6

Auditory Information Processing

Dominic W. Massaro

University of Wisconsin

I. INTRODUCTION

A model of information processing and memory must account for the memory structures and psychological processes intervening between stimulus and response. Memory structures are storage components that hold information. Psychological processes are operations in the model that involve a transformation of information in one structure to information in another. The goal of a model is to define the nature of the information held in each structure, its life span, and the rules by which it is lost. Similarly, the operations of the intervening processes must be described in detail. For example, can a given process allocate a limited processing capacity to one task at the expense of another? Structure and process are highly interdependent and the theorist must specify the structures utilized at each processing stage.

The goal of this chapter is to selectively review auditory information processing and short-term memory research in order to make sense of the structures and processes involved. I utilize a general processing model to incorporate data and to contrast different theories. Recent experiments have revealed a number of facts about perception and retention processes in short-term memory tasks. One goal of this paper is to incorporate these facts into a general information-processing model and to show that the rules describing these processes can be illuminated in a successive stage analysis. The model has been developed over the past five years hand in hand with experimental work (Massaro, 1970b, 1972a, 1975a, b). Figure

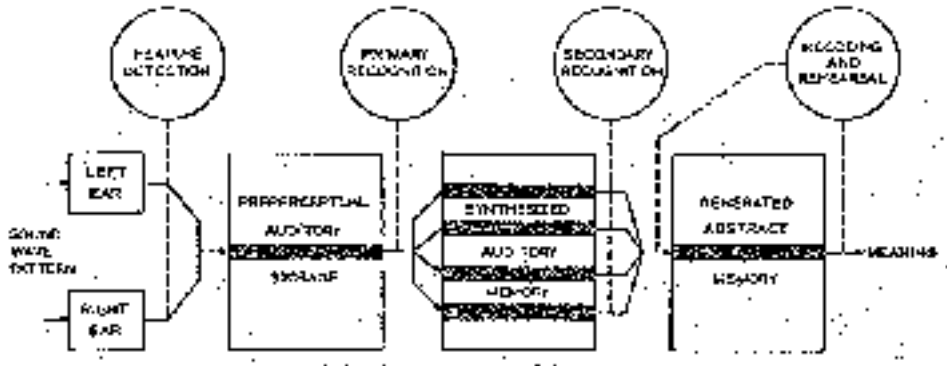

FIGURE 1 Components of auditory information processing model. See text for explanation.

1 presents in schematic form the most recent elaboration of the auditory information-processing components of the model.

Auditory input is funneled through the ears so that although the atmosphere is filled with sound, our contact with it is limited to these two organs. This observation makes apparent the differences between auditory and visual receptor systems. Whereas the relative spatial location of objects is represented in the retinal projection of the scene, the location of a sound is not correlated in any way with the position stimulated on the basilar membrane. Therefore, the relative position of a noisy and visible object can be located with one eye but not with one ear. Two auditory inputs to the two ears are needed to localize sound, and subtle differences in the intensities and temporal arrivals of the sound to the two ears are cues to the localization of a sound (Mills, 1972).

The transduction of the sound pressure fluctuations passes on featural information which is detected and stored centrally in preperceptual auditory storage. Feature detection can occur independently at each of the two ears and also after the separate inputs are combined in the auditory system. Although the sound inputs to the two ears are usually fused so that only one sound is perceived, different sounds to the two ears can be heard separately given the proper stimulus conditions. Recently, studies have shown that features of different speech sounds to the two ears can be integrated or combined centrally at the level of preperceptual auditory storage (Pisoni, 1975; Pisoni & McNabb, 1974). I will return to the feature detection problem in Section II.

Preperceptual auditory storage holds the featural information given by the detection process for a short time (about 250 msec) after a stimulus is presented. The primary recognition process involves a resolution of this information, producing a synthesized percept held in synthesized auditory

memory. Primary recognition accomplishes the phenomenological outcome of perceiving a particular sound of a particular loudness and quality at some location in space. The properties of preperceptual storage and the dynamics of the primary recognition process are discussed in Section III.

Section IV provides a discussion of the properties of preperceptual and synthesized auditory codes. Preperceptual auditory storage can be considered to be a buffer with only one central storage channel. It has a limited capacity of one item; a second sound overwrites the information stored there about the previous sound. The output of primary recognition produces a percept held in synthesized auditory memory. Synthesized auditory memory has a larger capacity (of 1 or 2 sec) and information can be held here along separate channels such as spatial location, sound quality, or intonation.

Given a synthesized auditory code, the task of the secondary recognition process is to transform this perceptual experience into meaning. Secondary recognition is limited in the number of auditory channels it can process per unit of time. Listeners cannot derive meaning from information that is perceived at two auditory channels as well as when the same information is perceived along one auditory channel. In Section V, I discuss traditional shadowing experiments and a relatively new task in which subjects are asked to count the number of sounds in a short sequence. In both tasks, the variable of interest is whether the auditory information is presented along a single auditory channel or alternated between channels.

Perceptual and conceptual codes are utilized at two successive stages in our model. Several experimental tasks have tapped the utilization of these codes. One phenomenon, categorical perception, has been important for arguments supporting the special and unique processes in speech perception. In Section VI, I argue that categorical perception might be explained in terms of the availability of perceptual and conceptual codes in the task. Similarly, the perceptual codes of consonant and vowel sounds might differ in resolution, producing differences in processing these sounds. Auditory recency and suffix effects are also discussed in terms of the utilization of perceptual and conceptual codes.

Secondary recognition involves the transformation of percept into meaning in generated abstract memory. The concept of different channels does not seem to be necessary in generated abstract memory since we appear to follow only one train of thought at a time. A simple limited-capacity rule describes processing at the level of this abstract memory. Recoding and rehearsal operations allow the maintenance of information in generated abstract memory indefinitely. The recoding operation can also transform a conceptual code into a perceptual one. Forgetting in generated abstract memory is discussed in Section VII. I also discuss some

memory. Primary recognition accomplishes the phenomenological outcome of perceiving a particular sound of a particular loudness and quality at some location in space. The properties of preperceptual storage and the dynamics of the primary recognition process are discussed in Section III.

Section IV provides a discussion of the properties of preperceptual and synthesized auditory codes. Preperceptual auditory storage can be considered to be a buffer with only one central storage channel. It has a limited capacity of one item; a second sound overwrites the information stored there about the previous sound. The output of primary recognition produces a percept held in synthesized auditory memory. Synthesized auditory memory has a larger capacity (of 1 or 2 sec) and information can be held here along separate channels such as spatial location, sound quality, or intonation.

Given a synthesized auditory code, the task of the secondary recognition process is to transform this perceptual experience into meaning. Secondary recognition is limited in the number of auditory channels it can process per unit of time. Listeners cannot derive meaning from information that is perceived at two auditory channels as well as when the same information is perceived along one auditory channel. In Section V, I discuss traditional shadowing experiments and a relatively new task in which subjects are asked to count the number of sounds in a short sequence. In both tasks, the variable of interest is whether the auditory information is presented along a single auditory channel or alternated between channels.

Perceptual and conceptual codes are utilized at two successive stages in our model. Several experimental tasks have tapped the utilization of these codes. One phenomenon, categorical perception, has been important for arguments supporting the special and unique processes in speech perception. In Section VI, I argue that categorical perception might be explained in terms of the availability of perceptual and conceptual codes in the task. Similarly, the perceptual codes of consonant and vowel sounds might differ in resolution, producing differences in processing these sounds. Auditory recency and suffix effects are also discussed in terms of the utilization of perceptual and conceptual codes.

Secondary recognition involves the transformation of percept into meaning in generated abstract memory. The concept of different channels does not seem to be necessary in generated abstract memory since we appear to follow only one train of thought at a time. A simple limited-capacity rule describes processing at the level of this abstract memory. Recoding and rehearsal operations allow the maintenance of information in generated abstract memory indefinitely. The recoding operation can also transform a conceptual code into a perceptual one. Forgetting in generated abstract memory is discussed in Section VII. I also discuss some

methodological problems in free recall experiments. The similarities and differences of the present model to other approaches are briefly considered in Section VIII.

II. FEATURE DETECTION

We assume that features are detected in the sound signal and placed in preperceptual auditory storage. The primary recognition process involves a resolution of these features, producing a synthesized percept. Pisoni and McNabb (1974) provide a nice demonstration of how these processes operate in the recognition of synthetic speech sounds. Subjects were required to identify a test syllable presented to one ear while simultaneously another interference syllable was presented to the other ear. The test syllable was a 300-msec stop consonant vowel (CV) chosen from the set /ba, da, pa, ta/ whereas the interference syllable was chosen from the set /ga, ka, ga, ka, ga, ka/. Figure 2 shows schematized sound spectrograms of the six stop consonants paired with the vowel /a/. The sounds are characterized by rapid formant transitions at the onset of the stop consonant followed by the steady-state vowel, and can be described along two acoustic dimensions. First, the sounds /ba, da, ga/ are voiced but /pa, ta, ka/ are voiceless. Voicing is determined by the delay of vocal cord vibration relative to the onset of the stop consonant sounds. The acoustic cue to voicing in these synthetic speech sounds is the presence or absence of the first formant transition (F_1) and presence or absence of aspiration in the higher formants (F_2 and F_3). Second, the sounds differ in place of

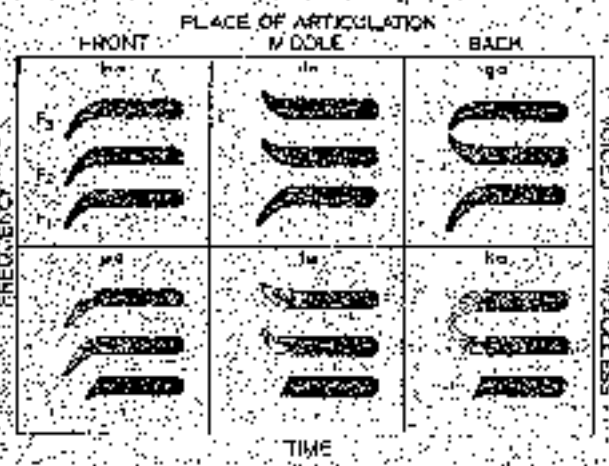

FIGURE 2. Schematized sound spectrograms of six stop consonant voiced syllables (synthetic speech). The voiceless sounds are characterized by the absence of a F_1 transition and the presence of aspiration in the F_2 and F_3 transitions.

articulation, the point of closure in the vocal tract in real speech. The acoustic cue to the place of articulation in these synthetic speech sounds is the direction and extent of the second and third formant transitions.

In our model, different features detected at the two ears are placed in preperceptual storage together. In Pisoni and McNabb's task, the test and interference syllable always differed in place of articulation but could be the same or different in voicing. We might expect that the interference syllable would disrupt recognition of the test syllable only to the extent that they have different acoustic features. If both syllables are voiced, subjects should have no difficulty determining the voicing of the test syllable since only the voiced feature (F_1 transition) is present in preperceptual auditory storage. If one of the syllables is unvoiced, however, identification of voicing should be degraded because of conflicting acoustic cues (F_1 transition and aspiration in F_2 and F_3). Pisoni and McNabb tested recognition of the test syllable as a function of whether the interfering syllable shared the voicing feature. For example, the syllables /ba/ and /ga/ share voicing whereas /ba/ and /ka/ do not. When the test and interference syllables had the same vowel, the test syllable was identified correctly about 90% of the time if the interference syllable shared the same voicing feature, and only 60% of the time if it did not. In terms of our model, the voicing of each syllable was detected at each ear and placed in preperceptual auditory storage. The featural information differs in the cases in which the simultaneous syllables share or do not share the voicing feature. When the syllables are both voiced or both unvoiced, there is no conflicting voicing information in preperceptual auditory store. If the syllables differ in voicing, however, conflicting acoustic features will be present in preperceptual storage and performance should be much poorer.

When one syllable is presented to one ear simultaneously with another syllable presented to the other ear, their features can be confused at the level of preperceptual auditory storage. Pisoni and McNabb's subjects were likely to hear the simultaneous syllables as one complex sound in the middle of the head. The two sounds were presented simultaneously at the same intensity. These acoustic cues signify one sound directly in *front* of or behind the observer. The syllables also had the same fundamental frequency (the acoustic cue responsible for voice pitch). This cue along with the same vowel context would serve to give the impression of a single speaker (who can be saying only one thing at a time). We might expect that identification of the test syllable would improve to the extent that the acoustic cues give information about two separate sounds instead of just one. One cue would be the relative onset time of the syllables. If the test sound is presented before the interference sound, the second sound will be less likely to fuse with the first sound. Pisoni and McNabb systematically varied the relative onset times of the syllables. Recognition of

articulation, the point of closure in the vocal tract in real speech. The acoustic cue to the place of articulation in these synthetic speech sounds is the direction and extent of the second and third formant transitions.

In our model, different features detected at the two ears are placed in preperceptual storage together. In Pisoni and McNabb's task, the test and interference syllable always differed in place of articulation but could be the same or different in voicing. We might expect that the interference syllable would disrupt recognition of the test syllable only to the extent that they have different acoustic features. If both syllables are voiced, subjects should have no difficulty determining the voicing of the test syllable since only the voiced feature (F_1 transition) is present in preperceptual auditory storage. If one of the syllables is unvoiced, however, identification of voicing should be degraded because of conflicting acoustic cues (F_1 transition and aspiration in F_2 and F_3). Pisoni and McNabb tested recognition of the test syllable as a function of whether the interfering syllable shared the voicing feature. For example, the syllables /ba/ and /ga/ share voicing whereas /ba/ and /ka/ do not. When the test and interference syllables had the same vowel, the test syllable was identified correctly about 90% of the time if the interference syllable shared the same voicing feature, and only 60% of the time if it did not. In terms of our model, the voicing of each syllable was detected at each ear and placed in preperceptual auditory storage. The featural information differs in the cases in which the simultaneous syllables share or do not share the voicing feature. When the syllables are both voiced or both unvoiced, there is no conflicting voicing information in preperceptual auditory storage. If the syllables differ in voicing, however, conflicting acoustic features will be present in preperceptual storage and performance should be much poorer.

When one syllable is presented to one ear simultaneously with another syllable presented to the other ear, their features can be confused at the level of preperceptual auditory storage. Pisoni and McNabb's subjects were likely to hear the simultaneous syllables as one complex sound in the middle of the head. The two sounds were presented simultaneously at the same intensity. These acoustic cues signify one sound directly in front of or behind the observer. The syllables also had the same fundamental frequency (the acoustic cue responsible for voice pitch). This cue along with the same vowel context would serve to give the impression of a single speaker (who can do saying only one thing at a time). We might expect that identification of the test syllable would improve to the extent that the acoustic cues give information about two separate sounds instead of just one. One cue would be the relative onset time of the syllables. If the test sound is presented before the interference sound, the second sound will be less likely to fuse with the first sound. Pisoni and McNabb systematically varied the relative onset times of the syllables. Recognition of

the test syllables that did not share voicing with the interference syllables improved as the relative onset time of the syllables was increased, reaching asymptote at roughly a 100 msec difference. Given that the differences in relative onset time signal two separate sounds, their features are not so readily confused in prepercepcual storage.

Pisoni and McNabb also showed that the detrimental effect of not sharing the voicing feature was attenuated as the similarity between the vowels of the test and interference syllables was decreased. Correct identification of the voicing of the test consonant in the vowel context /a/ improved from 60% to 73% to 85% when the vowel in the interference syllable was changed from /a/ to /æ/ to /i/. Analogous to the temporal order cue, we might expect that the sounds would be perceived as two separate sounds to the extent the syllables presented to the two ears had different vowel contexts. Pisoni and McNabb's results demonstrate how featural information from both ears is mixed at the level of prepercepcual auditory storage as described by our processing model. Future studies with this paradigm should illuminate the operations of feature detection and the properties of prepercepcual storage.

III. PRIMARY RECOGNITION

In 1958 Broadbent proposed that the human's interface to the stimulus world was organized along a series of independent sensory channels. Stimuli entered along each of these channels independently and in a simultaneous or parallel manner. Examples of channels were the auditory versus visual modality, locations in space within either of these modalities, or a speaker's voice quality, and so on. Broadbent (1958, 1971) viewed man as having a limit to the rate at which he could process information. Given this limited capacity, Broadbent assumed that this portion of the nervous system is preceded and protected by a selective device or filter which would only pass information along one channel at a time. The filter would reject all events except those possessing some common stimulus feature such as a given spatial location. The filter could change its basis for selection but this change would take a significant time called switching time. Figure 3 presents a flow diagram of Broadbent's model showing the two successive stages of processing.

Broadbent assumed that the short-term store held information along separate channels for 1 or 2 sec. Given simultaneous inputs to the two ears, the input coming in the left ear would be funneled to a different channel than a different input coming in the right ear. Processing must occur on one ear (channel) followed by a switch to the other ear since the inputs could not be processed simultaneously or in parallel. Broadbent's model

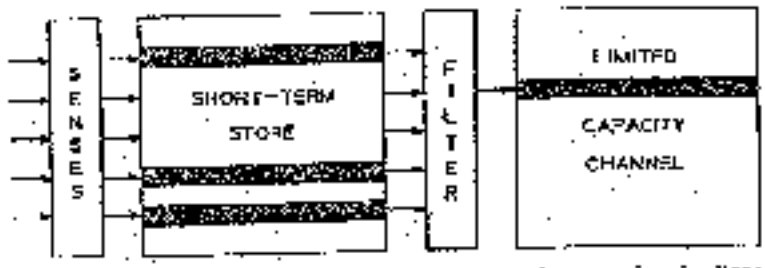

FIGURE 5 Flow diagram of two successive stages of processing in Broadbent's (1958) model.

contrasts nicely with our view of information processing. The first critical difference is whether, in fact, preperceptual storage (short-term store in Broadbent's model) is organized along independent sensory channels. Second, we assume that the span of this storage structure is on the order of .25 sec, whereas it is at least a factor of four times this in Broadbent's model. Third, we assume that a second input can interfere with the preperceptual storage of an earlier input whereas information in Broadbent's short-term storage simply decays passively with the passage of time. A series of experiments is now available which allows us to choose between these two pictures.

A. Backward Recognition Masking

One paradigm developed to study the properties of preperceptual auditory storage and the temporal course of primary recognition is temporal recognition masking (Massaro, 1970c). In this task, it is possible to evaluate the influence of one stimulus on the recognition of another when they occur at different points in time. In the recognition paradigm, the observer's task is to recognize two or more test signals. The test signals might be short tones, differing in pitch, timbre or loudness, vowels, or short syllables. In backward masking, one of the test signals is randomly presented on each trial, followed by a silent interval, followed by a masking stimulus. The masking stimulus precedes the test stimulus in the forward masking task. The observer's task is to identify the test signal as one of a fixed set of alternatives.

If preperceptual storage were simply subject to a passive decay as assumed by Broadbent, the test and masking sounds should not interfere with each other at the level of preperceptual storage. In contrast, we assume that the capacity of preperceptual storage is one sound; a second sound will displace the first at this level. Accordingly, we predict that a second sound will interfere with recognition of an earlier sound if it occurs

before the first sound has been recognized. In contrast, a masking sound will not interfere with perception of a following test sound since the test sound will replace the earlier masking sound in preperceptual storage. In a recent experiment of backward and forward masking (Massaro, Cohen, & Idson, 1976), the test signals were two tones that differed in frequency, sine waves of 790 and 860 Hz. The masking tone was a sine wave of 900, 825, or 750 Hz, and the duration of the test and masking tones was 20 msec. On some trials, no masking tone was presented. The test and masking tones were equated for loudness and presented at a normal listening intensity. The masking tone preceded or followed the test tone after a variable silent intertone interval. The observers were practiced in the task and all experimental conditions were presented randomly within a given test session. Each trial began with a visual cue signifying whether the first or second tone should be identified.

Figure 4 shows the average recognition performance of eight observers. In backward masking, recognition was critically dependent on the intertone interval showing that primary recognition was not complete at the end of the 20-msec test tone presentation. The poor performance at short intertone intervals is consistent with the idea that the masking tone interfered with the preperceptual auditory storage of the test tone. Performance continued to improve with increases in the silent intertone interval to

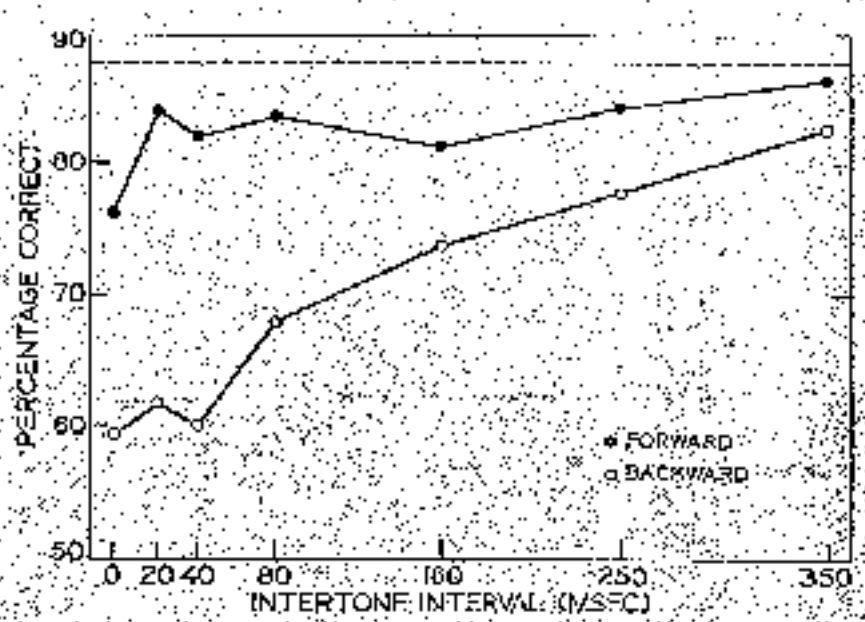

FIGURE 4 Percentage of correct recognition as a function of the intertone interval under backward and forward masking. The dotted line gives performance when no masking tone was presented. (After Massaro, Cohen, & Idson, 1976.)

250–350 msec. In forward masking, very little interference occurred when the masking tone preceded the test tone by 20 msec or more. This result supports the idea that a second sound replaces an earlier sound in preperceptual storage. Since the test tone replaced the earlier masking tone in storage, no forward masking of test tone recognition should occur except for any mutual degradation occurring during the roughly 60 msec period of temporal integration (Massaro, 1973, 1975a).

The results can be interpreted in terms of the capacity and duration of preperceptual auditory storage and the temporal course of recognition. Recognition of the test tone was not complete at the end of its 20-msec presentation. Accordingly, some preperceptual storage was necessary in order to hold the information in the tone presentation for the primary recognition process. A masking tone presented after the test tone appeared to terminate the primary recognition process. The resolution of the synthesized percept, passed on by primary recognition, is therefore a direct function of the processing time available before the onset of the masking tone. Given that recognition leveled off at somewhere around 250 to 350 msec, this is the best estimate of the effective duration of the preperceptual storage of a 20-msec tone.

The primary recognition process can be described quantitatively by the equation

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

where d' provides a measure of how well the high and low tones can be recognized, α is the amount of information in the stimulus, θ is the rate of processing the information, and t is the available processing time before the onset of the masking stimulus. The equation says that performance improves as a negatively accelerated function of processing time with an asymptote at α. In terms of features, the equation might be interpreted as saying that a fixed proportion of those features that remain unprocessed are processed per unit of time.

The observed performance at intertone intervals of 0 to 40 msec in backward masking is usually better than would be expected from the negatively accelerated function described in Equation (1) and reflects the integration of the two sounds at very short silent intervals (see Figure 4). When two sounds are integrated, the second does not serve to interrupt processing of the first as is the case when the second falls outside the period of integration. At very short intervals, the sounds are integrated into a single two-sound composite and primary recognition operates on this complex sound. Performance at the short intervals reflects maximal processing time of a complex sound that has reduced the signal-to-noise ratio of the test sound. The upturn in performance at the short intervals can be described by a smaller value of α with unlimited processing time.

Under these conditions, it is assumed that the masking syllable is integrated with the test syllable decreasing α, the amount of information in the stimulus. Processing time is unlimited because the integration prevents masking. Massaro (1975a) presents a more detailed discussion of integration in auditory and visual masking.

Stages of Processing

One critical dimension of information-processing research is to specify exactly the stage of processing responsible for observed results (Massaro, 1975a). Each stage of processing must be accounted for in the backward recognition task in order to conclude that the results do, in fact, measure the dynamics of primary recognition. Perceptual, memory, and decision processes are important in the recognition task. We want to attribute the results to perceptual changes so that synthesized auditory memory and decision differences must be eliminated as possible causes. Subjects must learn and remember what the high and low tones sound like in order to recognize them correctly. Performance in this task is a direct function of the subject's memory for the test stimuli. If the masking conditions were tested between blocks of trials, differences in memory might contribute to performance differences. Consider the comparison between a no-masking condition and a masking one when the conditions are tested between blocks of trials. The masking tone can interfere with the subject's synthesized auditory memory of the test tones even if it does not interfere with primary recognition (Massaro, 1975a). In this case, subjects may have more difficulty in recognizing the test tones in the masking than no-masking condition, not because of perceptual differences but because of memory differences in the two conditions. To control for memory differences, we randomize all of the experimental conditions within a session so that memory does not change with changes in the level of the independent variable (see Massaro, 1975a, c).

Control is exercised for the decision process by using a measure of performance that is relatively independent of response biases. The forced choice task with a fixed set of alternatives allows the investigator to derive a valid index of the perceptual process. The percentage correct averaged across trial types is a good measure since it adjusts for any bias to respond with a particular response alternative. Analogously, it is possible to treat the probability of responding high on high and on low trials, respectively, as hit and false alarm rates in the formal theory of signal detectability. These values can then be translated into d' values. The experiments are, therefore, applicable to theories that utilize d' values to index perceptual processing. The recognition probabilities can also be used to measure any decision bias to respond high or low across the experimental conditions.

In this case, the investigator has a measure of both perceptual and decision processes. With these methodological and analytic procedures, the backward masking paradigm appears to allow the investigator to measure the dynamics of the primary recognition process without a confounding of memory and decision processes.

It is possible that changes in the intertone interval in the masking task influence the decision rather than the perceptual process. Treisman and Rostron (1972), for example, state that the backward masking results "are susceptible to an explanation which does not assume sensory storage, namely that the masking tone adds variance to the judgmental processes at an early vulnerable period [p. 162]." Treisman and Rostron do not expand on this interpretation but it seems to mean that backward recognition masking is caused by changes in the decision (judgmental) system with changes in the intertone interval. Can the recognition masking results be explained by differences at the level of the decision system?

We assume that two stages of processing are necessary in the recognition masking task. The first stage of processing corresponds to a perceptual resolution of the information available in the preperceptual storage. This process takes time and a second masking stimulus is assumed to interfere with the process by interfering with the information available in preperceptual storage. The perceptual process outputs what it knows about the test tone to the decision system which translates this information into an appropriate response. The first perceptual operation should be influenced by variables such as the discriminability of the test tone and the amount of time its preperceptual storage is available for perceptual processing. The decision operation is usually assumed to be affected by judgmental variables such as the payoffs in the experimental task and the a priori probability of a particular signal.

It is logically possible, however, that the decision system is responsible for the backward masking functions. Assume that the subject adopts a criterion along the high–low continuum and responds high if the perceptual observation is on the high side of the criterion and low if the observation is on the low side of the criterion. It is possible that the criterion point varies from trial to trial and varies more at the short masking intervals. The increased variability in the criterion at the short masking intervals could account for the poorer performance observed in the task at these intervals. A criterion explanation must be post hoc since it has no mechanism to predict when masking will and will not be observed. This becomes apparent when one tries to account for all the available results on masking in terms of a decision process. For example, a decision explanation has no mechanism to account for the differences observed when the masking tone precedes or follows the test tone presentation (see Figure 4). Dorman, Kewley-Port, Brady, and Turvey (1973), Massaro (1970c, 1973),

and Wolf (1974) have also found significantly less forward than backward masking. These results are consistent with a two-process theory of masking (for example, Massaro, 1975a), and are not easily explained by a decision operation. Also Massaro and Kahn (1973) showed that the masking stimulus must be an auditory one to produce recognition masking. If subjects are required to process a light at varying times after the test tone presentation, no backward masking is observed. If processing a second stimulus adds variance to a decision operation, as assumed by Treisman and Rostron, we should also expect backward masking with the light mask. This result and the differences between backward and forward masking argue against a decision interpretation of the backward masking results.

B. Number of Channels

The recognition masking paradigm permits discovery of a number of properties of preperceptual auditory storage. We have seen that a second stimulus interferes with storage of an earlier stimulus in preperceptual storage, a result opposed to Broadbent's decay principle. We also assume that all inputs converge on a single channel in preperceptual auditory storage whereas Broadbent assumed that inputs are stored along separate channels according to the ear of presentation. One critical test between the two models, therefore, involves presenting the test and masking tones at the same ear or different ears and seeing if backward masking occurs in both cases. From Broadbent's model, we would expect much less, if any, backward masking when the test and masking tones are presented to opposite ears than when they are presented to the same ear or to both ears. The same sound presented to both ears localizes the sound in the middle of the head. We expect no difference from our model since both ears converge at preperceptual auditory storage.

Massaro and Cohen (1975) included two masking conditions in a study of recognition of the synthetic consonant vowel (CV) syllables /ba/, /da/, and /ga/. The syllables were synthesized at a duration of 42 msec; the first 20 msec constituted the CV transition and the last 22 msec the steady-state vowel (see Figure 2). The masking stimulus was also chosen randomly from this set of three CV stimuli. In the dichotic condition, the test and masking stimuli were presented to opposite ears. In the monaural condition, both CV syllables were presented to the same ear.

The results in Figure 5 display the similarity in the shape of the curves under the dichotic and monaural masking conditions. A further study (Massaro, 1975c) discussed in Section III.C replicated the dichotic masking results even when the subject was given advance knowledge of

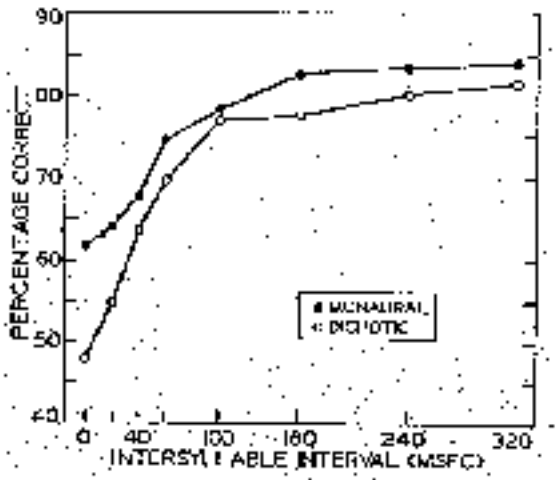

FIGURE 5 Percentage of correct recognition as a function of the intersyllable interval between the test and masking syllables under the dichotic and monaural masking conditions. (After Massaro & Cohen, 1975.)

the ear of the test sound presentation. The masking with dichotic stimuli allows a rejection of Broadbent's model in favor of a single channel preperceptual storage. One difference between the two masking functions in Figure 5 should be noted. Performance is significantly better at the zero and 20 msec silent intervals in monaural than in dichotic masking. This outcome is related to the earlier finding that performance is better than might be expected at backward masking intervals less than 40 msec (see Figure 4). The result reflects an integration of the test and masking stimuli at short intervals so that the masking stimulus does not replace the test stimulus. Instead, the masking stimulus decreases the information in the test stimulus but does not terminate the processing of the information in the test–mask composite. This integration does not occur as much in the dichotic case so that a steeper monotonic masking function is observed. The two curves come together within 40 msec showing that the inputs from the left and right ears must converge on the same storage area, producing the same interference in the dichotic and monaural case.

The dichotic masking results indicate that the two ears do not function as separate stimulus channels at the primary recognition stage of processing. In fact, localizing a sound in space should also require time for primary recognition to occur. A second sound presented before the first is localized will either be integrated with the first or serve to interrupt primary recognition processing of the first. Subjects were asked to localize a 20-msec test tone as coming from the right or the left (Massaro, Cohen, & Idson, 1976). The test tone could be followed by a 20 msec masking tone that came from the right, center, or left with respect to orientation of

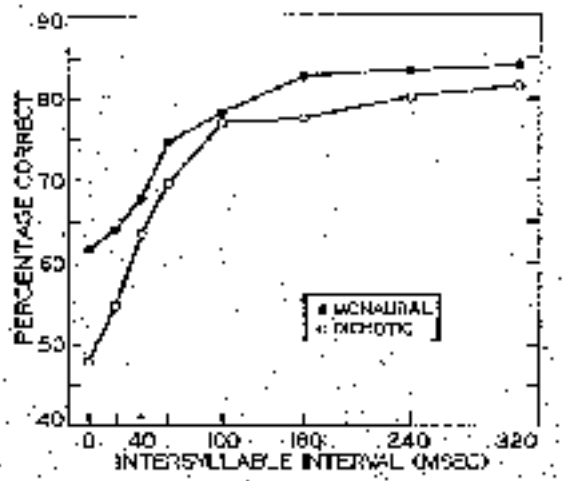

FIGURE 5 Percentage of correct recognition as a function of the intersyllable interval between the test and masking syllables under the dichotic and monaural masking conditions. (After Massaro & Cohen, 1975.)

the ear of the test sound presentation. The masking with dichotic stimuli allows a rejection of Broadbent's model in favor of a single channel pre-perceptual storage. One difference between the two masking functions in Figure 5 should be noted. Performance is significantly better at the zero and 20 msec silent intervals in monaural than in dichotic masking. This outcome is related to the earlier finding that performance is better than might be expected at backward masking intervals less than 40 msec (see Figure 4). The result reflects an integration of the test and masking stimuli at short intervals so that the masking stimulus does not replace the test stimulus. Instead, the masking stimulus decreases the information in the test stimulus but does not terminate the processing of the information in the test–mask composite. This integration does not occur as much in the dichotic case so that a steeper monotonic masking function is observed. The two curves come together within 40 msec showing that the inputs from the left and right ears must converge on the same storage area, producing the same interference in the dichotic and monaural case.

The dichotic masking results indicate that the two ears do not function as separate stimulus channels at the primary recognition stage of processing. In fact, localizing a sound in space should also require time for primary recognition to occur. A second sound, presented before the first is localized will either be integrated with the first or serve to interrupt primary recognition processing of the first. Subjects were asked to localize a 20-msec test tone as coming from the right or the left (Massaro, Cohen, & Idson, 1976). The test tone could be followed by a 20-msec masking tone that came from the right, center, or left with respect to orientation of

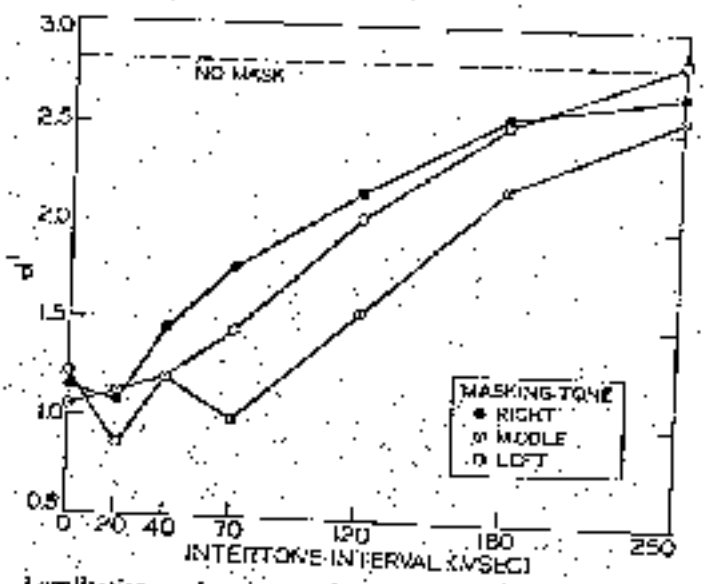

FIGURE 6. Localization performance measured in d' values as a function of the duration of the silent intertone interval. The parameter right, middle, or left gives the location of the masking tone. (After Massaro, Cohen, & Idson, 1976.)

the head. The direction of the tones was varied by producing interaural intensity differences to the two ears. Subjects classified the test tone as coming from the right or left.

Figure 6 presents identification performance, in terms of d' values, as a function of the intertone interval and the location of the masking tone. Localization performance improved dramatically with increases in the silent interval between the test and masking tones. These results show that processing time is required to localize a sound in the same way that it is required to recognize the pitch of a tone or the sound quality of a CV syllable (see Figures 4 and 5). The results support the idea that inputs to the two ears converge on a central preperceptual storage and primary recognition must occur before the sound is experienced at some location in space.

C. Selective Perception

The concept of a single-channel preperceptual auditory storage illuminates recent studies of attentional effects in auditory recognition. Moore and Massaro (1973) asked whether observers could selectively process one dimension of an auditory stimulus. On each trial, subjects identified either

the loudness, timbre, or both dimensions of a short test tone in a backward masking task. Performance improved with increases in the silent intertone interval, but was not dependent on whether subjects identified one or two dimensions of the test tone. This result shows that the primary recognition process could not selectively attend to certain features of timbre or loudness to improve its performance relative to the situation in which both dimensions must be identified. Given that the features are stored centrally along a single channel in our model, we would not expect any attention effects at this stage of processing.

Shiffrin, Pisoni, and Castañeda-Mendez (1974) asked whether subjects could selectively attend to one ear at the expense of processing information at the other. Subjects identified CV syllables under two attention conditions. In the divided attention condition, the test syllable could occur in either of the two ears. In the selective attention condition, the subjects monitored first the left ear and then the right ear. The syllable was presented to one of the ears during the appropriate monitoring interval. If it is possible to allocate a limited amount of processing capacity to one ear, we would expect that syllable identification would be better in the selective than in the divided attention case. If preperceptual storage does not hold the inputs to the two ears along separate channels, however, one would not expect selective attention effects. Shiffrin *et al.* (1974) found no attentional effects in two experiments, supporting our representation of preperceptual auditory storage.

Three procedural details of the Shiffrin *et al.* (1974) study might limit the generality of the answer to their question. First, a continuous background of white noise from separate noise generators was fed into each of the two ears independently of one another. With uncorrelated noise sources to the two ears, the listener experiences changes in the number of noise inputs and movement of the apparent locations of the noise sources. The subjects used by Shiffrin *et al.* (1974), therefore, may have found it difficult to sustain their attention at a given ear location. It would have been preferable to present degraded syllables without background noise to test the attention hypothesis. Second, Shiffrin *et al.* (1974) did not follow the test syllable with a masking stimulus so that processing may have continued after presentation of the test syllable. Finally, the selective and divided attention conditions were compared at only one performance level. This procedure cannot eliminate the possibility of a ceiling and floor effect for some of the subjects and/or some of the test stimuli. To eliminate this possibility, attentional effects should be evaluated across a number of performance levels. This also allows the investigator to measure how attention interacts with the dynamic course of the perceptual process.

The backward recognition masking task provides an ideal paradigm for studying attentional effects in perception. The masking stimulus controls

the amount of processing time of the test stimulus. Accordingly, selective and divided attention can be compared at processing times too short for switching of attention to be helpful. The backward masking task also allows the investigator to test for attentional effects at a number of performance levels, mitigating problems with ceiling and floor effects. The significant effect of processing time in the backward masking task would also allay potential criticisms of proving the null hypothesis when no selective attention is found. If attentional effects are found, the backward masking task allows one to measure the dynamic aspects of the process. In summary, the backward masking task gives the investigator precise control over the amount of stimulus information and processing time in the selective and divided attention tasks.

In order to provide a stronger test of selective perception with respect to spatial location, subjects were asked to recognize a test tone as high or low (Massaro, 1975c). The test tone could be presented to either of the two ears followed by a masking tone presented to the opposite ear after a variable silent interstone interval. Before each trial, subjects were cued that the test tone would be presented to the right, left, or to either ear. These practiced subjects also knew that the masking tone would be presented to the opposite ear. Therefore, given the cues right or left, subjects attempted to attend to the designated ear and to block out the unattended ear. Given the either cue, however, subjects had to divide their attention between the two ears since the test tone could occur on either ear. Figure 7 shows plots of performance under the selective and divided attention conditions as a function of the interstone interval. Supporting earlier dichotic masking studies, performance improved with increases in

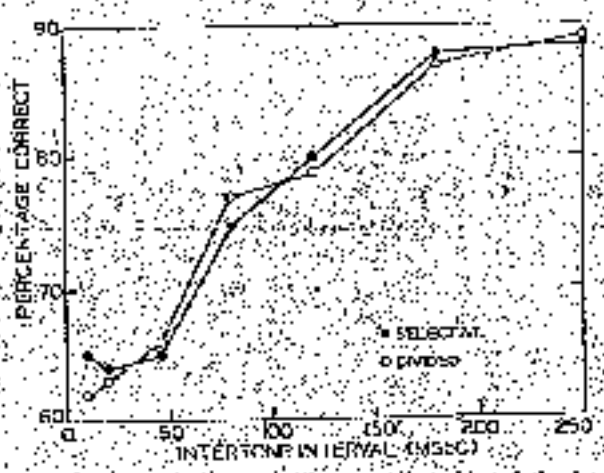

FIGURE 7. Percentage of correct recognitions as a function of the intertone interval under the selective and divided attention conditions. (After Massaro, 1975c.)

presenting tone before the masking tone. And in agreement with our results and the Shiffrin *et al.* findings, no differences were found between the selective and divided attention conditions.

The failure to find selective attention effects within the process of primary auditory recognition contrasts with positive attention effects when auditory and visual recognition are required. In a study by Massaro and Kahn (1978), simultaneously identifying the duration of a light decreased tone identification relative to the selective attention case in which just the tone had to be processed. This result shows that the process of primary recognition itself may be limited in capacity and dependent on allocation of a central processing capacity. Given separate preperceptual stores for auditory and visual inputs, identifying the duration of the light can subtract from the processing capacity available for primary auditory recognition. Within the process of primary auditory recognition, however, no further division of processing capacity may be possible. Figure 1 shows that primary recognition involves a readout of a single preperceptual channel of auditory features. Accordingly, it is not possible to attend to some of the features at the expense of other features. The nature of the storage structure is such that intensity cannot be processed independently of frequency and time spatial location cannot be processed independently of another. Figure 1 shows that spatial location information must be read out of preperceptual store along with sound quality. Accordingly, Shiffrin *et al.*'s (1974) and our subjects had no spatial location channel to selectively attend to at the primary recognition stage of processing. A positive result in that study would have been dependent on a Broadbentian short-term store. Similarly, in the Moore and Massaro study, both attributes of the sound must be read out of preperceptual storage together so that once again, selective attention cannot occur.

D. Partial Report

Preperceptual storage can also be studied in a task modeled after the partial report paradigm used in visual information processing (Averbach & Coriell, 1961; Dick, 1974; Sperling, 1960). Sperling (1960) presented subjects with a visual display of three rows of three letters each. A tone-report cue, indicating which of the three rows should be recalled, was presented either before or sometime after the 50-msec display presentation. Sperling found that recall of the cued row was almost perfect if the cue occurred simultaneously with the display presentation and decreased with increases in the delay of the cue. The advantage of the partial report at short delays was taken as evidence that the subject retains a visual representation of the display after the display presentation is terminated. This

visual representation allows the subject to access the items directly on the basis of spatial location.

In another of Sperling's experiments, the partial report cue referred to item category rather than spatial location. In this study, two letters and two numbers were presented in each of two rows. Either before or immediately after the display presentation, one of the two report cues was presented designating whether the numbers or the letters in the display were to be recalled. Partial report by category name was much inferior to partial recall by spatial location even when the category name was given seconds before the visual presentation. This result indicated that subjects could not access the items in the visual representation on the basis of item category. Logically, if the visual representation were preperceptual, category information would not be available and each item would have to be recognized before it could be accessed on the basis of category name. In this case, the subject would be limited by the span of immediate memory as in the whole report condition. This result, in conjunction with the positive results with recall by location, supports the hypothesis that the visual representation that remains after the short display presentation is a preperceptual one.

Darwin, Turvey, and Crowder (1972) utilized the Sperling partial report procedure with auditory stimuli. Three lists of three consecutive items (letters and digits) were presented to each of three spatial locations (left ear, center, and right ear). In the first study, subjects were cued to report the items at one of the three locations 0 to 4 sec after presentation of the auditory lists. In the second experiment, the report cue indicated whether the category letters or digits should be reported. Subjects were required to recall the items (digits or letters) indicated by the category cue and also had to report the spatial location of the items. The subjects had to know exactly as much in the second experiment as in the first to perform the partial report task accurately. In both experiments, the subject had to know the name and location of the item for correct recall. The only thing that differed between the two experiments was that the subject was directed to access the items either by spatial location or by category name. If we simply compute the number of items correctly recalled, the results indicate that partial recall is actually poorer when the subjects are cued by location than when cued by category name. However, the problem with a direct comparison of the spatial location and category name conditions is that the number of report cues and the number of items per cue differ in the two conditions. Accordingly, a series of studies was carried out to provide a direct comparison between cueing recall by spatial location and by category name (Massaro, 1972b, 1976b).

In one experiment two lists of four items each were recorded by different speakers at a rate of 4 items per second. These lists were presented simultaneously to the two ears. Each list on each channel contained 2 one-

syllable words and 2 letters chosen randomly without replacement from respective master sets of 25 one-syllable words and 25 letters. When subjects were asked to recall by location the cue indicated whether they should report the items presented to the left or right ear. When subjects were asked to recall by category, the cue indicated whether they should report the words or letters. Given these two report conditions (location and category) the results are directly comparable. In both report conditions there were two possible report cues and the subjects were required to report four items. The exact same lists and cues (pure tones) were used in both report conditions. The cue was presented immediately after the last item in the list. Therefore, if the items contain preperceptual auditory information that is more easily accessed along stimulus than name dimensions, recall should be superior in the location than in the category condition.

The results showed no advantage of recall cued by spatial location over recall cued by category name, when the report cue was given immediately after the list of auditory items (Massaro, 1976b). Two limitations are apparent in these findings. First, the study simply failed to reject the null hypothesis of no difference. The failure to find a difference may have been due to some weakness in experimental procedure. Second, there could be something inherently difficult about recalling items by spatial location in this task that has nothing to do with the nature of the memory storage of the test list. Treisman and Rostron (1972) point out that observers find it more difficult to locate items in auditory than visual space. To overcome these limitations, the partial report task described in the last paragraph was replicated in an experiment in which the report cue was presented seconds before or immediately after the test list. Figure 8 presents the mean number of items recalled when the location and category cues were given before or after the test list. Recall by spatial location was much better than recall by category name when the cue was given before the list presentation. Delaying the report cue until immediately after the test list decreased recall by spatial location by 25% but had an insignificant effect on recall by category name. Replicating the earlier studies, no difference in recall occurred when the cues were given after the test list.

The results show that two simultaneous lists of items are not maintained along separate auditory channels in a preperceptual form for 1 or 2 sec. Rather, the items are first stored centrally in preperceptual storage. Primary recognition resolves the items at a perceptual level and locates them at different locations in space. Secondary recognition then processes the items for meaning. If the report cue for spatial location is given before the test list, secondary recognition can devote most of its processing capacity (but not all of it) to the items perceived at a given location in space. In recall by category name, the cue before the list does not reduce the

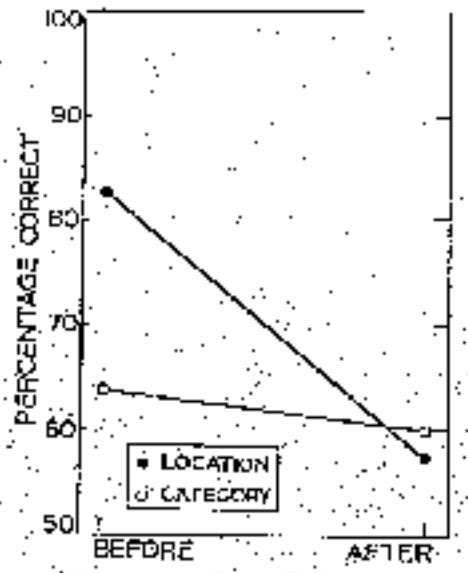

FIGURE 8. Percentage of items correctly recalled as a function of whether the report cue was given seconds before or immediately after the test list. The parameter specifies when recall was cued along spatial location or category name. (After Massaro, 1976b.)

processing required by secondary recognition. Each item must still be processed for meaning to determine if it belongs to the appropriate category for recall. Accordingly, more items can be reported correctly in recall by spatial location than recall by category name when the cue is given before the test list. If subjects are not cued until after the test list, the items have been processed by both primary and secondary recognition. In this case, retrieval by spatial location no longer shows an advantage over recall by category name.

Positive results in the partial report paradigm are dependent on the ability to selectively access preperceptual auditory information along some physical dimension such as spatial location. Broadbent's model allows for such selection whereas our model does not. Preperceptual storage is conceptualized as a single channel in our model preventing any selective access at this stage of processing. In agreement with this model, experimenters have failed to demonstrate an advantage of partial report by spatial location when the items are presented simultaneously to different locations in space and the recall cue is given after the test list. These results agree with the findings of dichotic interference in the backward recognition masking task and the results that primary recognition cannot selectively process the stimulus dimensions of timbre, loudness, and spatial location.

IV. PREPERCEPTUAL AND SYNTHESIZED CODES

The present model assumes that preperceptual auditory storage and synthesized auditory memory operate at two successive stages of auditory processing. A nice example of the different roles played by the two memories is seen in a memory for pitch task. In this task, observers are presented with a standard tone followed immediately by an interference tone and then a comparison tone. The observer reports whether the comparison tone is the same or different in pitch as the standard tone. According to the present model, preperceptual and synthesized auditory memory play different roles in this experimental task. The information available in the preperceptual auditory image and the time this information is available for perceptual processing determine the perception and storage of the standard test tone. The perceptual processing of the information in the preperceptual auditory image produces a synthesized percept that enters synthesized auditory memory. Synthesized auditory memory holds a representation of what the standard test tone sounded like during the interference tone presentation. The subject bases his decision on a comparison between the comparison tone and the representation of the standard tone in synthesized auditory memory.

If preperceptual and synthesized storage are different structures in the pitch memory task, these should be affected differently by independent variables in the experiment. Preperceptual memory is responsible for only the original perception and synthesis of the test tone; it should be affected by the signal-to-noise ratio of the test tone presentation and the time the information is available for perceptual processing. It should not be influenced by other variables that affect the synthesized memory for the test tone. In contrast, the temporal course of synthesized memory should be influenced by retention variables and not by the variables that affect the perception and storage of the standard tone.

Support for these assumptions comes from a number of studies that show perception and retention are successive processes in the pitch memory task. Increasing the duration of the standard test tone enhances the original perception and storage of that test tone but does not affect the rate of forgetting of the test tone during the interference tone presentation (Massaro, 1970b; Wickelgren, 1969). Similarly, increasing the duration of the interference tone increases the forgetting of the test tone but does not affect the original perception and storage of the test tone presentation. These conclusions are substantiated by a quantitative description of performance in the memory for pitch task (Massaro, 1970b).

Kinchla (1973) presented subjects with a compound sound made up of a 1000-Hz and a 500-Hz tone for 100 msec. A second pure tone of either

1000 Hz or 500 Hz followed the first compound sound after a silent interval of .5 to 2.0 sec in a delayed comparison task. Subjects decided whether the second tone was softer than or equal to the loudness of the same-frequency tone in the compound sound. The absolute loudness of each of the tones in the compound sound was varied over a wide range to insure that the listener's decision was based on memory for the tones of that trial. If only a few loudnesses were used, the observers would build up a long-term memory representation of the tones and would be able to categorize the second tone on this basis, bypassing the delayed comparison judgment. For a more detailed discussion of procedures in the delayed comparison task, see Massaro (1975a, Chapter 24). A visual cue, presented 1 sec before or .75 sec after the standard sound, instructed the listener to attend primarily to one of the two frequencies of the compound sound.

The attention cue was effective in that subjects performed significantly better on the cued than the uncued frequency component in the delayed comparison task. Performance in the task can be evaluated in terms of storage of the standard tones and the retention of this information during the interval between the standard and comparison tones. The cue presented before the standard sound improved both the storage and retention of the cued tone relative to the uncued tone. The cue presented after the standard sound had no effect on the storage of the two tones but did improve retention of the cued tone over that for the uncued tone.

Kinchla's results contrast with those of Moore and Massaro, who found no selective attention in the recognition of timbre and loudness. A difference in the processing stages tapped in the two studies might account for the different results. Subjects in Kinchla's task had to store and remember the test sound for a later comparison whereas subjects in the backward recognition masking task of Moore and Massaro had to recognize the timbre and/or loudness of the test sound. A person may not be able to selectively attend during recognition but might be able to selectively store and remember some sound. The temporal course of recognition in the absolute identification task in backward masking appears to be much shorter than the temporal course of storage in the delayed comparison task. A number of studies have shown that recognition performance asymptotes at roughly 250 msec in the backward masking task. This means that recognition processing is ineffective 250 msec after the onset of the test tone. In contrast, a number of studies have shown that memory for a test tone improves with increases in tone duration out to a couple of seconds (Massaro, 1975a; Wickelgren, 1969). In addition, Massaro (1970a) showed that the storage of a tone continued to occur during the silent interval after presentation of a 200 msec tone in a delayed comparison task. It is likely that subjects in Kinchla's study continued to process the 100 msec

standard tone after it was presented. Selective attention to the cued test tone may have only been effective after the initial recognition or perception of the standard tone had occurred. In terms of our model, subjects may have resolved both the cued and uncued test tones to the same degree at the level of primary recognition but then selectively attended to the appropriate percept at the level of synthesized auditory memory. Figure 1 shows that different tones could be processed along separate channels in synthesized auditory memory so that selective attention would be possible even though it is not possible in the readout of preperceptual auditory storage.

Preperceptual and synthesized storage may also be distinguished on the basis of similarity effects as predicted by the number of channels available in storage. The similarity between the test and masking stimulus does not appear to be critical for discrimination in the recognition masking task. Massaro (1970c) found no differences in the amount of masking as a function of the frequency-similarity between the test and masking tones. Massaro and Cohen (1975) found that a white noise stimulus produced the same amount of masking as a speech stimulus when the test stimuli were CV syllables in the backward masking task. Pisoni found no difference in backward masking of CV syllables as a function of an intensity difference between the test and masking syllables. These results provide some preliminary evidence that the preperceptual auditory storage of a stimulus can be disrupted by a second auditory stimulus regardless of the similarity relationship between the two stimuli. The results support the idea of a single channel at the level of preperceptual auditory storage.

In contrast to preperceptual storage, synthesized auditory memory is organized along separate channels and information held there is sensitive to the similarity of the auditory stimuli that follow. When a list of auditory items is followed by an auditory suffix, *zero*, that does not have to be recalled, the interference effect of the suffix on recall of the items in the test list is dependent on the physical similarity of the suffix to the test list. Presenting the test list in a different voice than the voice of the suffix or to a different location than the test list reduces the interference effect (Morton, 1970). Presenting the suffix at a different intensity or at a different duration than the items in the test list also reduces its interference (Crowder, 1973; Morton, 1970). The large similarity effects in the suffix paradigm, which supposedly taps synthesized auditory memory, and the lack of similarity effects in the backward masking paradigm, which is assumed to measure preperceptual auditory storage, argue for the separateness of these two memories.

The differences in preperceptual and synthesized memory are also seen in the different effects of the interference and masking tone in the pitch memory and recognition masking tasks. Performance in the recognition

masking task is assumed to be a direct function of the perceptual processing of the test tone before the onset of the masking tone. Given that the masking tone terminates perceptual processing of the test tone, increasing the duration of the masking tone beyond some minimal value should cause no further decrement to recognition. Two experiments varying the duration of the masking tone between 20 and 500 msec found no differences in the masking function (Massaro, 1971, 1975a). In contrast, increasing the interference tone duration in the memory for pitch task does not affect the original perception of the standard test tone, but does increase the forgetting of the standard tone (Wickelgren, 1969). This analysis of recognition masking, tone memory, and the suffix paradigm shows that preperceptual and synthesized memory play different roles in the processing of auditory information.

V. SECONDARY RECOGNITION

Primary recognition resolves the features held in preperceptual auditory storage into a percept held in synthesized auditory memory. The outcome of primary recognition is the phenomenological experience of perceiving a sound of a certain quality and loudness at a particular location in space. Secondary recognition resolves this percept into meaning held in generated abstract memory. We assume that primary recognition develops a series of perceptual dimensions or channels in synthesized auditory memory. The secondary recognition process can follow only one of these channels at a time. If the secondary recognition process is required to monitor two channels instead of just one, a performance decrement should be found.

A. Processing Speech Alternated between the Ears

Cherry and Taylor (1954) seemed to have provided a convincing demonstration of a limitation in secondary recognition. Their subjects repeated back (shadowed) messages read from literary texts at a rate of roughly 130 words per minute. It should be noted that success in shadowing is critically dependent on deriving meaning from the message. Unless encouraged to do otherwise, the shadower tends to repeat the message in phrases rather than shadowing syllable by syllable (Cherry, 1953; Marslen-Wilson, 1973). Shadowing performance improves with increases in the syntactic and semantic redundancy in the message (Rosenberg & Lambert, 1974; Treisman, 1965). We can be confident, therefore, that the shadowing task is tapping secondary recognition in our model.

Cherry and Taylor alternated the message at different rates between the ears. At relatively slow rates of alternation, producing speech segments

of 1 sec on each ear, shadowing performance was asymptotic. Performance decreased with increases in alternation rate, reaching a minimum when the duration of the segments to each ear was around 160 msec. Cherry and Taylor concluded that the shadower had to switch his attention alternately between the ears in order to recognize the message correctly. By assuming that switching attention takes time, shadowing performance will be disrupted to the extent switching time consumes most of the processing time available in the task. Performance will improve as the rate of alternation is slowed down because less time will be consumed by switching attention between the ears.

Given this explanation, one is surprised that shadowing performance improved with even faster rates of alternation that give less than 160 msec of signal to each ear before switching. Subjects were more accurate in shadowing when speech was alternated in 50-msec segments than for 100 msec segments. The reason for this, according to Cherry and Taylor, was that subjects only attend to one ear at the faster rates and are able to recognize the message from the small speech segments that alternate with small segments of silence on the attended ear. Neisser (1967) illustrates this point very nicely using a visual analog.

Huggins (1964) replicated Cherry and Taylor's experiment but used a variable speed tape recorder to present his passages at two different rates. He found that the rate of alternation required for minimal intelligibility was dependent upon the speech rate. Faster rates of alternation were needed for minimal intelligibility at the faster speech rate. This result weakened the attention switching hypothesis since a central switching mechanism should not be influenced by the rate of the speech signal but simply by the rate of alternation. Huggins (1964) used these results to argue that the size of the speech signal was critical and that alternating segments corresponding to about one half of a syllable provided the most interference. This led Huggins to argue for the syllable as the critical unit in speech perception (Huggins, 1964; Massaro, 1972a; Neisser, 1967). Huggins's results have been replicated and extended by Wingfield and Wheatle (1975), using time compressed speech to speed up the speech passages.

Although the experiments by Huggins and Wingfield and Wheatle show that attention switching alone cannot account for the effects of changes in speech rate, switching attention may still be involved in the task. In the next section, we show that a sound must be perceived at a given location before it requires a switch of attention. At relatively slow rates of alternation, the shadower has time to perceive the location of the message and to switch his attention there so that its meaning can be processed. As the rate is speeded up, less time is available for switching so that shadowing deteriorates, reaching the minimum at 160 msec of speech per ear. With

even faster rates, however, performance improves, now reaching perfect performance when the speech is alternated in small enough chunks. The reason for the improvement with faster rates may be due to insufficient time to localize the sounds correctly at fast rates. To the extent the sounds are not localized at the different ears, there is no need to switch between the ears to process the meaning of the message. According to this analysis, the subject does not listen to one ear at the very fast rates, but to the total message at some amorphous location somewhere inside the head.

Recent experiments show that there is something to the idea of switching of attention. Treisman (1971) presented subjects with a list of 6 or 8 digits to the middle of the head, or alternating between the ears. The list was presented at 4 or 6.7 items per second. Subjects recalled significantly fewer digits in the alternating than binaural presentation, especially at the fast rate of presentation. This result was independent of the size of the list. It is unfortunate that Treisman compared the alternating case to the binaural case instead of a monaural case since the digits might be easier to perceive with two ears than one. Even so, it appears that this confounding cannot account for the large alternating and binaural differences. The results seem to indicate, instead, that the time needed to switch between the ears in the alternating case disrupted performance relative to the binaural presentation (Harvey & Treisman, 1973).

B. Counting Sounds Alternating between Channels

Guzy and Axelrod (1972) asked their subjects to count the number of clicks presented in a short sequence. The clicks were either presented to the same ear or alternated between the ears. Counting, in terms of our model, requires the operations of the secondary recognition process to update a counter held in generated abstract memory. If the clicks are perceived as coming from different locations in the alternating case, the secondary recognition process would need additional switching time to monitor the channels in this case relative to the monotic presentation. Guzy and Axelrod found that the subjects tended to underestimate the total number of clicks in the alternating sequences relative to their estimations of single ear presentations. They argued that additional switching time would force the subjects to fall behind the alternating case, missing some of the click presentations.

In terms of our model the poorer performance found when inputs are alternated between the ears relative to being presented to the same location is due to the additional time for the secondary recognition process to switch between the channels in synthesized auditory memory. Given the importance of the finding, I asked whether this phenomenon revealed a true processing deficit (Massaro, 1976a). In the counting experiments, investi-

gators have depended on a point of subjective equality as a measure of performance. Given that this measure could also be influenced by decision variables, it seemed necessary to substantiate these results, utilizing a dependent measure that indicates exactly what the subject knows on each trial. Also, as in other information-processing tasks we asked whether this phenomenon would hold with highly practiced observers who are given feedback throughout the experiment. Subjects heard a sequence of 5, 6, 7, or 8 20-msec tones, reported which of these 4 sequences occurred, and were given feedback about the actual number of tones presented. The subjects were given 600 trials of practice on the first day followed by four experimental days. The tones could be presented at any of 8 rates of presentation ranging between 20 and 3.5 per second. And, of course, the tones were either alternated between the ears or presented to the same ear.

Figure 9 shows plots of the percentage of correct counts as a function of the processing time for each tone. Processing time is the time between the onsets of successive tones. The results show that counting performance improves with increases in processing time but at a significantly faster

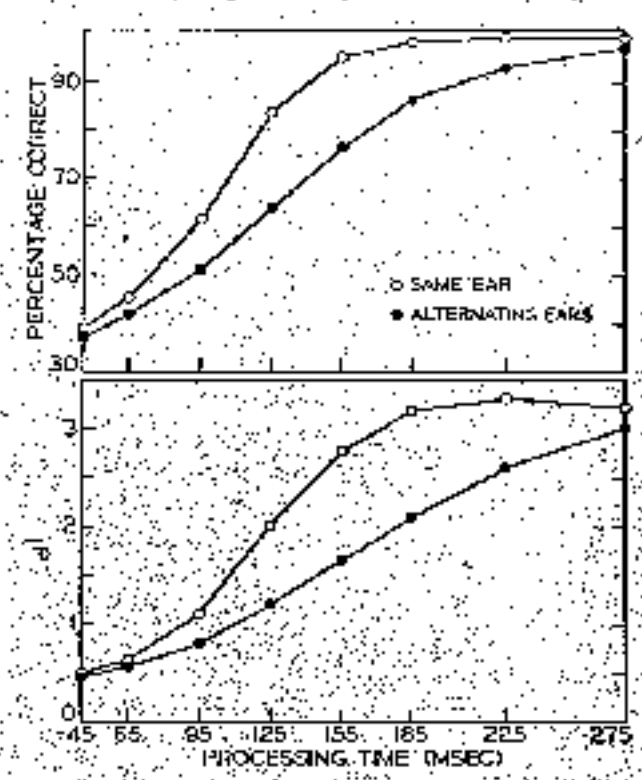

FIGURE 9 Percentage of correct counts and d' values as a function of the processing time for each tone for tones presented to the same ear or alternated between ears. (After Massaro, 1976a.)

rate when the tones are presented to the same ear than when alternated between ears. Given the four-alternative forced-choice task, the percentage of correct counts could be translated into d' values using the tables given by Elliot (1964). The observed d' values are also given in Figure 9. The d' value, like the percentage correct value, simply represents how well the sequences can be counted.

These results substantiate the idea of a switching time between the ears and can be described by a simple quantitative model incorporating the concept of switching time. Assume that counting performance for tones coming in the same ear is given by the equation

$$d'_{same} = \alpha(1 - e^{-\theta t_p}), \tag{2}$$

where α is the performance asymptote with maximal processing time t_p and θ gives the rate of approach to this asymptote. Processing time t_p is defined as the time between the onsets of successive tones. This equation describes performance in the single ear condition and can also be used to describe the alternating case by a simple incorporation of switching time

$$d'_{alternating} = \alpha(1 - e^{-\theta(t_p - t_s)}). \tag{3}$$

In the alternating case, switching time subtracts from processing producing a decrement in counting performance.

A second experiment explored the possibility that the frequency dimension operated similarly to spatial location. It replicated the counting experiment, but the 20-msec tones were presented binaurally, and were either presented at the same frequency or alternated between two different frequencies. The different frequencies were slightly over 1 octave apart; the notes A_4 and B_5 were used. All other experimental conditions were the same as in the counting experiment described earlier. The results showed that accuracy of counting improved with decreases in the rate of presentation of the tones. Furthermore, at rates between 4 and 8 tones per second, subjects showed about a 10% decrement in counting tones alternating between frequencies relative to being presented at the same frequency.

The decrement in counting performance when tones are alternated between the ears or between two frequencies is assumed to occur at the secondary recognition stage of processing. An alternative explanation would locate the counting deficit at the primary recognition stage. That is to say, it is possible that alternating tones are not heard as well as those coming along one location or frequency. However, the backward masking experiments discussed in Section III.C showed no decrement at the primary recognition stage when attention must be divided between the two ears or between two dimensions of the test tone. Secondary recognition, on the other hand, is necessary to abstract meaning from synthesized auditory memory. Given that sounds are perceived at different spatial locations or

different pitches in the alternating conditions, secondary recognition has more difficulty counting these sounds than sounds perceived at the same pitch and spatial location (see Figure 1). These two stages of processing illuminate the differences found in the backward masking and counting experiments and provide a processing model of when selective attention will and will not occur.

VI. AUDITORY AND ABSTRACT MEMORY

In the previous section, it was helpful to distinguish between two levels of processing auditory information. Primary recognition involves perceptual resolution; secondary recognition resolves information at a conceptual level. Primary recognition transforms information into a synthesized auditory form whereas secondary recognition generates an abstract memory representation.

Synthesized auditory and generated abstract memory play important roles in processing speech and can be used to account for a number of phenomena in speech perception and auditory short-term memory. This section begins with an illustration of how auditory and abstract codes can be differentiated in immediate memory tasks. The utilization of these codes in a number of tasks is then reviewed. This analysis appears to illuminate the processes that occur in the categorical perception of consonants, vowels, and nonspeech stimuli. Auditory recency and suffix effects in immediate memory can also be described in terms of the contribution of auditory and abstract codes.

A. Auditory and Abstract Codes

We have postulated different memory structures to describe the storage components available at successive stages of information processing. Auditory memory is needed to remember a person's voice since a conceptual representation is not sufficient for accurate recognition. Consider the number of voices you can recognize over the phone and how you might describe the differences between them. Ordinarily, your description would not be sufficient for another person to categorize the voices correctly. Abstract memory, on the other hand, is symbolic and is not modality specific. Accordingly, we might forget what someone's voice sounds like without forgetting what the person means to us. Auditory and abstract memories are manifest in a number of short-term memory tasks. In a number of experiments, it has been possible to isolate the memory storage responsible for performance at an auditory or abstract level. Massaro (1975a, Chapters 24–27) provides a more detailed description of modality specific and abstract representations in human information-processing tasks.

A few recent studies illuminate the operation of auditory and abstract codes in immediate memory tasks. The results can be used to pinpoint the utilization of an auditory or abstract code in the task. Smith and Burrows (1974) presented subjects with a relevant 4-item list to one ear simultaneously with an irrelevant 4-item list to the other ear. A probe item followed the test list on the same ear and the subject's task was to indicate whether or not the probe item was a member of the test list (Sternberg, 1966). In this task, only the test list and probe item had to be processed and subjects were instructed to ignore the irrelevant list, attending only to the relevant ear. The probe item was equally likely to be a member of the test list or another item. On some of the trials, the probe item was not a member of the test list but had been presented in the irrelevant list. In contrast to the 4% error rate to totally new probe items, subjects responded that these probe items had been in the test list on 20% of the trials. The memory storage and comparison could have occurred at the level of auditory or abstract codes in this task. If the codes were auditory, this meant that the auditory representation did not always maintain the spatial location of the sounds. Similarly, the codes could have been abstract ones with an incorrect tag or association about spatial location of the items.

Smith and Groen (1974) carried out an experiment that tended to locate the memory storage and comparison in this task at the level of abstract codes. They replicated the Smith and Burrows study but included trials in which the items in the test list belonged to a different semantic category (for example, animals) than the items in the irrelevant list (for example, furniture). If a probe item came from the irrelevant ear but belonged to a different category than the test list, subjects correctly categorized this item as a new item at the same rate as in the case of a totally new item. The meaning of the items was utilized to improve performance in the task. Subjects were able to respond "no" to an item from the irrelevant list because it belonged to a different category than the items in the test list. Given that an auditory code alone would not have information about category membership, it is safe to locate the memory storage and comparison in this task at the level of generated abstract memory in our model.

B. Categorical Perception

One well-documented phenomenon is that listeners can more easily discriminate between speech stimuli they identify with different names than those given the same name (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). Consider an experiment carried out by Pisoni (1973).

Pisoni synthesized seven consonant sounds on a continuum between /bæ/ and /dæ/. Figure 2 shows spectrograms of the synthesized sounds /ba/ and /da/. The five sounds between /bæ/ and /dæ/ were synthesized with second and third formant locations that divided the /bæ/–/dæ/ range into seven equal steps. Subjects were first given the sounds one at a time in an absolute identification task with the alternatives /bæ/ and /dæ/. These same subjects were given two of the sounds in a same–different task in which they were instructed to respond same or different with respect to the sound of the syllables. Sometimes the sounds were, in fact, the same. For the data analysis, the trials with two different sounds in the same–different task were classified according to whether or not they had been identified with the same label in the identification task. Pisoni's subjects could discriminate the difference between sounds assigned different labels much better than sounds assigned the same label in identification.

Pisoni's study illustrates the categorical nature of processing stop consonant syllables. Listeners appear to be limited in their ability to discriminate differences between speech sounds that have been assigned the same name. Categorical perception has been a central phenomenon in arguing for the motor theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Speech perception, according to this idea, differs from other perceptual phenomena, because at some level articulation intervenes in the decoding of a speech signal. Therefore, different sounds may be mediated by the same articulatory set of commands leading to the perceptual equivalence of the sounds. Subjects cannot discriminate differences between two sounds assigned the same name because they have perceived the sounds as identical. This means that, in speech, equal stimulus differences do not produce equal perceptual differences.

This strong inference from categorical perception research has come under heavy attack in recent years. Listeners can discriminate differences between speech sounds assigned the same name. Pisoni and Lazarus (1974) trained their listeners to listen to the differences between the successive syllables along the stimulus continuum from /ba/ to /pa/. Some of the subjects were able to discriminate between sounds normally assigned the same label as well as those sounds normally assigned different labels. Barclay (1970) has also shown that subjects can consistently perceive differences between sounds assigned the same label /d/. These two studies make viable an alternative interpretation of categorical perception results: subjects utilize the name they have given a sound which they forget the sound itself. In Pisoni's (1973) study, for example, listeners may have forgotten the sound of the first syllable and responded same or different according to whether the second syllable had the same name as the first. This strategy would have shown good discrimination between sounds

given different names and bad discrimination between sounds given the same name. For further discussion of this view of categorical perception see Fujisaki and Kawashima (1970), Massaro (1975a), and Paap (1975).

Categorical perception or the tendency to more easily discriminate differences in sounds given different labels than sounds given the same label is not limited to speech sounds. Cutting and Rosner (1974) used a Moog synthesizer to generate sawtooth waves that differed in their rise time. Rise time corresponds to the time it takes the wave to reach maximal intensity. The waves reached maximal intensity either immediately or in 10, 20, 30, 40, 50, 60, 70, or 80 msec after the onset of the sound. Sawtooth waves with short rise times sound like the plucking of a stringed instrument whereas those with longer rise times sound like the playing of the same instrument with a bow. Subjects were played the end stimuli and told to interpret the one with an immediate rise time as a plucked string of a musical instrument and the one with the 80 msec rise time as a bowed string similar to a violin. They then identified as *pluck* or *bow* the nine stimuli presented one at a time in a random sequence of trials. Discrimination between the stimuli was tested in the standard *ABX* task in which subjects indicated whether the last of a sequence of three stimuli was equal to the first or the second.

The results of the identification and discrimination tests are presented in Figure 10. The percentage of bow responses increased steadily with increases in the rise time of the sounds. Categorical perception is assumed to occur to the extent that performance in the *ABX* task is limited by performance in the identification task. If subjects assign two sounds to the same category, they should have difficulty discriminating them in the *ABX* task. For example, given that the stimuli with 0 and 20 msec rise times are usually called pluck, they should not be discriminated in the *ABX* task. The stimuli with 30 and 50 msec rise time, however, are usually assigned different labels so that performance with these stimuli in the *ABX* task should be very good. This was the outcome, exactly supporting the idea that the sounds were treated similarly to speech sounds previously shown to exhibit categorical perception. For example, Cutting and Rosner obtained the same results by varying the rise time in speech syllables giving a continuum between *cha* and *sha* as in chop and shop, respectively.

Cutting and Rosner's study appears to be at first glance a plus for any interpretation of categorical perception that rejected the idea that equal stimulus differences did not produce equal perceptual differences. That is to say, categorical perception may occur even though equal physical differences are perceived to be equally different in the context of speech sounds. The phenomenon of categorical perception may, in fact, be categorical memory. Subjects in the *ABX* task may try to remember the sound

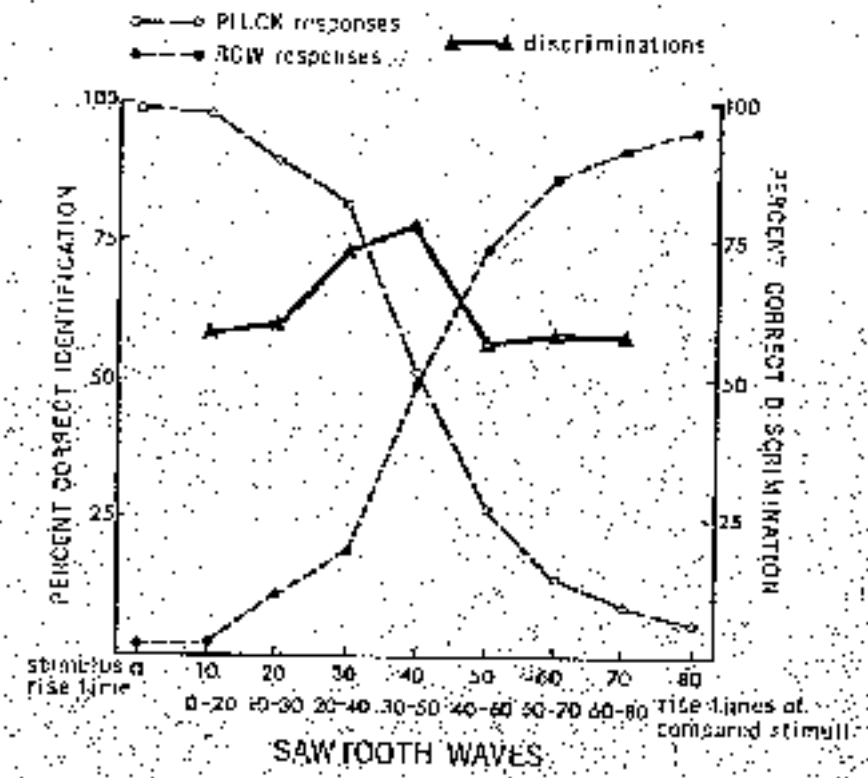

FIGURE 7.10. Percentage of correct identifications and discriminations of pluck and bow sounds. (After Cutting & Rosner, 1974.)

and the label assigned to the *A* and *B* sounds. When *X* is presented, they then try to match the sound of *X* with the remembered sounds of *A* and *B*. Given that sound memory is so fragile in this task, however, they more often than not forget the sounds of *A* and *B* (Massaro, 1975a; Pisoni, 1975). In this case, the subjects rely on the labels they assigned to the sounds A and B and choose the one that matches the label they assigned to X. This strategy will produce the results usually attributed to categorical perception. Sounds assigned different labels will be more likely to be discriminated in the *ABX* task simply because the subject depends on the *name* of the sounds rather than his memory for the sounds themselves. Accordingly, there is nothing unique about the perception of speech sounds when the role of memory is accounted for in the *ABX* task.

In accord with this viewpoint, Cutting and Rosner's subjects also employed the labels assigned to the sounds in the *ABX* task. Having spent a session labeling the sounds, it would only be natural to label them in the second session in the *ABX* task. In this way, musical sounds were treated like speech sounds in the *ABX* task because the musical sounds

were readily assigned labels. Cutting and Rosner, however, reduced the credibility of this interpretation in their second experiment. Rather than presenting the identification task before the *ABX* task, they presented them in reverse order. Subjects now discriminated the sounds in the *ABX* task without being told about the labels *pluck* and *bow* and with no experience in identifying the sounds. The results in the *ABX* task replicated exactly the results in the first experiment. Since subjects could have assigned labels anyhow, the experiment is not definitive. Asking the subjects whether *X* is equal to *A* or *B* may encourage the use of labels. But why did subjects use only two labels? If subjects employed three labels, two peaks instead of one would have been found. Whether or not the peak in performance in the *ABX* task represents the fact that equal stimulus differences do not produce equal perceptual differences remains to be answered.

The one solution to this problem may be direct comparison of *ABX* performance of pluck and bow sounds with steady-state sounds. Steady-state sounds such as vowels have not produced categorical perception in the *ABX* task (Fry, Abramson, Eimas, & Liberman, 1962). Cutting and Rosner should have included a continuum of steady-state sounds such as pure tones differing in frequency but not rise time. The steady-state sounds should be synthesized to give the same level of *ABX* performance on comparisons within each category as that given by the sounds differing in rise time. By equating the discrimination within the steady-state categories with discrimination within the sound categories differing in rise time, the listener would have the same level of auditory information in both tasks. We might expect, then, that the subject would utilize labels to the same degree in both tasks. If the between-category performance is the same for steady-state sounds and sounds differing in rise time, we can conclude that there is nothing unique about the pluck and bow sounds. Rather, the peak in Cutting and Rosner's data would appear to be due to the listener's tendency to utilize the name he assigns to a sound when he forgets what the stimulus actually sounded like. On the other hand, if no peak is observed between categories with steady-state sounds, where one occurs with changes in rise time, one would seem compelled to accept the idea that the equal stimulus differences in rise time do not produce equal perceptual differences.

Looking at the problem from a different perspective, there is no reason to expect that equal stimulus differences, as defined by the experimenter, are equal in terms of the auditory processing system. This conclusion is seen most clearly in a recent report by Stevens and Klatt (1974). Previous studies had shown that equal steps in the delay in voice onset time of synthetic stop consonant syllables do not appear to give equal perceptual differences. Short delays lead to the perception of voiced syllables whereas longer delays give rise to the perception of voiceless sounds. Syllables with

a 20 or 40 msec delay sound very similar and are perceived as voiced while 60 and 80 msec delays give similar percepts of voiceless syllables. Accordingly, the 20 msec difference between 40 and 60 msec appear to transmit more information than the 20 msec differences between 20–40 and 60–80 msec. Stevens and Klatt point out, however, that voice onset time also determines whether there are format transitions at the onset of voicing (see Figure 2). Significant transitions are present at 20 and 40 msec delays but not at 60 and 80 msec delays. By covarying the delay in voice onset time and the presence of transitions at the onset of voicing, Stevens and Klatt showed that the latter was a significant contribution to the perception of voicing. This demonstrates that equal stimulus differences in voice onset time as defined by the experimenter were not equal in terms of the auditory processing system. Notice that this explanation does not require any special perceptual operation such as that assumed by the motor theory of speech perception.

C. Consonants and Vowels

Another issue in speech perception research is whether vowels are perceived differently than consonants. In this section we consider the possibility that the perceptual resolution and the auditory memory of steady-state sounds like vowels are better than they are for sounds like stop consonants whose acoustic characteristics change rapidly over time. Figure 2 shows that vowel sounds are steady state since the formant values do not change during the sound as they do in the stop consonant portion. Employing the procedure outlined previously, Pisoni (1973) asked observers to identify a continuum of vowel sounds between /i/ and /I/ or /i/ or /r/ and then to discriminate the sounds in a same–different task. In contrast to the stop consonants, subjects were able to discriminate between different vowel sounds even though they were assigned the same label.

The differences found between stop consonants and vowels can be attributed to differences in synthesized auditory memory. Given their acoustic properties, the auditory memory for the sound of a vowel might be better than that of a stop consonant. A vowel is steady-state and tone-like, whereas stop consonants are characterized by rapid transitions across the frequency spectrum. In fact, it has recently been shown that the just noticeable difference for frequency transitions comparable to those found in speech is much larger than for smaller transitions or steady-state tones (Tsuzuraya, Sano, & Niimura, 1973).

Given that the sound of steady-state sounds is better resolved than that of stop consonants we would expect differences in the same–different task.

As demonstrated by the model given by Fujisaki and Kawashima (1970), the listener can utilize both perceptual and conceptual codes in the same different task. Subjects have information about the sound of the first sound in synthesized auditory memory and its name in generated abstract memory. If the sound is well remembered, the subject will make his decision on the basis of whether the second stimulus has the same sound. If the sound of the first stimulus is forgotten, however, subjects will tend to respond on the basis of whether the name of the second stimulus is equal to the name of the first. The observed differences between stop consonants and vowels appear to be due to the different degrees of resolution of these sounds in synthesized auditory memory rather than their phonetic class (Darwin & Baddeley, 1974). Paap (1975) discusses other studies of consonant and vowel differences in terms of differences in synthesized auditory memory.

D. Auditory Recency

One popular task in short-term memory research is the serial presentation of a short list of items followed by serial recall of the list beginning with the first item. The typical task involves a presentation rate of one or two items per second with seven or eight items in the list. The results, plotted in terms of probability of correct recall as a function of serial position, usually show that the first item is recalled best, with decreasing recall for each successive item until the last one or two items which are recalled somewhat better than what would be expected from a decreasing function (see Figure 11). Auditory recency refers to the finding that auditory presentation leads to better recall than visual presentation, especially at the end of the list.

Madigan (1971) compared auditory versus visual presentation of a list of eight one-syllable nouns presented at the rate of one per second. When subjects were asked to recall the items in the order in which they were presented, he found better recall for the auditory items, especially those at the end of the list. When subjects were asked to recall the items in backward order, however, there was no difference between the two presentation conditions, even at the end of the list. The recency advantage of auditory over visual presentation may be unique to the serial presentation and serial recall tasks or in free recall tasks where the last few items are recalled in serial order. If synthesized auditory memory is responsible for the auditory recency in serial recall, it may mean that the auditory code is helpful when it can be employed in a forward moving direction but not in a backward direction as in backward recall. Unfortunately, the exact contribution of an auditory presentation cannot be evaluated until the stor-

age, retention, rehearsal, and retrieval processes are accounted for in these recall tasks.

E. Suffix Effect

A related finding that has generated an abundance of experimentation is the suffix effect. Dallett (1965) asked subjects to serially recall a list of digits presented auditorily at a rate of 2.2 items per second. Subjects were given an eight-item list or a seven-item list followed by the predictable and, therefore, redundant digit "zero." The "zero" did not have to be recalled. Figure 11 shows that the seven-item list with the redundant "zero" functioned very much like the eight-item list. Although subjects had to recall one less item in the seven-item list, recall performance was exactly the same as recall of the first seven items of the eight-item list. Crowder (1967) asked subjects to recall a serial list of digits presented auditorily at 2 items/sec. In two conditions, the list contained either eight or nine items, and in a third, the eight-item list was followed by the redundant digit "zero." The results showed that the eight-item list with a redundant suffix functions like the nine-item list, not the eight-item list. These results show that an n-item list with a redundant suffix functions in the same way

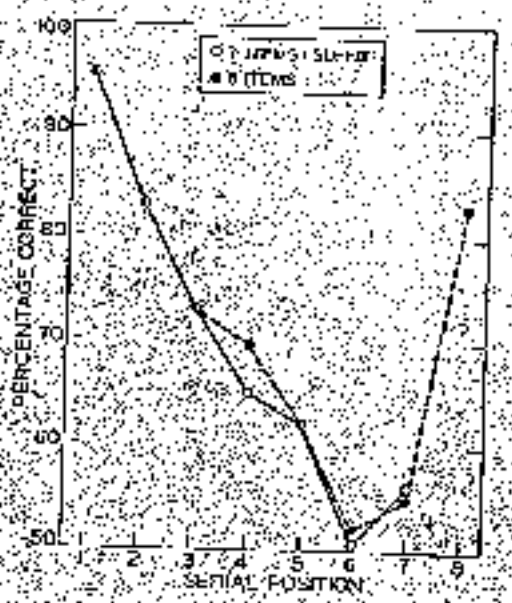

FIGURE 11 Percentage of correctly recalled items as a function of serial position in the list. The parameter refers to recall of an eight-item list presented alone or a seven-item list followed by the redundant suffix "zero." (After Dallett, 1965.)

as an $(n + 1)$-item list. Although the subject knows that the suffix is redundant, it cannot be ignored and is processed along with the rest of the list. This means that the interference produced by the redundant suffix functions in the same way as an additional auditory item to be remembered and must be accounted for in this way.

The suffix effect can be attenuated by reducing the physical similarity between the suffix and the list presentation. Presenting the suffix to the opposite ear as the stimulus list or in a different voice than the list reduces the suffix effect but does not eliminate it entirely. The suffix must be speechlike; white noise or a buzzer does not produce a suffix effect although playing the suffix "zero" in reverse does, in fact, function like a zero suffix. Using the model in Figure 1, the suffix effect seems to occur when the suffix emerges from synthesized auditory memory along the same channel as the test list. By varying some physical property of the suffix, it is perceived at a different location in auditory space or in voice quality so that the semantic analysis given by secondary recognition is reduced. This reduction, of course, will then reduce the amount of interference that the suffix usually has when it is processed as one of the items in the list (Massaro, 1970b, 1972a).

F. Summary

In contrast to the analysis of primary and secondary recognition, the research discussed in this section has failed to isolate exactly the processing stages responsible for many of the observed phenomena. The processing operations and structures that contribute to categorical perception must be defined in future research. It is clear that both auditory and abstract codes play a role and that a decision process may differentially utilize one code or another depending on task demands. In a related problem, the differences observed between consonants and vowels appear to be due to the acoustic properties of these sounds rather than differences in their phonetic class. The lesson to be learned from both of these areas of research is that the observed performance must be dissected in order to discover the structures and processes responsible for the phenomena. The information-processing analysis has been extremely valuable in this effort but additional work is needed before any final evaluation can be made.

One of the major tasks facing short-term memory researchers is the development of a process model for serial recall tasks. Process models have been developed for probe recognition and probe recall tasks, which eliminate the contribution of output interference (Massaro, 1970b, 1975a; Norman, 1966; Wickelgren, 1970). Until such a model can be utilized in recall tasks, it is difficult, if not impossible, to arrive at the mem-

ory structures and processes responsible for the phenomena of auditory recency and suffix effects.

VII. GENERATED ABSTRACT MEMORY

Generated abstract memory in our model corresponds to the primary, immediate, or short-term memory in the prototypical memory model (Atkinson & Shiffrin, 1968; Murdock, 1974; Waugh & Norman, 1965). Rehearsal and recoding operate at the level of abstract memory in our model. Rehearsal processes maintain information in immediate memory whereas recoding allows a reduction of the number of units in memory.

A. Forgetting

One persistent goal has been to determine if forgetting is due to time alone without rehearsal or to active interference. Reitman (1971) and Shiffrin (1973) used a signal detection task to prevent rehearsal during the forgetting interval. For example, Shiffrin's subjects were given five letters to remember and then were required to detect tonal signals for 1, 8, or 40 sec. After the signal detection task, the subjects were required to perform an arithmetic task for 5 or 30 sec. In the arithmetic task subjects saw a three-digit number followed by a single digit to be added every two seconds. The duration of the signal detection task had no significant effect on performance whereas increasing the duration of the addition task lowered performance significantly. Signal detection performance also was not affected by the addition of the memory task. One might argue that rehearsal did not occur during the signal detection task since the subject's detection performance was not disrupted by the memory task. Unfortunately, Reitman (1974) has shown that rehearsal can occur during the signal detection task without affecting signal detection performance. Therefore, it is possible that the subjects were able to rehearse during the signal detection period, preventing any forgetting that might otherwise have occurred. This possibility prevents a critical test between time decay and interference theories using this paradigm.

Recent findings by Reitman (1974), however, reveal than an intervening signal detection task interferes with memory for a list of five words. Although these results might be consonant with a pure decay process, they are more compatible with the view of a limited-capacity memory. The detection task requires detection, memory, decision, and response selection and execution processes during the retention interval. To the extent these processes draw on a limited capacity of resources, less processing capacity

is available for rehearsal and recoding allowing interference to take place. Similarly, Anderson and Craik (1974) showed that a choice reaction time task concurrent with the list presentation interfered with memory for the items on the list. The limited-capacity rule appears to offer a good description of forgetting in generated abstract memory.

Massaro (1970b) formalized a "limited capacity" view of forgetting in short-term memory studies. The central concept in the model is the effect of perceptual processing and rehearsal on memory. The two main assumptions of the model describe changes in memory for an item as a function of perceptual processing and rehearsal. The first assumption is that memory for an item is directly related to the amount of perceptual processing and rehearsal of that item. Accordingly, memory for an item will increase with increases in the time a subject has to process that item. The second assumption is that memory for an item is inversely related to the amount of processing of other items. This "limited capacity" rule provides a reasonable description of the acquisition and forgetting of information in generated abstract memory.

B. Free Recall

Consider a typical experiment in which subjects are given a list of items followed by a free recall of the list. The serial position curve usually shows that the last three or four items in the list are recalled best, the middle items poorest, and the first two or three items intermediate between these two. The advantage of the last and first items is called recency and primacy, respectively. Current experimentalists are interested in whether experimental variables such as rate of presentation, phonological and semantic similarity, and modality affect portions of the serial position curve differently (Murdock, 1962; Watkins, 1972; Watkins, Watkins, & Crowder, 1974). One problem with the free recall experiment, however, is that the experimenter has no control over the rehearsal strategies of the subjects and the order in which the subject reports back the items. For example, it has been demonstrated that reporting items in recall interferes with memory for the other items in the list (Norman & Waugh, 1968; Tulving & Arbuckle, 1963). Therefore, the experimenter cannot evaluate exactly whether one item is recalled better than another because the first was learned better than the second or simply received less output interference than the second.

A number of studies have purported to show that some variable affects the recency and earlier portions of the serial position curve differently (Craik, 1969; Glanzer & Razel, 1974; Murdock & Walker, 1969; Watkins, 1972). Given the problem of rehearsal strategies and order of report in

these experiments, however, infirms these conclusions. Consider the apparent findings of varying the rate of presentation of the test list. Previous studies have shown that rate of presentation has a large effect on the early and middle portions of the serial position curve, but no effect at the end of the list (Glanzer & Cunitz, 1966; Murdock, 1962). This result is contrary to the findings in a probe recognition or probe recall task that eliminate differences due to rehearsal strategies or output interference (Massaro, 1970b; Waugh & Norman, 1965; Wickelgren, 1970). In these latter studies the rate of forgetting of an item as a function of the number of items intervening between its presentation and test is inversely related to the rate of presentation (Massaro, 1970b).

Bernbach (1975) has illuminated the contribution of rehearsal strategies on the recency portion of the serial position curve in free recall. If subjects were given no information about the number of items in the test list, rate of presentation had a large effect on the recency portion of the curve. Bernbach also found a large effect of presentation rate in a continuous paired-associate task that eliminates output interference and differential rehearsal strategies. These results agree with the probe recognition and probe recall studies which control for rehearsal strategies and output interference. However, when Bernbach told the subjects the number of items in the test list in the free recall task, rate of presentation had a substantial effect on recall except at the last four serial positions. Bernbach's explanation of the discrepant results is straightforward. Subjects who know the number of items in the test list can reduce processing of the last few items and concentrate on rehearsing earlier items in the list. The last items will receive little interference from new test items and little output interference since they will be recalled first. Accordingly, when the subject knows the length of the test list, the last items presented at a slow rate receive the same amount of processing as those at a fast rate and no effect of rate is observed. Bernbach's work makes apparent the need to account for the processing stages in free recall tasks before theoretical conclusions can be reached.

VIII. SIMILARITY TO OTHER MODELS

The present model follows in the tradition of Broadbent's (1958) seminal information-processing model. Many, but not all, of the differences would be eliminated by inserting a preperceptual stage between the sources and his short-term store (see Figures 1 and 3). Broadbent (1958, 1971) has, in fact, studied and acknowledged the initial preceptual stage of processing but has not incorporated it into his flow diagram (see Broadbent, 1971, Chapter 4; Broadbent & Ladefoged, 1957). The present model is also

similar to other recent descriptions of auditory processing and memory. Aaronson (1974) postulates two stages of processing in immediate memory tasks. The stimulus input is placed in an initial store, transformed into some auditory representation, and then identified or coded based on some meaning retrieved from long-term memory. The initial storage buffer corresponds to our preperceptual storage; the auditory representation is analogous to our synthesized auditory memory; and the output of identification corresponds to generated abstract memory in our model. Similarly, Neisser's (1967) echoic storage and Crowder and Morton's (1969) precategorical acoustic storage correspond to synthesized auditory memory. Fujisaki and Kawashima (1970) and Pisoni (1975) have postulated auditory and phonetic codes in categorical perception studies. These codes would be held in synthesized auditory and generated abstract memory in our model.

IX. CONCLUDING REMARKS

The line between experiment and theory is at best wavy and sometimes contains gaps and sinkholes. The information-processing approach offers a promising paradigm that tightens the relationship between experiment and theory. The approach may be losing some of the novelty and romance since the beginning of its recent "zeitgeist" and, in fact, some disfavor is apparently emerging (Allport, 1975; Craik & Lockhart, 1972). Even if someone offers something of a paradigm shift, however, we must be content with a gradual accumulation of knowledge about the structures and processes responsible for how man interacts with his environment.

ACKNOWLEDGMENTS

The author gratefully acknowledges the helpful comments of Wendy L. Idson and William K. Estes.

REFERENCES

Aaronson, D. Stimulus factors and listening strategies in auditory memory: A theoretical analysis. *Cognitive Psychology*, 1974, 6, 108–132.

Allport, D. A. Critical notice. The state of cognitive psychology. *Quarterly Journal of Experimental Psychology*, 1975, 27, 141–152.

Anderson, C. M. B., & Craik, F. I. M. The effect of a concurrent task on recall from primary memory. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 307–313.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *Advances in the psy-*

ology of learning and motivation research and theory. Vol. II. New York: Academic Press, 1968.

Averbach, E., & Coriell, A. S. Short-term memory in vision. *Bell System Technical Journal*, 1961, 40, 309–328.

Barclay, J. R. Noncategorical perception of a voiced stop consonant: A replication. *Proceedings of the 78th Annual Convention of the American Psychological Association*, 1970, 9–10.

Bernbach, H. A. Rate of presentation in free recall: A problem for two-stage memory theories. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, 104, 18–22.

Broadbent, D. E. *Perception and communication*. New York: Pergamon Press, 1958.

Broadbent, D. E. *Decision and stress*. London: Academic Press, 1971.

Broadbent, D. E., & Ladefoged, P. On the fusion of sounds reaching different sense organs. *Journal of The Acoustical Society of America*, 1957, 29, 708–710.

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 two ears. *Journal of the Acoustical Society of America*, 1954, 26, 554–559.

Craik, F. I. M. Modality effects in short-term storage. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 658–664.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 671–684.

Crowder, R. G. Prefix effects in immediate memory. *Canadian Journal of Psychology*, 1969, 21, 450–461.

Crowder, R. G. Precategorical acoustic storage for vowels of short and long duration. *Perception and Psychophysics*, 1973, 13, 502–506.

Crowder, R. G., & Morton, J. Precategorical acoustic storage. *Perception & Psychophysics*, 1969, 5, 365–373.

Cutting, J. E., & Rosner, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, 16, 564–570.

Dallett, K. M. "Primary memory": The effects of redundancy upon digit repetition. *Psychonomic Science*, 1965, 3, 237–238.

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

Darwin, C. J., Turvey, M. T., & Crowder, R. G. An auditory analogue of the Sperling partial report procedure: evidence for brief auditory storage. *Cognitive Psychology*, 1972, 3, 255–267.

Dick, A. O. Iconic memory and its relation to perceptual processing and other memory mechanisms. *Perception & Psychophysics*, 1974, 16, 575–596.

Dorman, M., Kewley-Port, D., Brady-Wood, S., & Turvey, M. T. Forward and backward masking of brief vowels. *Haskins Laboratories Status Report on Speech Research*, 1973, 33, 93–100.

Elliot, P. B. Tables of *d'*. In J. A. Swets (Ed.), *Signal detection and recognition by human observers*. New York: Wiley, 1964.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. The identification and discrimination of synthetic vowels. *Language and Speech*, 1962, 5, 171–189.

Fujisaki, H., & Kawashima, T. Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute, Tokyo*, 1970, 29, 207–214.

Glanzer, M. Storage mechanisms in recall. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*. Vol. 5. New York: Academic Press, 1972.

Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 351–360.

Glanzer, M., & Razel, M. The size of the unit in short-term storage. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 114–131.

Guzy, L. T., & Axelrod, S. Interaural attention shifting as response. *Journal of Experimental Psychology*, 1972, 95, 290–294.

Harvey, N., & Treisman, A. M. Switching attention between the ears to monitor lines. *Perception & Psychophysics*, 1973, 14, 51–59.

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

Kinchla, R. A. Selective processes in sensory memory: A probe-comparison procedure. In S. Kornblum (Ed.), *Attention and performance*. Vol. IV. New York: Academic Press, 1973.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431–461.

MacLeod, S. A. Modality and recall order interactions in short-term memory for serial order. *Journal of Experimental Psychology*, 1971, 87, 294–296.

Marslen-Wilson, W. Linguistic structure and speech shadowing at very short latencies. *Nature*, 1973, 244, 522–523.

Massaro, D. W. Consolidation and interference in the perceptual memory system. *Perception & Psychophysics*, 1970, 7, 153–156. (a)

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

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

Massaro, D. W. Effect of masking time duration on preperceptual auditory images. *Journal of Experimental Psychology*, 1973, 87, 146–148.

Massaro, D. W. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 1972, 79, 124–145. (a)

Massaro, D. W. Perceptual and synthetic auditory storage. *Studies in human information processing*. University of Wisconsin (Tech. Rep. 70-1), 1972. (b)

Massaro, D. W. A comparison of forward versus backward recognition masking. *Journal of Experimental Psychology*, 1973, 100, 434–436.

Massaro, D. W. *Experimental psychology and information processing*. Chicago: Rand-McNally, 1975. (a)

Massaro, D. W. *Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press, 1975. (b)

Massaro, D. W. Backward recognition masking. *Journal of the Acoustical Society of America*, 1975, 58, 1059–1065. (c)

Massaro, D. W. Perceiving and claiming forms. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, in press. (a)

Massaro, D. W. Perceptual processing in dichotic listening. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, 2, 331–339. (b)

Massaro, D. W., & Cohen, M. M. Preperceptual auditory storage in speech recognition. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception*. Berlin: Springer-Verlag, 1975.

Massaro, D. W., Cohen, M. M., & Idson, W. L. Recognition masking of auditory lateralization and pitch judgments. *Journal of the Acoustical Society of America*, 1976, 59, 434–441.

Massaro, D. W., & Kahn, B. J. Effects of central processing on auditory recognition. *Journal of Experimental Psychology*, 1973, 97, 51–58.

Mills, A. W. Auditory localization. In J. V. Tobias (Ed.), *Foundations of modern auditory theory*. Vol. 2. New York: Academic Press, 1972.

Moore, J. J., & Massaro, D. W. Attention and processing capacity in auditory recognition. *Journal of Experimental Psychology*, 1973, 99, 49–54.

Moray, N., Bates, A., & Barnett, T. Experiments on the four-eared man. *Journal of the Acoustical Society of America*, 1965, 38, 196–201.

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

Murdock, B. B., Jr. The serial effect of free recall. *Journal of Experimental Psychology*, 1962, 64, 482–488.

Murdock, B. B., Jr., *Human memory: Theory and data*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1974.

Murdock, B. B., Jr., & Walker, K. D. Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 665–676.

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., & Waugh, N. C. Stimulus and response interference in recognition memory experiments. *Journal of Experimental Psychology*, 1968, 78, 551–559.

Pisoni, D. B. Theories of speech perception. In D. W. Massaro (Ed.), *Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics*. New York: Academic Press, 1975.

Pisoni, D. B. Perceptual processing time for consonants and vowels. *Journal of the Acoustical Society of America*, 1999, 53, 1, 369. Also appears in *Haskins Laboratories Status Report on Speech Research*, 1972, SR-31/32, 81–92.

Pisoni, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, 1973, 13, 253–260.

Pisoni, D. B. Dichotic listening and processing phonetic features. In R. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory*. Vol. 1. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975.

Pisoni, D. B., & Lazarus, J. H. Categorical and nonsystematic modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 1974, 55, 328–333.

Pisoni, D. B., & McNabb, S. D. Dichotic interactions of speech sounds and phonetic feature processing. *Brain and Language*, 1974, 1, 351–362.

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

Reitman, J. S. Without surreptitious rehearsal, information in short-term memory decays. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 365–377.

Rosenberg, S., & Lambert, W. E. Contextual constraints and the perception of speech. *Journal of Experimental Psychology*, 1974, 102, 178–180.

Shiffrin, R. M. Information persistence in short-term memory. *Journal of Experimental Psychology*, 1973, 100, 39–49.

Shiffrin, R. M., Pisoni, D. B., & Castañeda-Mendez, K. Is attention shared between the ears? *Cognitive Psychology*, 1974, 6, 150–215.

Smith, M. C., & Burrows, D. Memory scanning: Effect of unattended input. *Journal of Experimental Psychology*, 1974, 102, 722–725.

Smith, M. C., & Green, M. Evidence for semantic analysis of unattended verbal items. *Journal of Experimental Psychology*, 1974, 102, 595–603.

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

Sternberg, S. High-speed scanning in human memory. *Science*, 1966, 153, 652–654.

Stevens, K. N., & Klatt, D. H. Role of formant transitions in the voiced–voiceless distinction for stops. *Journal of the Acoustical Society of America*, 1974, 55, 653–659.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, 77, 234–249.

Treisman, A. M. Verbal responses and contextual constraints in language. *Journal of Verbal Learning and Verbal Behavior*, 1965, 4, 118–128.

Treisman, A. M. Shifting attention between the ears. *Quarterly Journal of Experimental Psychology*, 1971, 23, 157–167.

Treisman, A. M., & Rostron, A. B. Brief auditory storage: A modification of Sperling's paradigm applied to audition. *Acta Psychologica*, 1972, 36, 161–170.

Tsumura, T., Sone, T., & Nimura, T. Auditory detection of frequency transitions. *Journal of the Acoustical Society of America*, 1973, 53, 17–25.

Tulving, E., & Arbuckle, T. Y. Sources of intrusion in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1963, 1, 321–324.

Watkins, M. J. Locus of the modality effect in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 644–648.

Watkins, M. J., Watkins, O. C., & Crowder, R. G. The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 430–447.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review*, 1965, 72, 89–104.

Wickelgren, W. A. Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology*, 1969, 6, 13–61.

Wickelgren, W. A. Time, interference, and rate of presentation in short-term recognition memory for items. *Journal of Mathematical Psychology*, 1970, 7, 219–235.

Wingfield, A., & Wheeler, J. L. Word rate and intelligibility of alternated speech. *Perception & Psychophysics*, 1975, 18, 317–320.

Wolf, C. D. G. An analysis of speech processing: Some implications from studies of recognition masking. Unpublished doctoral dissertation, Brown University, 1974.