Psychological Review
1972, Vol. 79, No. 2, 124–145

PREPERCEPTUAL IMAGES, PROCESSING TIME, AND PERCEPTUAL UNITS IN AUDITORY PERCEPTION [1]

DOMINIC W. MASSARO [2]

University of Wisconsin

The present paper provides a theoretical account of the auditory recognition process. The theory describes recognition in terms of the information in a preperceptual auditory image and the time it is available for perceptual processing. Auditory recognition processes are assumed to be analogous to those operating in visual recognition. Necessary distinctions are drawn between auditory detection, recognition, and short-term memory. Studies of recognition provide direct support for a preperceptual auditory image that outlasts the sensory input. The processing of the preperceptual auditory image corresponds to a readout of the information available in a temporal or perceptual unit of information. Studies of speech perception support these conclusions. The syllable, not the phoneme, is implicated as the perceptual unit for speech perception. Thus, a framework is provided for the recognition stage of auditory information processing. This stage of perceptual processing outputs a synthesized percept that is utilized by succeeding stages of cognitive processing.

The perceptual process is characterized by the temporal course of identification or recognition. Recognition of a stimulus requires an analysis and synthesis of the information available in the sensory input. In vision, the visual image during an eye fixation keeps the information available for the recognition process. In contrast, an auditory input continuously changes over time and the information in the stimulus might not remain available. However, if the auditory information was held in a preperceptual auditory store, auditory perception might also involve a readout of the auditory image of the stimulus. The important characteristics of the auditory image will, of course, differ from those in the visual image. The major difference between the two images appears to be the critical dimension of the stimulus necessary for feature recognition. Whereas the spatial pattern is the important dimension in visual stimuli, the sequential pattern is critical in audition.

The temporal course of auditory perception is analyzed here in the framework of processes found in visual perception (Haber, 1969; Neisser, 1967). Similar processes may occur in auditory and visual information processing even though few similarities are found between visual and auditory psychophysics. The lack of correspondence between the modalities is mainly due to the different psychophysical characteristics of light and sound. For example, nothing exists in acoustics similar to color mixing in vision. It is not surprising that psychophysical approaches to the study of the two senses have little in common.

Once the information is available in the preperceptual image, the similarities between visual and auditory perception become noticeable. Similar to the visual stimulus, the auditory input must contain distinctive features that characterize a meaningful pattern. These features should be determined by the microstructure of the auditory input; that is, modulations of sound pressure over time. However, the features cannot be recognized as they arrive, since this requires that perception be immediate. Rather, it is assumed that a temporal unit of the auditory stimulus is stored in a preperceptual auditory store

[1] This work was supported by Grant MH 19399-01 from the National Institute of Mental Health, United States Public Health Service. I would like to thank John Barresi and Richard L. Venezky for helpful discussions. I also appreciate comments on particular points by Gordon Bear, Ronald A. Cole, Allan L. Fingeret, John Morton, Gordon Redding, John Theios, and Lucinda Wilder.

[2] Requests for reprints should be sent to Dominic W. Massaro, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

for processing. A visual stimulus produces an image that can persist after the stimulus is terminated (Averbach & Coriell, 1961; Haber, 1969; Kahneman, 1968; Neisser, 1967; Sperling, 1960, 1967). The preperceptual image preserves the information for recognizing the meaningful pattern (Haber, 1970). Correspondingly, the auditory image is needed for feature recognition of the temporal pattern.

Since the time of the Gestalt psychologists (Hochberg, 1970; Koffka, 1935), form and organization have played an important role in the description of the perception of visual inputs. When a series of dots is organized to form a circle, the units or features for processing are curvilinear segments rather than dots. Similarly, with an auditory preperceptual store, fluctuations of sound pressure temporally organized as a syllable should determine the appropriate acoustic distinctive features (e.g., Jakobson & Halle, 1956).

A number of questions are assumed to be important in auditory perception. What stimulus variables determine the formation of a perceptual unit of auditory information? Given that there is a unit, what influences its size? In speech, for example, is the unit a phoneme, syllable, or word? Further, if the unit of auditory information persists in some preperceptual auditory storage system, what is the life span of the image? What is its vulnerability to stimuli that precede or follow the stimulus in time? Are the features of the acoustic pattern processed in parallel or sequential order? Finally, considering the psychological moment theory of perception (Stroud, 1955), are perceptual units determined by some internal clock or does the stimulus input determine the unit?

The present paper provides a theoretical account of auditory perceptual processing. Although the theory directs itself to the recognition process, auditory detection and short-term memory must also be considered. The two main assumptions of the theory describe the perceptual process in terms of the information in the sensory input and the time the information is available for perceptual processing. The first assumption is that an auditory input produces a preperceptual auditory image that contains the information in the stimulus. The preperceptual image can be thought of as a perceptual unit of information. Therefore, the auditory image persists beyond the stimulus presentation and preserves its sequential information. The second assumption is that the recognition process entails a readout of the information in the preperceptual auditory image. This readout takes time and is referred to as the temporal course of perceptual processing. The time required for perceptual processing is directly related to the complexity of the identification task.

If a short auditory stimulus produces a preperceptual auditory image, a second stimulus should interfere with this image and, therefore, interfere with perception. Accordingly, the literature on auditory masking provides the first tests of the assumption of preperceptual auditory images. Auditory masking refers to any observation that information in a test auditory stimulus is reduced by presentation of another masking auditory stimulus. The present discussion is concerned with temporal masking conditions in which the test and masking stimulus do not coexist in time. Backward masking refers to the paradigm of following the test stimulus presentation by a masking stimulus. Forward masking refers to the paradigm of preceding the test stimulus presentation with a masking stimulus. Kahneman (1968) recently reviewed the visual masking literature, and Raab (1963) provides a review of masking in all sensory modalities. Although Raab recommends investigating "analogous" phenomena in the different senses, he does not discuss any similarities in detail.

Auditory detection refers to the psychological experience that an acoustic stimulus is or was present (Galanter, 1962). In auditory recognition, the subject identifies which of several alternatives matches the auditory input. Although it can be argued that detection and recognition have a number of properties in common, there are important differences between these two

processes. As an example, any language spoken at a normal intensity is easily detected but impossible to understand without the appropriate knowledge in long-term memory. Similar to visual masking (Haber, 1969; Kahneman, 1968; Raab, 1963), the auditory masking literature also supports the distinction between detection and recognition.

Auditory Detection Masking

A number of studies have been carried out studying backward masking of auditory detection (Deatherage & Evans, 1969; Elliot, 1962a, 1962b, 1967; Homick, Elfner, & Boothe, 1969; Raab, 1963; Samoilova, 1959). For example, Elliot (1967, Experiment I) employed a backward masking paradigm to study the development of auditory frequency contours. Durations of 10 and 100 milliseconds (msec.) were used for the test tone and masking noise, respectively. The steady-state level of the masking noise was 70-decibels (db.) sound pressure level (SPL). Thresholds for the detection of the test tone were computed at backward masking intervals of 10 to 100 msec. Backward masking decreased with increases in the masking interval. Very little masking was observed when the masking noise followed the test tone by 100 msec.

Elliot's (1967) results are supported by a number of recent detection masking experiments (Deatherage & Evans, 1969; Homick et al., 1969). The results indicate that a loud masker occurring within 100 msec. of a test stimulus presentation increases the detection threshold of the test stimulus. Further, the detection threshold is inversely related to the duration of the silent interval between the test and masking stimuli. However, these studies of backward masking do not necessarily provide evidence for a preperceptual auditory store. Rather, the masking stimulus could overtake the test stimulus in the auditory pathways decreasing the signal-to-noise ratio and, therefore, decreasing detection performance. In detection masking studies, the masking stimulus is much louder than the test stimulus. The question of interest, then, is how the loudness of a stimulus relates to the time it takes to be detected. McGill (1961, 1963) has demonstrated an inverse relationship between stimulus loudness and simple reaction time. Assuming that simple reaction time to a tone contains independent detection, decision, and response components (Donders, 1969; Sternberg, 1969), stimulus loudness should only affect the detection component. Therefore, it is reasonable to account for the decision and response components and to measure relative detection time as a function of stimulus loudness.

McGill (1961) presented a subject with a 1,000-hertz (Hz.) tone of various random intensities in a simple detection task. McGill's data indicated that increasing the amplitude of the test tone from 30- to 100-db. SPL decreased the median reaction time from 216 to 120 msec. This difference, which represents differences in detection time, approximates 100 msec. which is exactly the value usually found as the effective duration of backward detection masking. Therefore, the temporal course of backward detection masking could be equal to the difference in the detection times of the test and masking stimuli. If the masking stimulus overtakes the test tone in the auditory pathways, the difference in reaction times to the test and masking stimuli should equal the temporal course of backward masking.

Most quantitative studies of backward detection masking have employed test tones and masking noise as stimuli. Simple reaction times to noise of a certain intensity may not equal that of a tone of the same intensity. Keeping this in mind, it still may be helpful to interpret the effect of backward masking in terms of the detection times given by McGill (1963). Homick et al. (1969) studied backward masking of a tonal signal as a function of the intensity of the masking noise. The results indicated that increasing the masking noise from 70-db. to 90-db. SPL increased the temporal course of backward masking. For example, the 90-db. masker produced about the same amount of masking as the 70-db. masker if the 90-db.

masker was delayed an extra 20 msec. For example, subjects could detect a given test signal if there was silent interval of 30 msec. before the 70-db. masking noise. However, with a masking noise of 90 db. subjects needed a silent interval of 50 msec. for correct detection of the test signal. Accordingly, it is possible that the 70-db. noise takes about 20 msec. longer to be detected than the 90-db. noise. McGill's relative reaction times to 70- and 90-db. tones support this. It took McGill's subjects about 20 msec. longer to make a simple response to a 70-db. tone than to a 90-db. tone. The results indicate that backward detection masking occurs because the second masking stimulus overtakes the test stimulus and decreases the signal-to-noise ratio.

The results of backward detection masking do not provide evidence for the existence of a preperceptual auditory store. The present interpretation of backward detection masking must now be reconciled with a different account of forward detection masking. Forward masking of detection has also been demonstrated in a number of different experiments (Deatherage & Evans, 1969; Elliot, 1962a, 1962b; Homick et al., 1969). Although Elliot (1962a, 1962b) found less forward masking than backward masking, Homick et al. (1969) found at least as much forward masking as backward masking. It was argued that backward detection masking occurs because the masking stimulus overtakes the test stimulus decreasing the signal-to-noise ratio. In forward masking, this cannot be the case since the louder masking stimulus is presented first and would be detected before the test stimulus. Since the masking stimulus is terminated before the test tone is presented, no masking should take place unless an auditory image of the masking stimulus outlasts its presentation.

Forward masking can be accounted for by assuming that an auditory image outlasts the masking stimulus presentation. Results that will be presented later provide direct demonstrations of an auditory image that remains after a stimulus is presented. The results also indicate that the auditory image that outlasts the stimulus does not differ qualitatively from the image during the stimulus presentation. Therefore, the auditory image of a forward masking stimulus should interfere with detection of a test stimulus in the same way as a simultaneous masking stimulus (Green & Swets, 1966). Accordingly, a forward masking stimulus interferes with detection, since the auditory image of the masking stimulus decreases the signal-to-noise ratio of the test stimulus.

Erikson and Johnson (1964) have been credited with demonstrating that preperceptual echoic memory might last as much as 10 seconds (sec.) (Bryden, 1971; Neisser, 1967). Erikson and Johnson (1964) presented subjects with a barely detectable tone while they were engaged in reading a novel. The reading light was turned off either simultaneously with the tone or from 1 to 10 sec. after the tone presentation. To keep the subjects honest, the light was also turned off at times when no tone had been presented. At the offset of the light, the subjects had to report "whether or not the tone had sounded sometime within the last 10–15 sec. [p. 30]." The results indicated that the ability to correctly report whether a tone had been presented decreased with increases in the tone–test interval. However, subjects could still reliably report a tone presentation 10 sec. after its presentation. Neisser (1967) concludes from these results that "If we assume that attention to the novel precluded any encoding the of 'beep' we can infer that recall in this procedure must have been based on echoic memory alone [p. 204]."

However, Neisser's (1967) assumption cannot be justified and probably could easily have been disproved by Erikson and Johnson's own subjects. A phenomenological report could have indicated whether subjects noticed the tone when it was presented. Furthermore, at the offset of the light did they search for an auditory image or for a recent awareness or covert identification of the tone presentation? As Massaro (1970b) points out, the recall decision might have been based on memory

for the covert identification rather than on storage of the preperceptual image. Fortunately, one of the experimental conditions helps resolve the issue. Introducing a roving frequency tone sometime after the tone presentation did not decrease accuracy of report. In contrast, the masking results discussed below indicate that preperceptual store is easily disrupted by subsequent auditory inputs. Therefore, the reports of Erikson and Johnson's subjects were not based on a preperceptual image present at the time of the test.

Forward and backward detection masking studies do not provide direct evidence for the existence of a preperceptual auditory store. Furthermore, since most auditory inputs outside the laboratory are easily detected, the detection masking studies are not directly relevant to the temporal course of processing familiar auditory information. Quantifying the temporal course of masking would be relevant to the temporal course of processing auditory information only if the masking stimulus terminated processing of that information rather than prevented the detection of that information. For example, a second input in speech does not prevent detection of the first, but could terminate the perceptual processing of the first. This phenomenon would be similar to one in vision in which an eye movement while reading is sufficient to erase the earlier pattern. As in vision, the detection masking paradigm is not appropriate for studying the temporal course of auditory identification or recognition. Accordingly, a recognition masking task was devised to determine the properties of the auditory information storage system and to quantify the temporal course of processing auditory information (Massaro, 1970b, 1971).

Auditory Recognition Masking

In the recognition paradigm, the subject first learns to identify or recognize

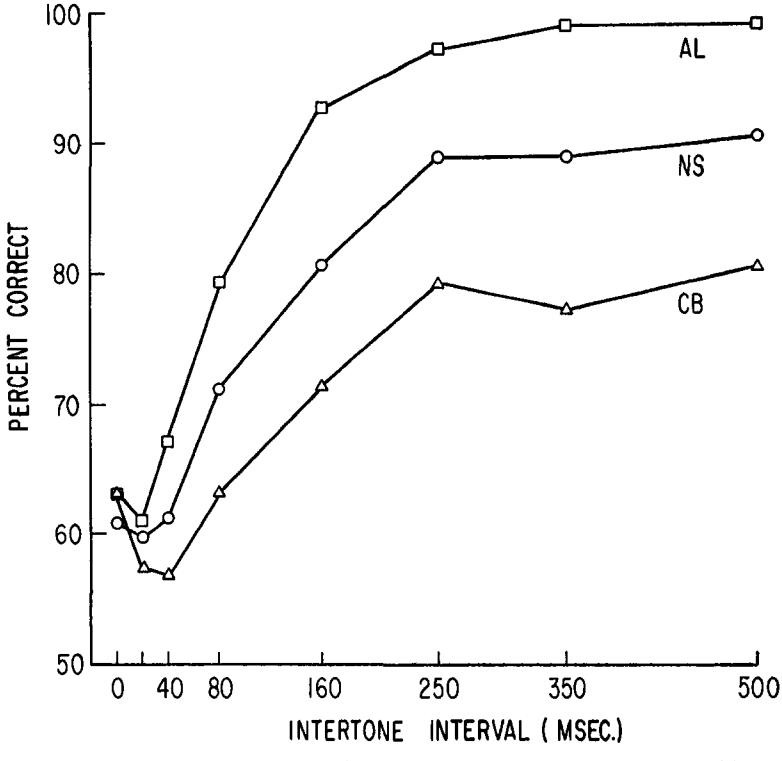

FIG. 1. Percentage of correct identifications of the test tone for subjects AL, NS, and CB as a function of the duration of the silent intertone interval.

two or more test signals. For example, the test signals could be two short tones differing in pitch. The subject's task is to identify the higher tone as high and the lower tone as low. In the backward masking paradigm, one of the test tones is presented followed by a silent interval followed by a masking tone. The test and masking tones are presented at the same loudness so that the masking stimulus will not overtake the test tone as in backward detection masking studies.

Massaro (1970b) presents an experiment that can serve as a prototype for discussion. In this study, one of two pure tones (a 20-msec. sine wave of 770 or 870 Hz.) was presented and the subjects' task was to identify the tones as low and high, respectively. The masking tone was equal to 820 Hz. All tones were presented at 81 db. The silent intertone interval lasted 0, 20, 40, 80, 160, 250, 350, or 500 msec. The masking tone lasted 500 msec.

Figure 1 shows that for each of the three subjects in the task, recognition performance improved with increases in the silent intertone interval up to 250 msec. Further increases in the silent interval beyond 250 msec. did not significantly facilitate recognition performance. These results provide information about the preperceptual auditory image of the test tone, the vulnerability of the auditory image to new inputs, and the temporal course of recognition. Given that the test tone lasted only 20 msec., some preperceptual image must have remained for perceptual processing necessary to improve recognition performance with increases in the silent intertone interval. This same result indicates that the masking tone terminated perceptual processing of the image. Since recognition performance levels off at about 250 msec., the image probably decayed within this period. These results also indicate that the subjects required at least 250 msec. of silence besides the 20 msec. of the test tone presentation for the recognition process.

Differences between Detection and Recognition Masking

A number of comparisons indicate that the backward masking observed in the recognition paradigm reflects different processes than those assumed to operate in detection masking. First, Elliot (1967) and Homick et al. (1969) have shown that the amount of backward detection masking is dependent on the frequency similarity of the test tone to the center of the masking noise. The results indicated that at all masking intervals, increasing the similarity of the narrow band masking noise to the test tone increased the total amount of masking. This result agrees with the assumption that the masking stimulus in detection masking decreases the signal-to-noise ratio. Simultaneous masking results also indicate that increasing the similarity of the test and masking stimulus decreases detection performance. In contrast, Massaro (1970b) found that the disrupting effect of the masking tone was not dependent on the frequency similarity of the masking tone to the test tone. This result agrees with the assumption that the masking tone in the recognition task terminates perceptual processing of the auditory image.

Second, presenting the masking tone contralateral to the test tone presentation does not decrease backward masking in the recognition paradigm (Massaro, 1970b). This result indicates that the preperceptual auditory image is located centrally rather than at the peripheral level. In contrast, a dichotic stimulus produces very little backward masking in the detection paradigm (Deatherage & Evans, 1969; Elliot, 1962a, 1962b). In agreement with simultaneous detection masking (Deatherage & Evans, 1969; Fletcher, 1953), a dichotic masker does not reduce the signal-to-noise ratio and detection is disrupted very little.

Third, although foward masking decreases detection performance significantly (Homick et al., 1969), Massaro (1970b) found no forward masking in the recognition paradigm. It was concluded earlier that forward masking of detection was due to the decreased signal-to-noise ratio pro-

duced by the auditory image. In contrast, forward masking does not occur in recognition, since the new image of the test tone terminates processing of the image of the masking stimulus. This occurs since both stimuli in the recognition task are equal in loudness. In contrast, the test stimulus in forward detection masking does not terminate processing of the image of the masking stimulus, since the test stimulus is much softer than the masking stimulus. The comparisons between detection and recognition masking show that detection masking is a phenomenon that is closely tied to the psychophysical processes operating in simultaneous masking studies. Recognition masking, on the other hand, is closely tied to the temporal course of perceptual processing.

The recognition masking results indicate that a short tone presentation leaves an auditory image which decays very rapidly. Further, this image can be processed for correct identification. A second new stimulus interferes with perceptual processing necessary for correct recognition. Two important questions raised by the recognition masking results need to be answered. First, what is the nature of the image that remains after presentation of a short auditory stimulus? Second, how does the masking stimulus interfere with perceptual processing of the image?

Preperceptual Auditory Images

Using a paradigm similar to one employed by Sperling (1967) and Haber and Standing (1969) for visual stimuli, Efron (1970a, 1970b, 1970c) provides an analysis of the perceived duration of tones and noise. In this paradigm, subjects judge the temporal overlap between two stimuli presented sequentially. The subjects state whether or not the onset of the second index stimulus (e.g., a light) occurred before the offset of the first auditory stimulus. The interval between the offset of the auditory stimulus and the onset of the index stimulus is adjusted until a subject perceives offset–onset simultaneity. The independent variable of interest is the duration of the first auditory stimulus. The results indicate that the minimal perception of an auditory stimulus lasts about 130 msec. Decreasing the duration of the first auditory stimulus below 130 msec. increased its perceived duration by a similar amount. For example, if the first noise burst lasted 30 msec., the subject did not perceive a temporal interval between its offset and the onset of the index stimulus until there was an interstimulus interval of 100 msec. Efron's (1970a, 1970b, 1970c) results support the assumption of preperceptual auditory images. If a short tone produces an auditory image for perceptual processing, the subject should be able to estimate its duration. Finally, analogous to visual information processing (Haber & Standing, 1969; Sperling, 1960, 1967), the auditory afterimage does not differ qualitatively from the image during the stimulus presentation.

Plomp (1964) studied the rate of decay of auditory sensation in a two-alternative task. He presented a 200-msec. pulse of noise followed by another noise burst either immediately or after a silent interval. Given that a noise burst produces an afterimage that decays over time, the subject should not notice a blank interval before a second softer burst of noise. Further, the length of the blank interval that goes unnoticed should increase with intensity differences between the two noise bursts. Plomp showed that a time interval of 2.6 msec. is noticed 75% of the time between two noise bursts of 65 db. In contrast, when the second noise burst is reduced to 15 db., the blank interval must be increased to 78 msec. to be noticed 75% of the time. This result indicates that the auditory image of the first noise burst decayed to 23% of its original value in 78 msec. The two-alternative task employed by Plomp controls for decision biases, whereas Efron's (1970a, 1970b, 1970c) paradigm does not. Therefore, it seems necessary to employ Plomp's task while varying the duration of the first noise burst to provide another test of Efron's results that the duration of the afterimage is inversely related to the duration of the test stimulus.

Some Russian studies have also provided evidence for an auditory image that remains after a short tone is presented (Gol'dburt, 1961). Gol'dburt has shown that a second tone can shorten the perceived duration of an earlier short tone. Perceived duration of the first tone increases with increases in the silent interval between the two tones. This result indicates that the auditory image extends the apparent presentation time of a short tone. A second tone can interfere with the auditory image and decrease the perceived duration of the first tone. Gol'dburt (1961) has also shown that the effect of the second tone decreases with increases in the duration of the first tone. This indicates that analogous to the visual image (Haber & Standing, 1969), the duration of an auditory image is inversely related to the duration of the stimulus producing the image. If perceptual processing usually takes about 250 msec., there would be no need for an image to remain if the presentation time of the stimulus exceeds this value. Processing the information in the stimulus seems to be sufficient to eliminate any afterimage of the stimulus presentation.

Von Békésy (1971), simulating backward inhibition in concert halls, measured the perception of a sound that was followed by a second sound in a different location. He presented a 1,000-Hz. tone for 35 msec. over a single speaker followed by a 1,500-Hz. tone over a ring of speakers surrounding the single speaker. If the second tone is presented 60 msec. after the onset of the first tone, the second tone reduces the perceived duration and loudness of the first tone. Von Békésy states,

> But if we assume that every stimulus starts a process in the brain which lasts perhaps 200 milliseconds, we can make backward inhibition acceptable if we further suppose that this process can be inhibited at any moment during the 200-millisecond interval by the onset of the second stimulus [p. 530].

The studies of Efron (1970a, 1970b, 1970c), Plomp (1964), Gol'dburt (1961), and von Békésy (1971) indicate that the auditory image remaining after a short stimulus does not differ significantly from the image present during the stimulus presentation. A short auditory stimulus usually requires perceptual processing that outlasts the life of the stimulus. The auditory image remaining after the stimulus presentation contains the necessary information for perceptual processing. Since the afterimage does not differ from the image present during the stimulus, the subject mistakes the afterimage for the image that is present during the stimulus presentation. Accordingly, the subject overestimates the duration of a short stimulus. With a longer stimulus, the subject has time to process the information available during the stimulus presentation and does not overestimate the duration of the stimulus.

In a different approach to measuring the duration of preperceptual auditory images, Guttman and Julesz (1963) repeated an identical section of wide-band noise. The authors varied the duration of the section and measured this effect on periodicity perception. It is difficult to judge whether this paradigm is appropriate for determining the duration of preperceptual auditory images. However, the results and an interpretation will be presented for completeness. If a section of white noise is repeated above 19 cycles per second (cps), the periodicity is heard as pitch. Between 4 and 19 cps, motorboating is heard. In these two cases, the stimulus sounds smooth or homogeneous within a period. Although listeners can hear a whooshing in the range of 1–4 cps, an intraperiod roughness is heard. Which of these estimates qualify for estimating the duration of auditory images? If subjects perceive a change within a period, they have already synthesized some of the preperceptual image. Therefore, more than the preperceptual image is responsible for the perception of whooshing at 1–4 cps. The motorboating at 4–19 cps sounds homogeneous within a period. The duration of preperceptual images should be measured with the restriction that the noise section sound homogeneous within a period since synthesis has not taken place. Employing this criterion, the results indicate 250 msec. as the

maximal duration for preperceptual auditory images.

Massaro (1972) has provided another demonstration that the auditory image does not differ qualitatively from the image present during the stimulus presentation. In the recognition masking paradigm, the question of interest was which variable predicts tone recognition best: test stimulus duration or processing time. Although increasing the duration of a tone had a slight facilitatory effect on identification performance, processing time was most critical for accurate identification. That is, with processing time held constant, increasing the duration of the test tone had very little effect on identification performance. On the other hand, increasing processing time improved identification performance independent of the duration of the test tone. In terms of a perceptual processing model (Massaro, 1970a), the rate of processing the information in a test stimulus presentation did not differ as a function of following a short tone presentation by a silent period or by simply leaving the test tone on for the processing interval.

Kahn and Massaro [3] have shown that the afterimage of the test tone must have acoustic properties similar to the image of the test tone presentation. They reasoned that it was logically possible that the masking tone simply distracted the subject's attention and reduced recognition performance. If masking was due to switching of attention, the processing of a nonauditory stimulus should also function as a masking stimulus in the tone recognition task. To test this, light and tone masks were employed in the tone identification task. The subjects were also required to identify the duration (long or short) of the masking stimulus to make certain that they would attend to the masking stimulus when it was presented.

Following the test tone with a masking tone replicated earlier studies indicating that identification performance improved with increases in the silent intertone interval. On the other hand, when the test tone was followed by a masking light, test tone identification did not depend on the interstimulus interval and was at an asymptotic level. The accurate identification of the duration of the masking stimuli indicated that the subjects processed the masking stimuli. Therefore, these results support the conclusion that the auditory image is stored with the acoustic features of the test stimulus presentation and these features are not lost by a simple switch in attention. A second auditory input interferes with recognition by interfering directly with the auditory information available from the earlier test stimulus presentation.

The results support the first assumption that an auditory input produces a preperceptual auditory image that contains the information in the auditory stimulus. The image persists beyond the stimulus presentation and preserves its acoustic information. The second assumption is that the recognition process entails a readout of the information in the preperceptual auditory image. The recognition process takes time and is referred to as the temporal course of perceptual processing. The following discussion provides an analysis of the time required for perceptual processing.

Perceptual Processing Time

An experimental paradigm has been developed to establish the minimal duration necessary to perceive a repeated sequence of auditory items. Warren, Obusek, Farmer, and Warren (1969) have shown that subjects need at least 300 msec. of each of four nonverbal stimuli for accurate perception of their temporal order when the items are repeated in a loop. The experimental task involves identification of a repeated sequence of four successive sounds (e.g., high tone, buzz, low tone, and hiss). Although inexperienced subjects need much longer, experienced subjects cannot accurately identify the correct temporal order unless each stimulus lasts at least 300 msec.

The fact that four repeated digits (Warren & Warren, 1970) or four repeated vowels (Thomas, Hill, Carrol, & Bien-

[3] B. J. Kahn and D. W. Massaro. Backward recognition masking: Interference or shift of attention. Unpublished manuscript, 1971.

venido, 1970) can be recognized at much shorter durations (125–200 msec.) suggests that there may be some qualitative differences in perception of speech and nonspeech material. On the other hand, as Thomas et al. point out, the time to reach a decision for familiar sounds (speech) would be expected to be much less than the time required for unfamiliar nonspeech stimuli. A direct comparison of the processing times required for speechlike and nonspeechlike material while controlling for familiarity (i.e., discriminability) has not been reported. It seems unlikely that the temporal course of pattern recognition would be qualitatively different for speech and nonspeech material if the familiarity and distinctiveness of the material are taken into account.

A measure of processing time in speech might be found in the duration of vowels in normal speech which are in the range of 150 to 350 mesc. (Fletcher, 1953; House, 1961). Since the pattern of sound pressure changes in a steady state vowel repeats at the speaker's fundamental frequency (Fletcher, 1953), the extended duration of the vowel might be needed for processing the information available in the vowel presentation. Vowels at much shorter durations can be identified if followed by a silent retroactive interval (Gray, 1942). However, if processing is interfered with by following the short test vowel presentation with other vowels, the test vowel should not be identified. This result would provide evidence that the redundancy of the vowel in normal speech allows time for processing, since the extended duration of the vowel protects it from later speech until processing has been completed.

Massaro [4] has also employed short vowel stimuli in the recognition masking paradigm described earlier. The spoken vowels were recorded at the same fundamental frequency and amplitude. A steady-state segment of each vowel was stored digitally employing a computer-controlled analog-to-digital converter. During the experiment, the vowel segment was played back using a digital-to-analog converter. In the recognition masking task, 20-msec. segments of the vowels /i/ as in "heat" and /I/ as in "hit" were employed as test items. The masking stimulus was a 270-msec. nonsense vowel made up of two alternating vowel segments. These two segments were taken from the vowels /a/ as in "hat" and /U/ as in "put" and lasted 45 msec. each. The subjects had three days of practice identifying the test vowels in this task before the present experiment.

The results of the experiment are shown in Figure 2. For each subject, identification performance increased with increases in the silent intervowel interval. Therefore, the results indicate that processing

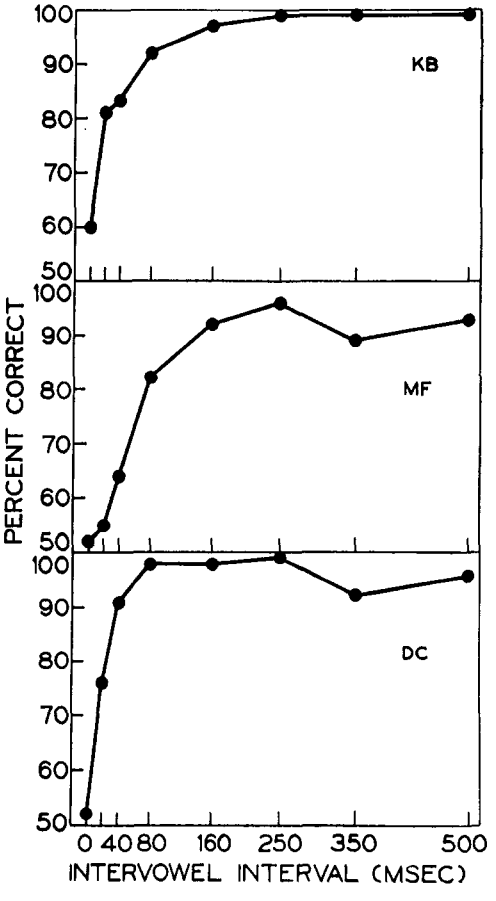

FIG. 2. Percentage of correct identifications of the test vowel for subjects KB, MF, and DC as a function of the duration of the silent intervowel interval.

[4] D. W. Massaro. Perceptual processing time of short vowels. Unpublished manuscript, 1971.

time is also critical for vowel identification. Performance improved much faster with increases in the silent interval in this task than in the tone identification task (cf. Figure 1). The effective masking interval for pure tones was almost twice as long as the masking interval for vowels. This comparison indicates that "i" and "I" are more distinctive than tones of 870 and 770 Hz. Therefore, the vowels could be identified much faster than the tones and were less sensitive to masking at longer silent intervals.

It should be possible to increase the effective masking interval of vowels by increasing the number of test alternatives. With only two vowels subjects can probably make their decision very rapidly, since few features of the vowels need be processed for accurate identification. With a larger number of alternatives, however, subjects would have to process more features before reaching a decision. Accordingly, increasing the number of alternatives should increase the temporal course of backward masking.

The estimates of perceptual processing time from a number of different tasks agree remarkably. The recognition masking studies indicate that perceptual processing can last between 120 and 250 msec. Identifying a sequence of four auditory stimuli requires anywhere from 125 to 300 msec. per stimulus depending on the nature of the material. Although these estimates were obtained in fairly simple experimental situations, they should not be unrelated to the temporal processing of complex auditory information. Therefore, these results will be important in the discussion of processing continuous speech.

The estimates of perceptual processing time are in the same range as the estimates of the duration of preperceptual auditory images. The first 100 to 250 msec. of a stimulus presentation is critical for the recognition process. An auditory stimulus presented for shorter periods produces an auditory image. Accordingly, perceptual processing of a short stimulus continues during the retroactive silent interval. In the recognition masking studies, the short tone or vowel presentation functions as a perceptual unit and recognition corresponds to a readout of information in the unit. If an auditory stimulus changes rapidly over time (e.g., during a consonant-vowel transition), perceptual processing must continue after presentation of the perceptual unit. Therefore, it is necessary to specify what stimulus variables determine the formation of a perceptual unit of information.

Perceptual Units

According to the present theory, a perceptual unit cannot exceed preperceptual auditory memory and must be followed by a steady-state or silent period so that perceptual processing can take place. Perceptual units must also be identified with respect to the temporal changes in sound pressure of the auditory signal. In recognition masking, the test stimulus functioned as the perceptual unit. It is interesting that the masking stimulus could not be integrated with the test stimulus, forming a perceptual unit, and therefore reduce the masking effect. Some integration of the test and masking tones may have occurred when the masking tone followed the test tone immediately. This would account for the initial rise in performance at the zero silent interval in Figure 1.

If the masking tone could be integrated with the test tone, it would not necessarily decrease identification performance. The stimulus situation can be constructed so that integration of two stimuli will be possible as can be seen in Hirsh's (1959) study. In Hirsh's task, the subject is required to judge which of two sounds came first. The onset difference between the two sounds is varied while keeping their offsets simultaneous. Employing two pure tones that differ in frequency, Hirsh showed that the first tone can be identified 75% of the time if it precedes the second by 17 msec. At first glance, this paradigm seems remarkably similar to the auditory recognition masking task discussed above. In this case, Hirsh's results seem discrepant with Massaro's (1970b) results that indicated that it takes about 1/4 sec. to process the pitch of a test tone. However,

it appears that subjects are able to integrate two tones that overlap in time, thus reducing the masking effect of the second tone. Accordingly, the perceptual unit in Hirsh's task was the quality of the pair of tones (Broadbent & Ladefoged, 1959). With high and low tones, the subject learns the two integrated stimuli: high-low and low-high. By simply identifying the integrated stimulus, the subject can indicate which tone came first. On the other hand, if a silent interval occurs between the two tones as in Massaro's (1970b) task, integration cannot take place and the single test tone functions as the perceptual unit. In this case, the second tone terminates perceptual processing of the test tone.

The overall duration of the tones in Hirsh's experiments varied between 440 and 560 msec. There is also evidence that subjects can perceive temporal order at even smaller intervals if the overall duration of the tones is shortened (Patterson & Green, 1970). By shortening the overall duration of Hirsh's low and high tones to 10 msec., Patterson and Green showed that subjects can identify a temporal onset difference of 1.5 msec. This temporal interval is now similar to the empty interval necessary for separation between two brief sounds for subjects to report there are two sounds instead of one (Miller & Taylor, 1948). Hirsh (1959) assumed that the perceptual process involved in identifying whether one or two sounds have occurred is at a lower level than the process involved in the temporal order task. Hirsh felt that since 20 msec. is needed for judging temporal order, more of the perceptual system is involved than the ear itself. However, we can reject Hirsh's distinction between these two tasks since Patterson and Green's results indicate that very short time differences can also be recognized in the temporal order task. Two milliseconds cannot be a measure of central processing time, since this would require that perception be essentially immediate.

Perceptual processing is the identification of acoustic information that is probably stored sequentially. The visual image prolongs the stimulus in its correct spatial pattern (Sperling, 1960). If the auditory image does something analogous to the visual image, it must preserve the stimulus in its original sequential pattern for perceptual processing. The time to process the auditory input in the temporal order task should not be in the range of the onset differences that can be discriminated, but rather should reflect the processing time measured in the pitch identification task (Massaro, 1970b), To measure central processing time, it is necessary to stop perceptual processing at varying times after stimulus presentation. Accordingly, to disrupt performance in the temporal order task, two brief tones differing in onset times could be followed with a third tone after a short silent period. From Massaro's results, the third tone should mask the identification of temporal order. Furthermore, the identification of temporal order might be vulnerable for about 250 msec. after the two tones are presented.

Studdert-Kennedy, Shankweiler, and Schulman (1970) studied the identification of two consonant-vowel syllables as a function of their differences in onset time. The two syllables only differed with respect to the initial consonant, since they were chosen from the set of six syllables /ba,da,ga,pa,ta,ka/. On each trial, the two syllables were presented for 250 msec. to the same ear or to different ears. For both monaural and dichotic presentation, subjects identified a given syllable about 65% of the time if the two syllables were presented simultaneously. Under monaural presentation, increasing the differences in onset time improved identification of the first syllable and disrupted identification of the second syllable. In contrast, increasing the differences in onset time improved identification of both syllables under dichotic presentation.

These results indicate different processes operating in the monaural and dichotic presentations. Monaural presentation of the syllables produces a stimulus situation similar to the paradigm employed by Hirsh (1959). Supporting this, the monaural data agree with Hirsh's results. Subjects were able to identify the first syllable 75%

of the time if it led the second by about 17 msec. If the first syllable led by 50 msec. or more, its identification was perfect. Other results provide evidence for simultaneous masking effects with monaural presentation. Identification of the second syllable decreased with increases in onset differences between the two syllables. Identification of the second syllable is near chance if it follows the first by 50 msec. or more. Studdert-Kennedy et al. (1970) point out that with onset differences of 50 msec. or more, the stimulus sounds like a single syllable with a superimposed click.

Identification of the syllables with dichotic presentation is a different matter. The second syllable is identified about 85% of the time if it follows the first by 50 msec. or more. The fact that the second syllable cannot be identified at these lag times in monaural presentation indicates that dichotic presentation preserves the integrity of the syllable at each ear. Accordingly, since dichotic presentation separates the two syllables into two perceptual units, this task is more similar to Massaro's (1970b) recognition paradigm. Therefore, presentation of the second syllable should interfere with identification of the first. This follows from the assumption that a second input that cannot be integrated with the first terminates perceptual processing of the first. The results support this assumption. Identification of the first syllable was lowest if it led the second by 20 msec. and improved with increases in lead time. Identification of the first syllable was still only 80% correct when it led the second syllable by 120 msec. Further increases in lead time would have increased identification of the first syllable. Extrapolating the results indicates that the optimal processing time for the first syllable would be close to the 250 msec. found by Massaro (1970b) for pure tones. Therefore, the results of Studdert-Kennedy et al. (1970) support the importance of perceptual units and perceptual processing time in auditory perception.

Creel, Boomsliter, and Powers (1970) present clinical evidence that indicates that the perception of tones or noise is temporally categorical. Patients with a deficient blood supply to the brain stem need as much as 200 to 400 msec. of a 1,000-Hz. tone to perceive it as tone rather than noise. A normal listener needs about 10 msec. of the same burst to achieve the tonal sensation (Stevens & Davis, 1938). A given patient hears a sine wave presented for 300 msec. as noise and the same sine wave presented for 400 msec. as tone. As the authors point out, given a tone of 400 msec., one might expect that this patient would hear 300 msec. of noise followed by a short tonal sensation. Since the perception is either noise or tone, the signal is perceived as a Gestalt and is therefore temporally categorical. This result provides convincing evidence that perception is not immediate and that auditory perception is the result of processing preperceptual units of information.

Even though 10 msec. is a sufficient duration for a tone to be heard as tone for normal observers, the results discussed earlier indicate that it takes much longer to process the information necessary for the tonal sensation. Patients with an impaired blood supply to the brain may need more processing time than normals for a tonal sensation. Since the image of a short tone decays rapidly, its information may not last long enough for the patient's perceptual processing necessary for a tonal sensation. In terms of analysis by synthesis (Neisser, 1967), the information may not be present long enough for synthesis of a tonal sensation. Extending the duration of the tone, of course, increases the life of the information necessary for processing and a pure tone is perceived.

Processing Continuous Speech

Studies of interrupted and alternated speech also provide evidence concerning perceptual units. In interrupted speech, half of the speech is eliminated by replacing segments of the speech signal with silence. Therefore, speech and silence are alter-

nated at a given rate. Miller and Licklider (1950) and Dirks and Bower (1970) have shown that intelligiblilty of speech was not lowered significantly by eliminating half of it if the empty intervals occur every 50 or 5 msec. However, if the speech was replaced with silence every other 500 msec., about half of the words were recognized. These results indicate that the durations of the perceptual units critical for the monosyllabic words employed are between 50 and 500 msec. Furthermore, the redundancy of the perceptual unit makes recognition possible even though alternate portions of it are replaced with silence.

Huggins (1964) studied the perception of speech passages taken from scientific essays for the layman. Under conditions of interrupted speech, he found that speech intelligibility was lowest at rates of 1.5 to 5 interruptions per second. Therefore, the poorest speech recognition is produced by replacing the speech signal with silent periods that last between 100 and 330 msec. Since the durations of syllables are within this range, the results implicate the syllable as the perceptual unit for speech. If the syllable functioned as the perceptual unit for speech recognition, its complete removal should and does lead to the poorest speech recognition.

Alternating continuous speech from ear to ear rather than removing segments of the speech signal also provides information about the perceptual unit of speech perception. Alternating speech from ear to ear about three times per second disrupts speech perception (Cherry, 1953; Cherry & Taylor, 1954). In contrast, alternation from ear to ear at much faster or slower rates does not disrupt speech perception. Huggins (1964) elucidated these findings by showing that the rate of alternated speech that led to the poorest speech recognition was dependent on the speed of the speech signal. If speech is speeded up, faster alternation rates are necessary for minimal intelligibility.

By covarying the rate of alternation and the speed of the speech signal, Huggins found that alternating segments of speech approximating the average duration of a syllable led to the poorest recognition. Alternating speech at this rate insures that a switch between ears will occur exactly once during most syllables. Assuming that a speech signal in one ear cannot be integrated with the signal arriving in the opposite ear, these results indicate the syllable as an important segment in speech perception. With only part of the syllable in each ear, subjects would have minimal information if identification of the syllable was critical for speech perception. If the alternation of the speech was slowed down, the subject would have more of the syllable in one ear and, therefore, should be able to identify it more accurately. Since processing time is constant across rates of alternation in Huggin's study, the results indicate the syllable as the optimal segment or perceptual unit in the processing of speech.

Huggins (1964) also presents some measurements of the temporal characteristics of his speech passages. First, it is interesting that 18% of the total duration of the passages was essentially silent. This free time probably helps establish perceptual units and gives the subject time for perceptual processing. The mean duration of the syllables was 200 msec. with an interquartile range of 150–250 msec. The durations of syllables, then, agree with the estimates of perceptual processing time. The agreement of perceptual processing time and the duration of the perceptual unit correspond to a similar finding in visual information processing. Since each eye movement in reading overwrites the preceding visual image, the fixation must and does last long enough for perceptual processing (Haber, 1970; Woodworth, 1938).

The results indicate the syllable as the perceptual unit in speech perception. Warren (1970), in fact, has shown that if another extraneous sound replaces a phoneme in a sentence, listeners actually report hearing the missing phoneme. Warren deleted a 120-msec. section corresponding to the first /s/ in legislatures in the sentence, "The state governors met with their respective legislatures convening in the capital city." The missing section

was replaced with a recorded cough or tone of the same duration. Only one of the 40 subjects reported hearing a missing sound and he selected the wrong one. In addition, subjects were not able to locate the position of the extraneous sound in the sentence. These results indicate that we can synthesize a unit of speech perception with a limited amount of information in the preperceptual auditory image.

The perceptual synthesis of the missing phoneme must be made on the cues from its auditory context. Relevant features exist before and after the missing phoneme and these features are held in preperceptual storage for perceptual processing. On the basis of limited information (since a phoneme is missing), the subject extracts what features are present and is able to synthesize the correct syllable or word. This synthesis must take place on a perceptual level rather than a verbal or abstract level since the subject actually believes he heard the missing phoneme. If the phoneme is replaced with silence rather than an extraneous noise, the subject notices the gap and yet is able to identify the word correctly. The temporal course of the information in the preperceptual store could be determined by varying the temporal separation between the degraded word and the necessary context for word synthesis. Phonemic restoration should not occur when the sensory features near the missing phoneme are no longer available for perceptual synthesis. Accordingly, if the context comes after this preperceptual information has decayed or has been masked by other sensory input, the observer should notice the missing phoneme.

In a different approach to disrupting speech, Cherry and Wiley (1967) and Holloway (1970) have shown that passing only the strongly voiced speech sounds decreases speech perception. This sequence of sounds and silent periods does not contain the rhythmic nature of normal speech. By adding a low level of white noise in the silent gaps between the separate sounds, the authors showed that the speech once again becomes intelligible. These results show that the nature of normal speech is important for establishing temporal or perceptual units for speech perception. By artificially segregating the information, the organization of the larger units normally identified is broken. Noise, then, can bridge the gap and provide the subject with an organized unit for speech perception.

Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) have presented a detailed and interesting account of speech perception that differs from the present approach. The basic assumption of Liberman et al. is that speech perception occurs at the level of the phoneme. Therefore, their emphasis is on the information processing necessary for phonemic perception from the acoustic signal. The analysis of the acoustic signal is in terms of the formant structure of speech patterns. Figure 3 presents the first two formants of the patterns /di/ and /du/. The darkened areas represent the concentration of acoustic energy at the indicated frequencies over time.

Using the example of /di/ and /du/, Liberman et al. show that the second formant structure is necessary for perception of the /d/ segment of the syllables /di/ and /du/. Therefore, they argue that /d/ perceived as the same phoneme is represented by two entirely different acoustic signals (cf. Figure 3). Liberman et al. (1967) state,

> It is, then, the second-formant transitions that are, in the patterns in Figure 1 [Figure 3 in the present article], the acoustic cues for the perception of the /d/ segment of the syllables /di/ and /du/. We would first note that /d/ is the same perceptually in the two cases, and then see how different are the acoustic cues. In the case of /di/ the transition rises from approximately 2200 cps to 2600 cps; in /du/ it falls from about 1200 to 700 cps. In other words, what is perceived as the same phoneme is cued, in different contexts, by features that are vastly different in acoustic terms. How different these acoustic features are in nonspeech perception can be determined by removing them from the patterns of Figure 1 [Figure 3 in the present article] and sounding them in isolation. When we do that, the transition isolated from the /di/ pattern sounds like a rapidly rising whistle or glissando on high pitches, the one from /du/ like a rapidly falling whistle on low pitches [p. 435].

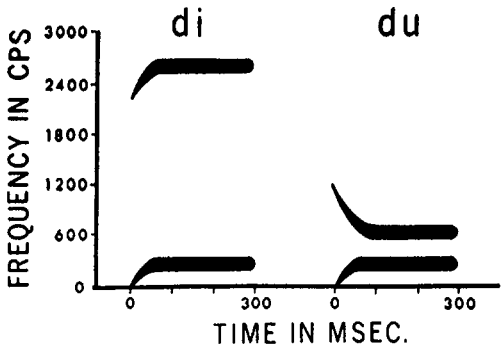

FIG. 3. Spectrographic patterns sufficient for the synthesis of /d/ before /i/ and /u/. (Reprinted with permission from an article by A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy published in the *Psychological Review*, 1967, Vol. 74. Copyrighted by the American Psychological Association, Inc., 1967.)

However, assuming that the syllables /di/ and /du/ function as perceptual units, there exists a direct correspondence between acoustic signal and percept.

Other results of Liberman et al. (1967) indicate that presenting the early part of the pattern /di/ alone produces a nonspeech sound. By adding successive parts of the pattern, the phoneme /d/ is never heard. Rather the perception changes categorically to the syllable /di/. Therefore, with increases in stimulus duration, perception changes from a nonspeech sound to a consonant-vowel cluster. Liberman et al. conclude from this evidence that the formant transition continuously provides information about the two phonemes in parallel.

In contrast to the conclusion of Liberman et al., their results provide convincing support for eliminating the phoneme as the perceptual unit for processing speech. Evidence reviewed above indicated the syllable as the perceptual unit critical for speech perception. Other recent results show that phoneme perception follows the perception of the syllables containing the phonemes. Savin and Bever (1970) presented subjects with a sequence of nonsense syllables. The subjects responded as soon as they heard a target in the sequence. They responded faster to a target that was a complete syllable than to a target that was a phoneme from that syllable. Warren (1971) has studied identification time for targets in continuous speech. Identification times for monosyllabic words or nonsense syllables were faster than identification times for phoneme clusters. Reaction times to phoneme clusters were shorter than reaction times to individual phonemes within the clusters. These results indicate that syllable identification precedes phoneme identification in speech perception.

These results suggest a different process for syllable identification than the process responsible for phoneme identification. If the syllable is the perceptual unit, identification of the syllable involves an analysis of the information in the preperceptual storage of the syllabic unit. The distinctive features necessary for recognition would be relevant to this unit. On the other hand, since the phoneme is not a perceptual unit, phoneme identification in speech probably does not occur as a result of analysis of preperceptual memory. After identification of the syllable, the subjects could simply scan the syllable for the appropriate phoneme target. In vision, this process would be analogous to finding the letter "a" in the word "cat" after the word was already identified. Warren's (1970) results indicated that subjects do not recognize that a phoneme is missing when it is replaced by an extraneous sound. These results support the interpretation given here that the phoneme is not perceived directly from preperceptual memory, but is inferred from the identification of the syllable or word.

The results indicate perceptual units that are larger than the phoneme in speech perception. As Liberman et al. (1967) point out, attempts to produce intelligible speech by a recombination of phonemes has been unsuccessful (Harris, 1953). In contrast, Petersen, Wang, and Sivertsen (1958) showed that intelligible speech can be produced by using units that are at least one-half syllable in length. Liberman et al. (1967) concluded that

> This parallel delivery of information produces at the acoustic level the merging of influences we have already referred to and yields irreducible acoustic

segments of approximately syllabic dimensions [p. 441].

Rather than parallel processing of phonemes, it is more parsimonious to assume that the syllable functions as the perceptual unit for speech perception.[5]

If the vowel is integrated with the consonant-vowel transition (as in Hirsh's task), the extended duration of the vowel in normal speech would provide time for perceptual processing. As mentioned earlier, vowels in normal speech last 150–250 msec. As can be seen in Figure 3, the last 200 msec. of /di/ and /du/ do not provide new information. The first 100 msec. of the pattern probably could be identified if it is presented alone and followed by a silent interval. However, if the consonant-vowel transition was followed by a speech sound that could not be integrated with it, perception should be disrupted and backward recognition masking should occur.

Throughout this paper it has been argued that there are perceptual or temporal units of auditory information, and that perception is determined by the processing of the information in the temporal units. The assumption of temporal units in auditory information processing may seem similar to the psychological moment theory proposed by Stroud (1955). However, the important difference is that the psychological moment assumes that time is the critical independent variable determining the perceptual experience (Kahneman, 1968; Schmidt & Kristofferson, 1963; Stroud, 1955). Contrasting this, the results indicate that the size of the temporal unit is determined by the properties of the physical stimulus. The duration of the perceptual unit in a consonant-vowel transition is longer than the 20-msec vowels used in the recognition paradigm. However, in both cases, we must process the information available and this processing requires time. During perceptual processing, the information is extracted from the stimulus image and the synthesized percept enters consciousness or short-term memory as a unit. The preperceptual features of the auditory input will usually be lost with the next auditory input. Therefore, short-term memory is necessary for retaining the synthesized information necessary for combination with the synthesized information from other perceptual units (Lashley, 1951; Wickelgren, 1969).

[5] It should be noted that the term "syllable" must be interpreted loosely. There are four basic syllable types: Vs, VCs, CVs, and CVCs where V is a vowel and C is a single consonant or a consonant cluster. It is unlikely that all syllables function as perceptual units. For example, CVCs probably contain two perceptual units.

Auditory Short-Term Memory

Morton (1970) and Crowder and Morton (1969) interpret some short-term effects as evidence for a precategorical acoustic storage that is assumed to have many of the properties of an auditory preperceptual store. Morton and Crowder's demonstration of precategorical acoustic storage is the suffix effect. The subjects are asked to recall serially a list of eight items presented auditorily at a rate of two per second. Performance on a control list is compared to performance on the same list followed by the suffix zero. Although the suffix is not recalled, it selectively decreases performance on the final two or three items (Crowder, 1967). Control studies show that the auditory suffix zero produces more interference with recall of the final items than a visual suffix (Morton & Holloway, 1970).

These experiments indicate that although the suffix interferes with auditory information from the stimulus list, this information is *not* necessarily precategorical. It was argued earlier that the information available in the auditory image of a short tone was preperceptual since the information in the image was present before the tone had been identified. On the other hand, in the serial recall task, each item in the list presentation should have been identified before the following item is presented. Therefore, the suffix effect is due to interference of the subject's short-term memory for the final items in the list. Recognition masking, on the other hand, is due to interference with the subject's perception

of the test item. If subjects were required to recall only the last two or three items in the list, the suffix effect should be eliminated. Since each item is perceived as it is presented, no interference should be observed in the recall task without some short-term memory loss.

Crowder and Morton's results are more consistent with interference effects in short-term memory than precategorical acoustic storage. A number of results indicate that decreasing the similarity of the suffix to the stimulus list decreases interference. First, presenting the list in a male voice and the suffix in a female voice reduces the suffix effect. Second, presenting the suffix contralateral to the stimulus list presentation also reduces the suffix effect. Finally, white noise or a buzzer presented as the suffix produces no interference. Crowder and Morton's results, then, demonstrate that similarity plays an important role in short-term memory. This interpretation agrees with recent studies of short-term memory that find that increasing the similarity of retroactive items to test items increases forgetting (Deutsch, 1970; Massaro, 1970c, Experiment II; Reitman, 1971; Wickelgren, 1965, 1966).

The size of the suffix effect is directly related to the similarity of the stimulus list and the suffix. Contrasting this, similarity effects have not been found in studies of interference with preperceptual auditory images. Massaro (1970b) showed that decreasing the similarity between the test and masking tones did not decrease backward recognition masking. Furthermore, presenting the masking tone contralateral to the test tone presentation did not reduce the interference effect of the masking tone. Finally, white noise interferes with preperceptual memory in speech recognition (Dirks & Bower, 1970). These results support the assumption that the suffix effect and backward recognition masking occur at two different stages of information processing. Crowder and Morton are studying forgetting of identified items and have not demonstrated conclusively that memory for these items contains "precategorical acoustic storage."

As Morton (1970) points out, many of the results demonstrating a suffix effect can be interpreted in terms of selective attention. For example, presenting the suffix at twice the subjective loudness of the stimulus list reduces the suffix effect. In terms of selective attention, the difference in loudness of the suffix provides a dimension for rejection of the suffix. On the other hand, if precategorical acoustic storage was responsible for the suffix effect, a louder suffix would be expected to produce more interference, since the suffix supposedly overwrites the information in precategorical storage. Morton (1970) presents the crucial test between selective attention and precategorical acoustic storage accounts of the suffix effect. Presenting the stimulus list monaurally, he compared a monaural to a binaural presentation of the suffix. Selective attention would predict a reduction in the suffix effect with the binaural presentation, since this could be used as a dimension for rejection. However, "precategorical acoustic storage" predicts no difference, since a binaural suffix should overwrite preperceptual information as well as a monaural suffix. The results provide evidence against a "precategorical acoustic storage" account of the suffix effect, since the binaural suffix actually produced less interference than the monaural suffix.

Bryden (1971), working in a dichotic listening situation, has also interpreted attention and short-term memory effects as preperceptual auditory memory effects. Bryden presents four pairs of digits, at two pairs per second, in a dichotic listening experiment, The subject is instructed to attend to a given channel. He then recalls all of the digits with instructions to recall the attended or unattended message first. Results indicate that, overall, the attended digits are recalled much better than the unattended digits. Furthermore, an analysis of the serial position curve indicates that there is a very small serial position effect for the attended digits. In contrast, memory for the unattended digits shows a huge recency effect. Recall of the last

unattended digit is twice as probable as recall for the first unattended digit.

Bryden (1971) interprets these results as indicating that the unattended items are in preperceptual storage. However, the digits presented in the unattended ear could have been identified at presentation. Supporting this interpretation, Norman (1969) has shown that items presented to the unattended ear in a shadowing task are recognized at presentation. These unattended items also show forgetting functions similar to typical short-term memory studies (Norman, 1966). Therefore, it is argued that the unattended items in Bryden's task were identified at presentation but received very little perceptual processing for memory and, therefore, showed more forgetting than the attended digits.

This interpretation requires a distinction between two stages of perceptual processing pointed out explicitly by Massaro (1970a). The first stage, perception or perceptual processing of the information in the preperceptual auditory image refers to the analysis of the sensory input in which physical features are examined in order to identify the input. After identification of the item, the second stage of perceptual processing is carried out in order to remember the item. In Bryden's task the unattended items could have been poorly recalled because they received very little perceptual processing for memory. Digits can be identified easily within 200 msec. as shown by Warren and Warren (1970). Therefore, Bryden's subjects had time to identify both the attended and unattended items at presentation. In the time remaining, they rehearsed the attended items. This interpretation is more compatible with the literature reviewed here than is an interpretation based on preperceptual storage.

Preperceptual auditory images and short-term auditory memory have their analogous counterparts in vision. A preperceptual auditory image is analogous to the positive afterimage in vision that keeps the stimulus in its original form (Sperling, 1960). Evidence presented earlier indicated that very little preperceptual information remains after readout of the auditory image which occurs within 250 msec. Accordingly, synthesized short-term auditory memory must be qualitatively different than preperceptual auditory images. Synthesized auditory memory is analogous to the visual memory studied by Posner and his colleagues (Beller, 1971; Posner, Boies, Eichelman, & Taylor, 1969; Posner & Keele, 1967). Their results indicate that the visual memory used to remember a capital A is qualitatively different than preperceptual visual images. Similarly, the auditory memory used to remember a list of digits is qualitatively different than preperceptual auditory images.

Conclusion

Temporal or perceptual units play an important role in auditory recognition. These units are held in a preperceptual auditory store for perceptual processing. If a second auditory pattern can be integrated with a first, they can form a single unit. If not, the second can interfere with the preperceptual image of the first. Perceptual processing refers to an analysis of information in the perceptual unit. This analysis requires an examination of the physical features of the stored sequential pattern in order to identify the input. The temporal course of perceptual processing depends on the complexity of the identification task. The more difficult the discrimination, the longer the time needed for reading out the necessary information.

The present results are also relevant to the distinction drawn between the processing of speech and nonspeech material (Liberman et al., 1967). The motor theory of speech perception assumes that articulatory mechanisms mediate the perceptual processing of speech. However, assuming a preperceptual auditory store eliminates some of the "complexities" of speech input that demand a unique processor for speech. Most importantly, the perceptual unit of speech cannot be at the phoneme level but must be more complicated. On the other hand, it cannot be as large as two or three

words as noted in the discussion of short-term memory experiments. Phonemes may become important when the features necessary for recognition are determined. However, this is an entirely different problem than determining the perceptual unit that remains intact for perceptual processing.

The present paper has not attempted to provide a total account of speech perception or complex pattern recognition where short-term memory, long-term memory, and decision processes become critical. With respect to speech perception, the contributions of syntax and semantics cannot be ignored (Miller, 1962; Miller, Heise, & Lichten, 1951). With complex auditory patterns, form, rhythm, melody, and organization become important at the level of synthetic memory (Cheatham & White, 1954; Dowling, 1971; Garner, 1951; Heise & Miller, 1951; White, 1963). Enjoying a symphony involves much more than simply analyzing the information available in preperceptual store. The present paper has, however, provided a conceptual framework for the recognition stage of perceptual processing. This stage of perceptual processing outputs a syntheiszed percept that becomes available to the next stage of auditory information processing.

REFERENCES

AVERBACH, E., & CORIELL, A. S. Short-term memory in vision. *Bell Systems Technical Journal*, 1961, **40**, 309–328.

BELLER, H. K. Priming: Effects of advance information on matching. *Journal of Experimental Psychology*, 1971, **87**, 176–182.

BROADBENT, D. E., & LADEFOGED, P. Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 1959, **31**, 1539.

BYRDEN, M. P. Attentional strategies and short-term memory in dichotic listening. *Cognitive Psychology*, 1971, **2**, 99–116.

CHEATHAM, P. G., & WHITE, C. T. Temporal numerosity: III. Auditory perception of number. *Journal of Experimental Psychology*, 1954, **47**, 425–432.

CHERRY, E. C. Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 1953, **25**, 975–979.

CHERRY, E. C., & TAYLOR, W. K. Some further experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 1954, **26**, 554–559.

CHERRY, C., & WILEY, R. Speech communication in very noisy environments. *Nature*, 1967, **214**, 1164.

CREEL, W., BOOMSLITER, P. C., & POWERS, S. R. Sensations of tones as perceptual forms. *Psychological Review*, 1970, **77**, 534–545.

CROWDER, R. G. Prefix effects in immediate memory. *Canadian Journal of Psychology*, 1967, **21**, 450–461.

CROWDER, R. G., & MORTON, J. Precategorical acoustic storage. *Perception and Psychophysics*, 1969, **5**, 365–373.

DEATHERAGE, B. H., & EVANS, T. R. Binaural masking: Backward, forward, and simultaneous effects. *Journal of the Acoustical Society of America*, 1969, **46**, 362–371.

DEUTSCH, D. Tones and numbers: Specificity of interference in immediate memory. *Science*, 1970, **168**, 1604–1605.

DIRKS, D. D., & BOWER, D. Effect of forward and backward masking on speech intelligibility. *Journal of the Acoustical Society of America*, 1970, **47**, 1003–1008.

DONDERS, F. C. On the speed of mental processes. (Trans. by W. G. Koster) In W. G. Koster (Ed.), *Attention and performance II*. Amsterdam: North Holland Publishing, 1969.

DOWLING, W. J. Recognition of inversions of melodies and melodic contours. *Perception and Psychophysics*, 1971, **9**, 348–349.

EFRON, R. Effect of stimulus duration on perceptual onset and offset latencies. *Perception and Psychophysics*, 1970, **8**, 231–234. (a)

EFRON, R. The minimum duration of a perception. *Neuropsychologia*, 1970, **8**, 57–63. (b)

EFRON, R. The relationship between the duration of a stimulus and the duration of a perception. *Neuropsychologia*, 1970, **8**, 37–55. (c)

ELLIOT, L. L. Backward and forward masking of probe tones of different frequencies. *Journal of the Acoustical Society of America*, 1962, **34**, 1116–1117. (a)

ELLIOT, L. L. Backward masking: Monotic and dichotic conditions. *Journal of the Acoustical Society of America*, 1962, **34**, 1108–1115. (b)

ELLIOT, L. L. Development of auditory narrow-band frequency contours. *Journal of the Acoustical Society of America*, 1967, **42**, 143–153.

ERIKSON, C. W., & JOHNSON, H. J. Storage and decay characteristics of nonattended stimuli. *Journal of Experimental Psychology*, 1964, **68**, 28–36.

FLETCHER, H. *Speech and hearing in communication*. Princeton: Van Nostrand, 1953.

GALANTER, E. Contemporary psychophysics. In, *New directions in psychology*. New York: Holt, Rinehart & Winston, 1962.

GARNER, W. R. The accuracy of counting repeated short tones. *Journal of Experimental Psychology*, 1951, **41**, 310–316.

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

717–724. (English translation: *Biophysics*, 1961, **6**, 809–817.)

GRAY, G. W. Phonemic microtomy: The minimum duration of perceptible speech sounds. *Speech Monographs*, 1942, **9**, 75–90.

GREEN, D. M., & SWETS, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.

GUTTMAN, N., & JULESZ, B. Lower limits of auditory periodicity analysis. *Journal of the Acoustical Society of America*, 1963, **35**, 610.

HABER, R. N. (Ed.) *Information-processing approaches to visual perception*. New York: Holt, Rinehart & Winston, 1969.

HABER, R. N. How we remember what we see. *Scientific American*, 1970, **222**(2), 104–112.

HABER, R. N., & STANDING, L. G. Direct measures of short-term visual storage. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 43–54.

HARRIS, C. M. Study of the building blocks in speech. *Journal of the Acoustical Society of America*, 1953, **25**, 962–969.

HEISE, G. A., & Miller, G. A. An experimental study of auditory patterns. *American Journal of Psychology*, 1951, **64**, 68–77.

HIRSH, I. J. Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 1959, **31**, 759–767.

HOCHBERG, J. Attention, organization and consciousness. In D. I. Mostofsky (Ed.), *Attention: Contemporary theory and analysis*. New York: Appleton-Century-Crofts, 1970.

HOLLOWAY, C. M. Passing the strongly voiced components of noisy speech. *Nature*, 1970, **226**, 178–179.

HOMICK, J. L., ELFNER, L. F., & BOOTHE, G. G. Auditory temporal masking and the perception of order. *Journal of the Acoustical Society of America*, 1969, **45**, 712–718.

HOUSE, A. S. On vowel duration in English. *Journal of the Acoustical Society of America*, 1961, **33**, 1174–1178.

HUGGINS, A. W. F. Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America*, 1964, **36**, 1055–1064.

JAKOBSON, R., & HALLE, M. *Fundamentals of language*. 's-Gravenhage: Mouton, 1956.

KAHNEMAN, D. Method, findings, and theory in the studies of visual masking. *Psychological Bulletin*, 1968, **70**, 404–425.

KOFFKA, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.

LASHLEY, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley, 1951.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431–461.

MASSARO, D. W. Perceptual processes and forgetting in memory tasks. *Psychological Review*, 1970, **77**, 557–567. (a)

MASSARO, D. W. Preperceptual auditory images. *Journal of Experimental Psychology*, 1970, **85**, 411–417. (b)

MASSARO, D. W. Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, 1970, **83**, 32–39. (c)

MASSARO, D. W. Effect of masking tone duration on preperceptual auditory images. *Journal of Experimental Psychology*, 1971, **87**, 146–148.

MASSARO, D. W. Stimulus information versus processing time in auditory pattern recognition. *Perception and Psychophysics*, 1972, in press.

MCGILL, W. J. Loudness and reaction time. *Acta Psychologica*, 1961, **19**, 193–199.

MCGILL, W. J. Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963.

MILLER, G. A. Decision units in the perception of speech. *IRE Transactions in Information Theory*, 1962, IT-8, 81–83.

MILLER, G. A., HEISE, G. A., & LICHTEN, W. The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, 1951, **41**, 329–335.

MILLER, G. A., & LICKLIDER, J. R. Intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 1950, **22**, 167–173.

MILLER, G. A., & TAYLOR, J. Perception of repeated bursts of noise. *Journal of the Acoustical Society of America*, 1948, **20**, 171–182.

MORTON, J. A functional model for memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

MORTON, J., & HOLLOWAY, C. M. Absence of a cross-modal "suffix effect" in short-term memory. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 167–176.

NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

NORMAN, D. A. Acquisition and retention in short-term memory. *Journal of Experimental Psychology*, 1966, **72**, 369–381.

NORMAN, D. A. Memory while shadowing. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 85–93.

PATTERSON, J. H., & GREEN, D. M. Discrimination of transient signals having identical energy spectra. *Journal of the Acoustical Society of America*, 1970, **48**, 894–905.

PETERSON, G. E., WANG, W. S.-Y., & SIVERTSEN, E. Segmentation techniques in speech synthesis. *Journal of the Acoustical Society of America*, 1958, **30**, 739–742.

PLOMP, R. Decay of auditory sensation. *Journal of the Acoustical Society of America*, 1964, **36**, 277–282.

POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. I. Retention of visual and name codes of single letters. *Journal of Experimental Psychology*, 1969, **79**(1, Pt. 2).

POSNER, M. I., & KEELE, S. W. Decay of visual information from a single letter. *Science*, 1967, **158**, 137–139.

RAAB, D. H. Backward masking. *Psychological Bulletin*, 1963, **60**, 118–129.

REITMAN, J. S. Mechanisms of forgetting in short-term memory. *Cognitive Psychology*, 1971, **2**, 185–195.

SAMOILOVA, I. K. Masking of short tone signals as a function of the time interval between masked and masking sounds. *Biophysics*, 1959, **4**(5), 44–52.

SAVIN, H. B., & BEVER, G. T. The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 295–302,

SCHMIDT, M. W., & KRISTOFFERSON, A. B. Discrimination of successiveness: A test of a model of attention. *Science*, 1963, **139**, 112–113.

SPERLING, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, **74**(11, Whole No. 498).

SPERLING, G. Successive approximations to a model for short-term memory. *Acta Psychologica*, 1967, **27**, 285–292.

STERNBERG, S. The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, 1969, **30**, 276–315.

STEVENS, S. S., & DAVIS, H. *Hearing*. New York: Wiley, 1938.

STROUD, J. The fine structure of psychological time. In H. Quastler (Ed.), *Information theory in psychology*. New York: Free Press, 1955.

STUDDERT-KENNEDY, M., SHANKWEILER, D., & SCHULMAN, S. Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. *Journal of the Acoustical Society of America*, 1970, **48**, 599–602.

THOMAS, I. B., HILL, P. B., CARROLL, F. S., & BIENVENIDO, G. Temporal order in the perception of vowels. *Journal of the Acoustical Society of America*, 1970, **48**, 1010–1013.

VON BÉKÉSY, G. Auditory inhibiton in concert halls. *Science*, 1971, **171**, 529–536.

WARREN, R. M. Perceptual restoration of missing speech sounds. *Science*, 1970, **167**, 392–393.

WARREN, R. M. Identification times for phonemic components of graded complexity and for spelling of speech. *Perception and Psychophysics*, 1971, **9**, 345–349.

WARREN, R. M., OBUSEK, C. J., FARMER, R. M., & WARREN, R. P. Auditory sequence: Confusion of patterns other than speech or music. *Science*, 1969, **164**, 586–587.

WARREN, R. M., & WARREN, R. P. Auditory illusions and confusions. *Scientific American*, 1970, **223**(6) 30–36.

WHITE, C. T. Temporal numerosity and the psychological unit of duration. *Psychological Monographs*, 1963, **77**(12, Whole No. 575).

WICKELGREN, W. A. Acoustic similarity and retroactive interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 53–61.

WICKELGREN, W. A. Phonemic similarity and interference in short-term memory for single letters. *Journal of Experimental Psychology*, 1966, **71**, 396–404.

WICKELGREN, W. A. Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review*, 1969, **76**, 1–15.

WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

(Received July 12, 1971)