5}$$

where $y > 1$. The y parameter determines the steepness of the middle section of the ogive. The values of *X* are assumed to be a linear function of perceived duration, *P*. Thus, $X = aP + b$, except that values of *X* less than zero are set to zero and values greater than one are set to one. Equation 5 is ogival in form when the *X* values fall between zero and one. The parameters a and b allow the ogive to shift horizontally along the dimension of perceived duration. FIGURE 3 plots three different ogival functions describing changes in *F* with changes in *P*. This relationship was used successfully to describe the change in the voicing feature value of a vowel–consonant syllable as a function of vowel duration and changes in fundamental frequency.[4]

PERCEPTUAL PROCESSING TIME

The model shown in FIGURE 4 has been utilized in our research in auditory information processing and speech perception.[18,19] The feature detection process transforms the sound wave pattern into acoustic features that are stored centrally in a preperceptual auditory storage. This storage accumulates the acoustic features until the sound pattern is complete and holds the features for a short time after a sound is presented. The primary recognition process resolves the featural information, which produces the phenomenological outcome of perceiving a particular sound of a particular loudness, quality, and duration, at some location in space.

The perceptual processing during the stimulus and a short period following is responsible for perceptual recognition and experience of the stimulus. A second pattern does not usually occur until the first pattern has been perceived. However, if the second pattern is presented before recognition of the first pattern is complete, the second pattern might interfere with recognition of the first. By varying the delay of the second pattern we can determine the duration of preperceptual auditory storage and the temporal course of perceptual recognition. The experimental task is referred to as a backward recognition masking paradigm. Pure tones differing in frequency, intensity, waveform, duration, and spatial location have been employed as test items in the backward masking task.[20,21]

Backward Masking

Consider a typical backward masking task. Consonant–vowel syllables /ba/, /da/, and /ga/ presented at a normal listening intensity were used as test and masking stimuli.[22] Only enough of the syllable was presented to make it sound speech-like. The 42-msec syllable had 30 msec of the consonant–vowel transition plus 12 msec of steady-state vowel. On each trial, one of the three syllables was presented followed by a variable silent interval before presentation of a second syllable chosen from the same set of three syllables. The subject's task was to identify the first syllable as one of the three alternatives. Subjects were told to ignore the second syllable, if possible.

FIGURE 5 plots the observed results in terms of discriminability (*d′*) values for each of the three test alternatives. The *d′* measure provides an index of how well the subject discriminates a given test alternative from the other test alternatives in the task. Correct identification of the first speech sound increased with increases in the silent interval between the two sounds. This result shows that recognition of the consonant phoneme was not complete at the end of the consonant–vowel transition or even at the end of the short vowel segment of the sound. Syllable recognition required perceptual processing after the speech sound was terminated. These results support the idea that the speech sound is held in preperceptual auditory storage while processing takes place. A second sound interferes with additional processing of the first sound.

Perceiving Speech

The role of backward masking in the perception of speech is nicely illustrated in the identification of intervocalic stop consonants (*VCVs*). FIGURE 6 presents a stylized

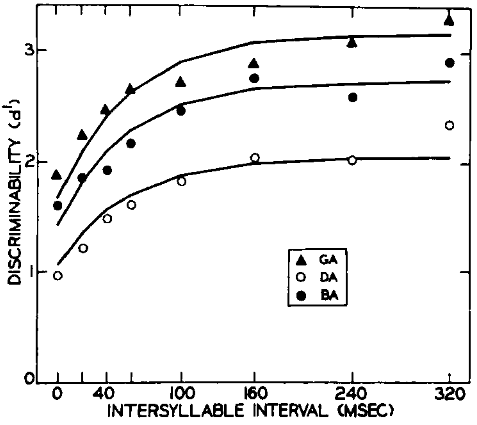

FIGURE 5. Predicted and observed discriminability of the test syllables as a function of the silent processing time available before the onset of the masking syllable. (After Massaro.[22])

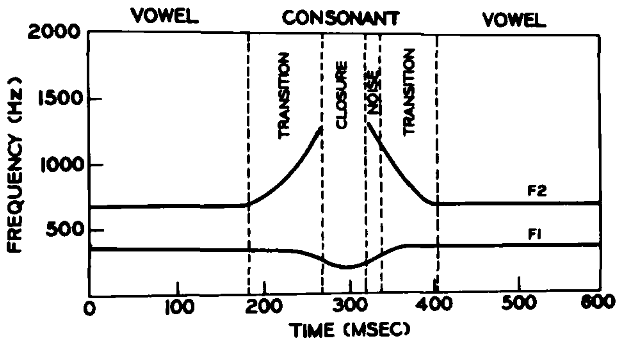

FIGURE 6. Stylized representation of a *VCV* sequence with a voiced-stop consonant containing steady-state vowel formants, transitions, closure, noise burst, and steady-state vowel formants.

spectrogram of a *VCV* with a voiced stop consonant. The steady-stated vowel is followed by a transition into the consonant, a closure period representing the closure of the vocal tract, a transition into the final vowel, and the final steady-state vowel. We refer to the segment before the closure as the *VC* and the segment after the closure as the *CV*. The *VCV* speech stimulus is an interesting one because of its redundancy and noncorrespondence with perceptual experience. We tend to hear a single consonant between the two vowels, although either the segment before closure or the segment after closure is sufficient when presented alone to produce the perception of the consonant.

Why doesn't the listener hear two consonants when processing a *VCV* speech stimulus? Following the logic of backward masking, we might expect the *CV* to "backward-mask" the processing of the *VC* to the extent the closure duration is short. Given insufficient processing time for the *VC* transition because of the quickly following *CV* transition, the *CV* transition has the largest impact on the perceptual experience.[19] More recent evidence for backward masking in perceiving speech comes from Repp.[23] Subjects had to discriminate normal *VCV* stimuli from modified *V–CV* stimuli in which the *VC* transitions were eliminated and replaced with the steady-state vowel. To the extent that the *CV* masks perceptual recognition of the segment preceding the closure, subjects should not be able to discriminate the *VCV* from the *V–CV* stimuli. The results showed that discrimination was poor at short closure durations and improved dramatically with increases in the closure interval. Repp also synthesized stimuli with contradictory *VC* and *CV* transitions; subjects had to discriminate /ab-di/ from /ad-di/ or /ab-bi/ from /ad-bi/. The results showed typical backward-masking effects. A *CV* segment can terminate processing of an earlier *VC* segment so that the *CV* segment provides the primary cues for the perception of the intervocalic consonant. Thus, the time available for processing plays an important role in perceiving speech.

Perceiving Time

The perceived duration of a stimulus is a function not only of its physical characteristics, but also of the amount of time spent processing it. A given stimulus, having constant physical characteristics, will have a variety of perceived durations, dependent upon the amount of available processing time. Massaro and Idson[24] used a backward-masking recognition paradigm, in which one of two target tones of differing durations was presented on each trial, followed after a variable silent interval by a masking tone of one of three durations. The subject identified the target tone as being long or short. The top panel of FIGURE 7 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 function of the intertone interval. This result was obtained for all three masking-tone durations. Thus, the average masking function for duration reflects similar processes to those occurring for other stimulus dimensions.

The bottom panel of FIGURE 7 illustrates that the short and long targets give different functions. Identification of the long target increases, whereas the short target gives a decreasing function with increases in the intertone interval. Massaro and Idson[24] accounted for these results within the model illustrated in FIGURE 4. The perception of duration is assumed to occur in essentially the same manner as for other

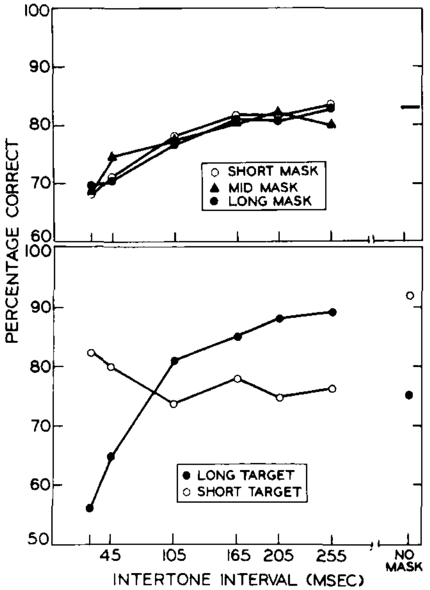

FIGURE 7. *Top panel:* Percentage of correct identifications of the target as a function of the duration of the mask and the intertone interval. *Bottom panel:* Percentage of correct identifications of each of the target tones as a function of the intertone interval. (From Massaro and Idson.[26])

attributes. However, the added assumption is made that the perceived duration of the target increases with the silent processing time after target presentation. That is, both discriminability and perceived duration are assumed to increase with additional processing time. If sufficient time is allowed for complete processing of the target, its perceived duration will be directly related to its temporal extent. If less than complete processing time is available, the target will be perceived as having a duration shorter than its asymptotic value. Therefore, with 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 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 exactly the results which were obtained (compare FIGURE 7).

It is also assumed that the mask adds to the perceived duration of the target. The addition is proportional to the mask duration. This assumption was supported by the strong target–mask interaction found in these studies. To the extent that the target and mask were both short, performance was better than if the target was short and the mask was long. A quantification of the model provided a good description of the perception and discrimination of tone and vowel duration.[21,24–26]

NORMALIZATION PROCESSES IN SPEECH PERCEPTION

It is now well documented that perceptual recognition of a segment of speech can be influenced by the duration of surrounding segments. Nooteboom[27] demonstrated a boundary shift between the long and short Dutch vowels /a/ and /a:/ with changes in speech rate. Longer durations were required for hearing the long vowel alternative as the speech phrase preceding the word containing the vowel was made slower. Normalization phenomena reveal the *relative* nature of acoustic cues in speech perception: The feature value of a given acoustic property must be normalized relative to other local context. In the fuzzy logical model, the context duration might lead to a normalization of acoustic feature values. This influence could result from an actual change of the psychophysical relationship between physical duration and perceived duration or from a more cognitive change of the function relating perceived duration to the feature value. Consider a typical mapping between physical duration and a feature value. For a given feature value, a shorter physical duration is necessary at a faster speech rate relative to a slower rate. As an example, it is well known that a relatively long silent closure interval cues a voiceless stop in a *VCV* context. A longer silent closure duration would be necessary for a given value of this cue to the extent that the stop is embedded in a context of a slow speaking rate.

Normalizing Vowel and Consonant Duration Cues

Port and Dalby[12] evaluated the role of speech tempo in the perception of the voicing distinction between the words "rapid" and "rabid." A trained speaker recorded the sentence "I'm trying to say 'rabid' to you," at a fast tempo and a slow tempo. The word "rabid" was removed from both sentences and cut immediately after closure of the medial /b/ and just before the release. The glottal pulsing during the closure interval was deleted and silent periods were inserted to create a continuum varying from 50 to 200 msec in steps of 10 msec. These modifications of the two versions of "rabid" were now embedded in both the fast and slow original sentences and presented to listeners for identification. Thus, listeners heard the word "rabid" taken from a fast or slow sentence and presented with one of 16 closure intervals embedded in either a fast or slow sentence. Listeners simply identified whether they heard "rabid" or "rapid" in each of the 64 possible test sentences.

FIGURE 8 gives the results of the experiment. In addition to the large effect of closure duration, both speaking rate of the original test word and the context sentence had significant influences on the identification judgments. The percentage of "rabid" judgments decreased with increases in the closure duration. There were more voiceless judgments for the test words originally spoken at a fast tempo relative to those spoken at a slow tempo. In addition, a longer closure was necessary to hear "rabid" in a fast than in a slow sentence. The predictions of the fuzzy logical model are also given in FIGURE 8. The derivation of these predictions is carried out in the next section.

A Model of Normalization

We define the two prototypes in the "rabid"–"rapid" identification task as

rabid: long vowel and NOT(long closure)

$$= V \times (1 - C) \tag{6}$$

rapid: NOT(long vowel) and long closure

$$= (1 - V) \times C, \tag{7}$$

where V and C stand for long vowel and long closure, respectively.

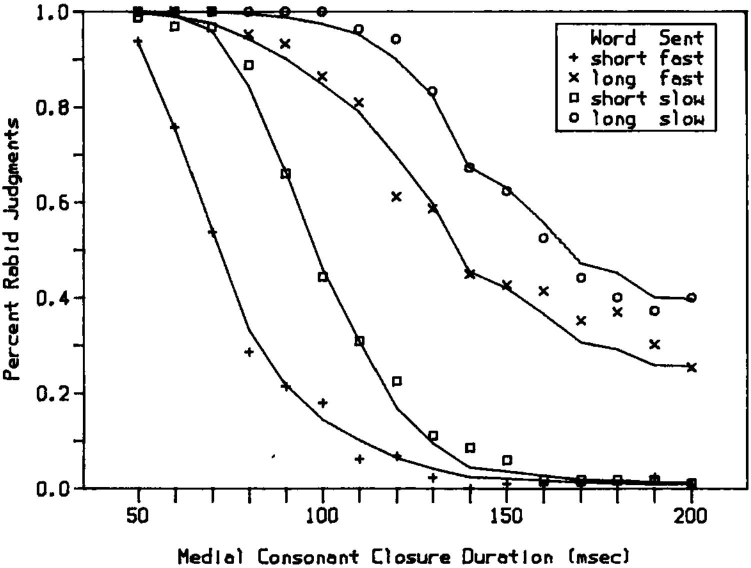

FIGURE 8. Predicted and observed identifications of word "rabid" as a function of closure duration; vowel duration and sentence duration are the curve parameters. (Data from Port and Dalby.[12])

To quantify the contribution of sentence tempo, a parsimonious assumption is that each feature value cueing duration is modified appropriately as a function of sentence tempo. A fast sentence rate will give a value more towards the long alternative for a given feature than will a slower sentence rate. The prototypes given by Equations 6 and 7 can be considered to be the representation for a normal rate; faster or slower speech rates will be represented by exponents on the feature values. In this case, the values V and C would be raised to exponents a and b, respectively.

The different exponents a and b simply allow different degrees of normalization for different features. The exponents will vary only with sentence rate. The exponents should be less than one if the sentence is faster than normal. As an example, a 0.7 value of V will give a .84 value when the exponent is .5. The slower sentence context has just the opposite influence and requires an exponent greater than one. An exponent of 2 reduces a .7 value to .49, resulting in a shift away from the long alternative.

The likelihood of a voiced response, P(rabid), would be based on the value of match to the rabid prototype relative to the sum of the values of prototype matches for rabid and rapid.

$$\mathrm{P(rabid)} = \frac{V^a \times (1 - C^b)}{[V^a \times (1 - C^b)] + [(1 - V^a) \times C^b]} \tag{8}$$

The predictions of Equation 8 for Port and Dalby's results are given in FIGURE 8. The model provides an adequate description of the normalization results with an RMSD of .027.

CONCLUSION

We have discussed three roles of time in auditory information processing: the information provided by time, the time necessary for information processing, and the normalization of processing due to time. These important aspects of auditory information processing can be described within a general auditory information-processing model. The model was developed to account for dimensions of our auditory experience other than the temporal one and thus auditory time offers a strong test of the basic assumptions of the model. The critical features of the model include memory structures and functional processes, the evaluation and integration of auditory information, and the classification of sound patterns. The model accounted for a wide variety of temporal phenomena contributing to our auditory experience. Thus, auditory time provides additional evidence for both the general information-processing model and the fuzzy logical model developed primarily in nontemporal domains of auditory perception.

ACKNOWLEDGMENTS

Michael M. Cohen provided valuable assistance at all stages of the project.

REFERENCES

1. JAMES, W. 1980. The Principles of Psychology. Holt. New York, NY. (Reprinted by Dover Publications. New York, NY.)
2. PETERSON, G. E. & I. LEHISTE. 1960. Duration of syllable nuclei in English. J. Acoust. Soc. Am. **32:** 693–703.
3. DENES, P. 1955. Effect of duration on the perception of voicing. J. Acoust. Soc. Am. **27:** 761–764.
4. DERR, M. A. & D. W. MASSARO. 1980. The contribution of vowel duration, FO contour, and frication duration as cues to the /juz/–/jus/ distinction. Percept. Psychophys. **27:** 51–59.
5. RAPHAEL, L. J. 1972. Preceding vowel duration as a cue to the voicing of the voicing characteristic of word-final consonants in American English. J. Acoust. Soc. Am. **51:** 1296–1303.
6. TSENG, C., D. W. MASSARO & M. M. COHEN. Lexical tone perception: Evaluation and integration of acoustic features in Mandarin Chinese. Unpublished manuscript.

7. LEHISTE, I. 1973. Phonetic disambiguation of syntactic ambiguity. Glossa: 107–122.
8. LEHISTE, I., J. P. OLIVE & L. A. STREETER. 1976. Role of duration in disambiguating syntactically ambiguous sentences. J. Acoust. Soc. Am. **60:** 1199–1202.
9. AINSWORTH, W. A. 1972. Duration as a cue in the recognition of synthetic vowels. J. Acoust. Soc. Am. **51:** 648–651.
10. BENNETT, D. C. 1968. Spectral form and duration cues in the recognition of English and German vowels. Lang. Speech **11:** 65–85.
11. MASSARO, D. W. & M. M. COHEN. 1976. The contribution of fundamental frequency and voice onset time to the /zi/ – /si/ distinction. J. Acoust. Soc. Am. **60:** 704–717.
12. PORT, R. F. & J. DALBY. 1982. Consonant/vowel ratio as a cue for voicing in English. Percept. Psychophys. **32:** 141–152.
13. MASSARO, D. W. & M. M. COHEN. 1983. Consonant/vowel ratio: An improbable cue in speech. Percept. Psychophys. **33:** 501–505.
14. CHANDLER, J. P. 1969. Subroutine STEPIT finds local minima of a smooth function of several parameters. Behav. Sci. **14:** 81–82.
15. LEHISTE, I. 1976. Influence of fundamental frequency patterns on the perception of duration. J. Phonet. **4:** 113–117.
16. MILLER, J. L. & A. M. LIBERMAN. 1979. Some effects of later-occuring information on the perception of stop consonant and semivowel. Percept. Psychophys. **25:** 457–465.
17. CARRELL, T. D., D. B. PISONI & S. J. GANS. 1980. Perception of the Duration of Rapid Spectral Changes: Evidence for Context Effects with Speech and Nonspeech Signals. Research on Speech Perception, Progress Report 6. Department of Psychology, Indiana University, Bloomington, Indiana.
18. MASSARO, D. W. 1975. Experimental Psychology and Information Processing. Rand-McNally. Chicago, IL.
19. MASSARO, D. W. 1975. Understanding Language: An Information Processing Analysis of Speech Perception, Reading and Psycholinguistics. Academic Press. New York, NY.
20. KALLMAN, H. J. & D. W. MASSARO. 1983. Backward masking, the suffix effect, and preperceptual auditory storage. J. Exp. Psychol. Learn. Mem. Cog. **9:** 312–327.
21. IDSON, W. L. & D. W. MASSARO. 1977. Perceptual processing and experience of auditory duration. Sensory Processes **1:** 316–337.
22. MASSARO, D. W. 1974. Perceptual units in speech recognition. J. Exp. Psychol. **102:** 199–208.
23. REPP, B. H. 1978. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. Percept. Psychophys. **24:** 471–485.
24. MASSARO, D. W. & W. L. IDSON. 1976. Temporal course of perceived auditory duration. Percept. Psychophys. **20:** 331–352.
25. IDSON, W. L. & D. W. MASSARO. 1980. The role of perceived duration in the identification of vowels. J. Phonetics **8:** 407–425.
26. MASSARO, D. W. & W. L. IDSON. 1978. Target-mask similarity in backward recognition masking. Percept. Psychophys. **24:** 225–236.
27. NOOTEBOOM, S. G. 1982. Speech rate and segmental perception or the role of words in phoneme identification. *In* The Cognitive Representation of Speech. T. Myers, J. Laver, & J. Anderson, Eds. North-Holland. Amsterdam.
28. ODEN, G. C. & D. W. MASSARO. 1971. Integration of featural information in speech perception. Psychol. Rev. **85:** 172–191.
29. ANDERSON, N. H. 1974. Information integration theory: A brief survey. *In* Contemporary Developments in Mathematical Psychology, Vol. 2. D. H. Krantz, R. C. Atkinson & P. Suppes, Eds. Freeman. San Francisco, CA.