

## On Short and Long Auditory Stores

Nelson Cowan  
University of Massachusetts—Amherst

This article examines research that indexes auditory sensory memory. The literature suggests that two separate forms of auditory sensory memory exist: (a) a "short auditory store," which extends the apparent duration of a stimulus up to about 300 ms and is used in stimulus recognition, and (b) a "long auditory store," which retains auditory information of a sound or sound sequence for at least several seconds. Theoretical descriptions of these two forms of storage, differences between them, and research strategies for determining their properties are explored. Finally, the implications of these two stores for models of information processing are discussed.

Auditory (or "echoic") storage refers to the retention of acoustic properties of sounds (e.g., Neisser, 1967). Because many sounds are brief or change rapidly over time, auditory storage probably is required to integrate them into unitary percepts. Moreover, there may be limitations on perception or recall due to the decay of auditory storage. An understanding of perceptual processes, therefore, would be improved by knowledge about the temporal properties of auditory storage. However, estimates of its temporal course have varied widely (cf. Crowder, 1976; Klatzky, 1980; Massaro, 1972). In the present article, it is proposed that there is evidence for two different types of auditory sensory memory: (a) an unanalyzed auditory trace lasting up to about 200 to 300 ms, and (b) a separate, longer memory of auditory information lasting at least several seconds. Although many previous commentaries have not explored this possibility, two-storage models were suggested by Aaronson (1974), Massaro (1975), and Kallman and Massaro (1979, 1983). The present article discusses evidence for such a model in many experimental situations and examines theoretical issues raised by this

approach. Issues include the properties of the shorter versus longer form of auditory storage, how storage might operate for complex stimuli such as connected speech, and optimal strategies for research on auditory memory. Finally, implications of the data for storage versus process models of memory are discussed. It is concluded that the brief trace possibly can be characterized as a "process" rather than a store, but that the longer form of auditory sensory memory seems less suitably characterized as a process. For clarity, both memory components are referred to as stores until the storage versus process issue is addressed in the final section.

### One Versus Two Stores: Introduction to a Topic of Debate

A guiding principle in theory construction is parsimony. Applied to memory, it requires the postulation of as few components as possible. Early information processing theories (e.g., Atkinson & Shiffrin, 1968; Neisser, 1967) included only three levels of memory: a literal sensory memory, a short-term categorical memory, and a theme-oriented long-term memory. However, nature may not equal this degree of parsimony. Taken together, auditory memory research findings suggest that there are two sensory stores with very different properties. The issue of two stores versus one store received attention largely on the basis of two paradigms. In one paradigm, Massaro (1972) found that when two brief sounds are presented in rapid succession, the second (masking) sound can prevent recog-

---

The author thanks Priscilla Cowan, Doris Aaronson, Murray Glanzer, David Balota, Rachel Clifton, Marsha Clarkson, Alexander Pollatsek, Dan Anderson, Marvin Daehler, Arnold Well, Keith Rayner, Robert Crowder, and Dominic Massaro for their helpful comments. The author also thanks Dr. Crowder for making his data available.

Requests for reprints should be sent to Nelson Cowan, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

inition of the first (target) sound. This "backward masking" effect is strongest at short delays and decreases to an asymptotic level when the stimulus onsets differ by about 250 ms. It was initially proposed that the minimal onset difference allowing a release from masking indicates the time at which auditory storage has decayed, and that a later mask is ineffective because no auditory storage remains to be masked. However, some investigators suggested that although 250 ms may be the duration of the process of information extraction from the auditory trace, it need not indicate the decay time of the trace. After 250 ms, the mask could be ineffective because auditory storage is no longer needed, although it might still exist.

Another paradigm (Crowder & Morton, 1969; Dallet, 1965; Morton, Crowder, & Prussin, 1971) was used to examine the effect of a suffix item on the recall of an auditory list (i.e., the "suffix effect"). In this approach, a nonrecalled item (e.g., the word *go*) was appended to a list of words to be recalled. The suffix interfered with recall when it followed the list by several seconds or less. Because the suffix was more effective when it was acoustically similar to the list items (Morton et al., 1971), it was said to interfere with auditory memory. Therefore, auditory memory seemed to last several seconds. Critics of this paradigm maintained that the memory being interfered with by the suffix was in some sense processed or categorical rather than sensory. There was no clear consensus on the duration or the number of sensory stores. Estimates of the duration of sensory storage from a variety of paradigms ranged from less than 1 s to over 10 s (Crowder, 1976; Klatzky, 1980).

#### Evidence for Two Auditory Stores

Despite the controversy about auditory storage, previous reviews (Crowder, 1976; Massaro, 1975) contain much of the evidence that there are two types of auditory sensory store. There are studies that support the finding of a rapidly decaying trace and those that support a sensory memory lasting at least several seconds. However, some studies on a brief store might be most familiar to investigators interested in psychoacoustics,

whereas some studies on the longer auditory store might be more familiar to those interested in memory. Although there have been discussions of the two-store hypothesis mentioning both types of evidence (Kallman & Massaro, 1983; Massaro, 1975), a more detailed examination of findings about the two stores is needed.

First, a reexamination of the evidence for two stores is required. Could all of the evidence actually index a single type of auditory storage? Some of the studies to be examined demonstrate that this is unlikely. Some perceptual properties (e.g., audible sensation) are not retained after about the first 300 ms following the stimulus. Other perceptual properties (e.g., the ability to distinguish a sound from other, similar sounds) seem to last much longer and do not even seem possible until part of a second has elapsed since the stimulus onset. Thus, there is an early storage of acoustic information that may mediate the initial recognition of stimulus properties, here termed "short auditory storage," and a later storage of auditory information that makes possible further stimulus comparison and retention, here termed "long auditory storage." The term "long" is used only in comparison with "short," and is to be distinguished from permanent or long-term storage. Some auditory qualities can be saved in long-term memory, and this is discussed later.

Studies of short and long auditory storage are discussed in turn. Some of the studies provide strong evidence for the particular type of auditory memory being discussed, whereas others are less conclusive but can be interpreted well with two stores. Several recent articles are especially relevant to the two-store hypothesis. Kallman and Massaro (1979) examined pitch comparisons in two types of three-stimulus sequence: (a) standard tone, backward mask, comparison tone, and (b) standard tone, comparison tone, backward mask. The mask could disrupt initial encoding of the previous stimulus (i.e., readout from its brief, literal trace) in either type of sequence, but only in the mask-second condition could it interfere with the longer storage needed for a comparison of tones. The results showed that the two types of auditory sensory store have different interference properties.

Target-mask similarity effects occurred only within the longer type of storage. In another study (Kallman & Massaro, 1983), multiple masking sounds were used in a recognition task. They did not result in a partial release from masking, as multiple suffixes do in list recall tasks (Crowder, 1978, 1982a). Therefore, two storage mechanisms with different properties seem to be involved. These two storage mechanisms, presumably short and long auditory storage, are now discussed in greater depth.

#### *Short Auditory Storage*

Short auditory storage refers to a literal store that decays within a fraction of a second following a stimulus. There are several types of evidence to be examined that index this short auditory store. The most direct of these is the measurement of the persistence of auditory sensation. A second type of evidence is the set of auditory tasks in which information in a stimulus is integrated across time (e.g., detection, loudness judgment, or pitch comparison). These paradigms indicate that information can be integrated only within a time window about the duration of short auditory storage, and they imply that such a form of storage should in fact exist. A third type of evidence is the backward and forward masking of auditory detection. The exact estimates of the trace duration differ among these paradigms, due at least partly to different task demands (e.g., whether the subject can directly compare two or more stimuli or must judge isolated stimuli). Nevertheless, the different estimates fall within a reasonably narrow range.

*Auditory persistence paradigms.* Several paradigms have been used to obtain direct estimates of the persistence of auditory sensation. Efron (1970a, 1970b, 1970c) presented subjects with a target sound (tone or noise) followed after a variable period by a probe stimulus (in one experiment, a light flash, and in another, a different sound presented in the other ear). The subject judged the point of perceived simultaneity between the target offset and the probe onset. It was found that the point of target offset was overestimated for brief targets by an amount inversely related to the target duration. Any target

shorter than 130 to 180 ms was incorrectly perceived to be as long as sounds that actually were 130 to 180 ms long. (The 130-ms value was obtained when the probe was a light, and the 180-ms value when the probe was a contralaterally presented sound.) Stimuli longer than 130 to 180 ms were overestimated only by a few milliseconds. Thus, 130 to 180 ms may be taken as a rough estimate of the minimal duration of an auditory percept, which consists of the stimulus period plus the period of sensory persistence. This sensory persistence may be due to a constant-duration neural process originating at the stimulus onset (but see Efron, 1970b, for an examination of how perceptual onset lags might modify the estimates of sensory persistence). Allan (1975) proposed a different model of simultaneity judgment based on attention switching between one stimulus and another, but attention switching cannot explain the stimulus duration effects Efron obtained.

Another measure of sensory persistence (not an entirely satisfactory one) was devised by Békésy (1933, discussed by S. S. Stevens & Davis, 1938). He gradually increased the suddenness of the offset ramp of an 800-Hz tone until the subject could hear no further change in the stimulus. Regardless of the original tone intensity, the point at which no further change in the stimulus could be heard was when the offset ramp was shortened to 140 ms. It was apparently assumed that there could be no perceptible difference between tones that physically decay more rapidly than sensation decays. However, such an assumption is unreasonable if one considers that the ramp offset could summate with the residue or auditory storage from prior stimulation. Miller (1948) used noise stimuli in a similar task and obtained a minimal detectable decay duration of 60 to 90 ms, but the stimulus durations were not given in this or in Békésy's experiment.

Miller (1948) conducted another experiment that may have indexed the persistence of audibility. Subjects listened to continuous white noise interrupted at regular intervals by fixed durations of silence or by fixed durations of noise that was attenuated rather than switched off completely. Subjects were to adjust the attenuation level until they could just detect the difference between silent

versus attenuated intervals. The duration of interruptions varied. It was found that the lowest detectable level of sound during attenuated intervals decreased with increases in the duration of the intervals, and an asymptotic level was approached when the interruptions were 300 to 350 ms long. An account of these findings based on auditory storage is that the sound before an interruption leaves an audible trace. Subjects would not be able to hear an attenuated sound until the trace of prior sound decayed to a level low enough that it would not produce masking. Thus, in this paradigm the trace does not seem to decay entirely before 300 to 350 ms. However, performance in this paradigm may also be affected by the time needed to integrate sound within the attenuated period or to distinguish a silent gap from a local minimum in the fluctuating noise signal.

A variation of this paradigm was used in a classic experiment on sensory persistence by Plomp (1964). Subjects received pairs of noise bursts varying in intensity, but always well above the detection threshold when presented singly. The first burst always lasted 200 ms and the pair always filled a 400-ms window, but a variable silent gap was included by manipulating the start time and duration of the second burst. It was found that when the first burst was sufficiently more intense than the second, and the silent gap sufficiently brief, subjects failed to detect the gap. Presumably, on these trials the second burst was mistaken for a continuation of the sensory persistence of the first burst. In the limit (i.e., when the first burst was maximally intense relative to the second burst), the just-noticeable gap approached 200 ms, which estimates the duration of audible trace persistence.

Penner (1977) replicated this finding and also ran a condition in which the two noise bursts were presented dichotically. In this condition, it was found that the just-noticeable gap was always about 30 ms. The latter finding does not provide evidence for or against a short storage, because many different complex processes could be involved. Perceptual possibilities exist that would not be possible with binaural stimuli (e.g., the perception of a single sound traveling from one ear to the other).

In sum, two types of research on audible persistence have been discussed. In one type (Efron, 1970a, 1970c), the subject must judge the perceived offset of a sound. This research has resulted in estimates of persistence of 130 to 180 ms. In the second type of research on audible persistence (e.g., Miller, 1948; Plomp, 1964), the subject must judge a perceptible trace of the stimulus such as its offset decay rate or continuity across a gap. From a number of studies of this sort, the estimates of persistence ranged from 60 ms to 350 ms. Assuming that task demands account for these differences, the maximum estimate may be the most important, because it could only occur in the paradigm sensitive to the smallest auditory trace or residue. Nevertheless, some of the discrepancies between paradigms should motivate further research. Efron's (1970a, 1970c) results suggest that sensory persistence is constant if measured from the stimulus onset, and that sounds longer than 130 to 180 ms produce little, if any, persistence. In contrast, Plomp (1964) and Penner (1977) found that a 200-ms noise burst produces sensory persistence for 200 ms past the offset. It would be useful to repeat the Plomp-Penner experiment using a range of stimulus durations (e.g., 20 to 200 ms) to determine if the estimate of sensory persistence following any stimulus can be derived from Efron's estimate plus a correction factor due to differences in task demands. One way that the tasks might differ is in the subjects' criteria for sensory persistence. At a certain intermediate stage of decay, an auditory trace might be intense enough to mask a gap (in Plomp's paradigm) and yet sufficiently decreased from its original level that a stimulus offset is registered (in Efron's paradigm). Later in the article, these persistence paradigms are reconsidered briefly with respect to a quantitative model.

*Integration phenomena.* It is well known that the perceived loudness of a sound increases as the sound's duration increases. Similarly, the detection threshold of a sound decreases as the duration increases. In sounds above threshold, increasing duration causes improved accuracy of loudness or pitch comparisons. This temporal dependence of perception most often has been conceptualized

as the integration of incoming information across time (e.g., Munson, 1947; Zwillocki, 1960, 1969). Notice, moreover, that the integration of information between any two times, *A* and *B*, requires the retention of information from time *A* at least until time *B*. An important aspect of these integration phenomena is that they are limited to a certain time window. If one adopts the view that temporal integration requires retention of incoming information, then the time window should indicate the limits of retention. The limits of auditory temporal integration seem to be consistent with short auditory storage.

One source of evidence about integration is in detection studies, although factors other than integration are involved in some detection tasks. Many studies indicate that the time/intensity trade-off for detection of a tone or noise is limited to stimulus durations of about 200 to 300 ms (Békésy, 1933; Garner & Miller, 1947; Green, Birdsall, & Tanner, 1957; Miskolczy-Fodor, 1959; J. C. Stevens & Hall, 1966). Zwillocki (1960) obtained a similar limit of about 330 ms for the contribution of stimulus duration to the detectability of rapidly repeating sound pulses. However, other studies suggest that the time/intensity trade-off can last up to 2,000 ms (Hamilton, 1957; Miller, 1948; Small, Brandt, & Cox, 1962). The difference between these sets of studies seems to be in the response measure. If the subject is allowed an extended interval in which to listen for the sound, then a positive response can result from detection during only one small part of the interval. Miller (1948) and Small et al. (1962) used a method of adjustment that did provide an extended period of observation, whereas studies in which a brief period of trade-off was obtained used discrete intervals of observation.

The role of perceptual integration versus nonperceptual factors in detection seems clear in the work of Green et al. (1957). Subjects listened for tones embedded in noise in a signal-detection task. In one experiment, tone duration was varied while either the signal intensity or the signal energy was held constant. With constant intensity, detectability increased with increasing tone duration up

to 2,000 ms. However, because constant intensity is achieved by increasing the total stimulus energy along with the duration, this condition is inappropriate for an examination of stimulus integration. In the condition with stimulus energy held constant, detectability was maximum with stimuli between 20 and 200 ms long and declined sharply with further increases in stimulus duration. Thus, with sounds longer than 200 ms, subjects were unable to integrate the stimulus energy completely.

It has also been found that both backward and forward masking of detection increase as the duration of the mask is increased, but only until its beginning or end is about 200 ms away from the target (Weber & Green, 1978). Presumably, this limit occurs because the stimulus energies of the target and mask are integrated only within a 200-ms interval.

When subjects are asked to compare the pitch of two tones in a same-different task, the sensitivity to a frequency difference increases to an asymptotic level as the tonal duration increases to about 200 to 300 ms (Chih-an & Chistovich, 1960/1979; Doughty & Garner, 1948; Turnbull, 1944). Similarly, an asymptote at about 250 ms is found in intensity discrimination for a continually presented standard tone with transient increases in intensity that the subject must try to detect (Garner & Miller, 1944; Harris, 1963, p. 37). Studies using discrete standard and comparison tones similarly suggest that there is a point of asymptote for intensity discrimination that is reached when the tones are lengthened to 200 to 500 ms (Garner, 1947, 1949), although the point of asymptote cannot be determined more precisely from these particular studies. Thus, the studies of frequency discrimination are consistent with the view that integration is based on short auditory storage.

A final type of perception that involves stimulus integration is the judgment of loudness. Scharf (1978) reviewed a large number of studies and concluded that, in general, a sound reaches its maximum loudness when it is at least 150 to 250 ms long. However, loudness judgment is a sophisticated task that depends upon multiple factors. When the task is made simple by requiring direct mag-

nitide estimates of individual sounds, an asymptotic level of loudness is obtained with sounds 150 to 200 ms long for all stimulus intensity levels (J. C. Stevens & Hall, 1966). On the other hand, in tasks requiring loudness matching, the stimulus duration needed for asymptotic judgment levels is intensity-dependent (e.g., Miller, 1948; Small et al. 1962).

In summary, a number of perceptual attributes of sound stimuli are dependent upon the integration of auditory information within a window of roughly 200 to 300 ms. Although nonperceptual factors also can influence these tasks, some studies appear to eliminate nonperceptual factors (especially the signal detection results of Green et al., 1957). These studies suggest that information from the beginning of each sound is retained for several hundred milliseconds, allowing its beginning to be integrated with later portions of the sound.

*Detection masking.* One possible consequence of auditory persistence is the masking of detection of one sound by a prior or subsequent sound. For this effect to occur, the mask must be longer or more intense than the target. Both backward and forward detection masking effects have been obtained with intervals up to about 200 ms (Deatherage & Evans, 1969; Elliott, 1971; Wright, 1964; Zwislocki, 1972; Zwislocki, Pirodda, & Rubin, 1959). This maximum masking period of roughly 200 ms has been observed with ipsilateral, binaural, and contralateral presentation of target and mask, although the masking effect is smaller with contralateral presentation. Both forward and backward masking may be consequences of the auditory persistence of the first stimulus in a pair. In forward masking, the audibility of the mask persists through the target presentation and consequently may drown out the target. In backward masking, the mask may interfere with the persistence that results from a prior target. Auditory persistence of the target probably contributes to its detection, so interference with this persistence hinders detection.

Finally, a number of other phenomena may result from short auditory storage, although this is not certain. For example, when one counts tones presented in rapid sequence, they are underestimated with presentation

rates faster than about 250 ms per tone (Garner, 1951). With more rapid rates, subjects may not have sufficient exposure to the auditory trace of each tone to enable complete encoding, and incompletely encoded tones may not be counted. Another study (Massaro, 1976a) suggests that tone counting depends upon encoding and also on attention to the sounds. Both counting and perception of simple stimulus properties (frequency and duration) were affected by the processing time per tone, but counting was additionally affected by alternation of tone frequency or spatial location. Another paradigm is a "syllabic reassembly" task in which the phonemic components of a syllable are pronounced separately (so far as is possible) and presented in sequence with silent intervals between phonemes. In general, subjects are best able to identify syllables when the silent gaps between components are each 200 ms or shorter (Beasley & Beasley, 1973; Shriner & Daniloff, 1970). This suggests that for adequate identification, subjects may need to compare each component with the auditory trace of the previous component.

To summarize, three kinds of paradigms provide evidence of a short auditory store: studies of auditory sensory persistence, studies of temporal integration, and studies of detection masking. Other data exist that seem easily explained in terms of a short auditory store. In a later section of the article, possible theoretical applications of this short store in a number of complex situations are explored.

#### *Long Auditory Storage*

It seems unlikely that the brief form of auditory storage discussed above could suffice for all types of auditory perception. For example, one often may be exposed to sequences of stimuli that cannot be identified immediately (e.g., words within a sentence spoken in a noisy environment). Optimal comprehension in this situation probably requires the retention of auditory information from multiple unidentified speech sounds. This auditory memory would permit later recognition with the help of contextual cues. It is largely because of this sort of phenomenon that one may initially suspect that there is a

form of auditory storage lasting several seconds. It is referred to here as long auditory storage.

*Modality and suffix effects.* The two phenomena that have been used most often to study auditory memory occur within list recall. The first is the auditory modality superiority effect (Burrows, 1972; Conrad & Hull, 1968): There is an advantage for the recall of items within auditory as opposed to visual lists, confined to the most recent list items. The second is the suffix effect (Crowder & Morton, 1969; Dallet, 1965; Morton et al., 1971): A not-to-be-recalled, list-final item (i.e., suffix) impairs recall of the recency portion if the list and suffix are presented auditorily. (Related to this is the "prefix effect." The recall of recent items within an auditorily presented list is impaired if the subject is required to pronounce a word before recalling the list. However, a suffix selectively impairs performance at the last one or two serial positions, whereas a spoken prefix is not selective in this way.) Crowder and Morton (1969) suggested that both modality and suffix effects indicate the presence of "precategory acoustic storage," although more recent work presents a modified view (e.g., Crowder, 1978, 1983). When one is examining auditory memory, modality and suffix effects present a formidable problem. The findings do not yield many solid conclusions about auditory memory, but there are important leads that should not be neglected. Several tentative conclusions emerge from the present examination of this complex literature: (a) auditory sensory memory is not the only factor involved in modality and suffix effects; (b) in comparison with most factors in modality and suffix effects, auditory sensory memory is more selectively found at the final list position; (c) even at the final list position, it is not yet possible to distinguish auditory memory from the memory for other features, such as the articulatory movements that occur in speech; and (d) either auditory or articulatory memory lasts 10 to 20 seconds.

Morton et al. (1971) conducted 17 experiments to examine suffix effects with various stimulus and task factors. One general finding was that the disruptive effect of the suffix depended upon the acoustic similarity of the

list and suffix (defined in terms of pitch, voice quality, and spatial location) but not upon the semantic similarity. This suggested that the suffix disrupted auditory storage. Crowder (1971) found that the modality and suffix effects occurred with speech items whose vowels differed, but not with items differing only in the initial consonant. Subsequently, Darwin and Baddeley (1974) further clarified the basis of this finding. If the items to be remembered are too similar acoustically (no matter whether they are consonants or vowels within a narrow phonetic range), then auditory memory sufficient for the task will decay rapidly, so that suffix and modality effects may not occur.

There are multiple mechanisms within the suffix effect, however. One indication of this involves serial position effects observed in different conditions. Baddeley and Hull (1979) confirmed the finding of Morton et al. (1971) that prefixes, unlike suffixes, do not selectively impair performance at the end of the list. Further, they found that long suffixes (e.g., "Yugoslavia") were more detrimental than short suffixes (e.g., "Spain") for medial list positions, but short suffixes were more detrimental for the final list position. This result obtained no matter whether the words in the list equaled the short or the long suffix in length. However, the explanation of this effect is unclear; some of the more obvious possibilities were systematically eliminated. The findings do suggest that list-final and list-medial suffix effects may have different causes.

Another study (Balota & Engle, 1981) suggested that auditory sensory memory can be clearly indexed by the suffix effect only in the last list position. The suffix effect was modulated by varying the presentation rate or amount of practice, which are presumed to influence subjects' strategies, but the suffix effect remained unchanged at the final serial position. Thus, it seems likely that masking of auditory memory is a mechanism of suffix effects that may be unambiguously observed only in the final serial position. Accordingly, other experiments (Greenberg & Engle, 1983; Morton et al., 1971) indicate that varying the physical properties of the suffix results in a modification of the suffix effect localized at the last 1 or 2 serial positions.

In contrast, a number of studies have examined aspects of the suffix effect other than auditory memory, and they have resulted in effects that were not localized at the last serial position. For example, requiring the subject to attend to the suffix increases the amount of interference within the list (Greenberg & Engle, 1983; Morton et al., 1971; Nairne & Crowder, 1982). Subjects' classification of the suffix also makes a difference. Ayres, Jonides, Reitman, Egan, and Howard (1979) found that an ambiguous "wa" sound that could be heard as speech or music acted as a more effective suffix to a word list when it was perceived to be speech, and the effect was not localized.

The modality effect also is complex and not easily interpreted as due solely to auditory storage rather than attention and encoding (see Penney, 1975, for a review). The role of auditory memory has been demonstrated by Hitch (1975), who found that training subjects to ignore the suffix (by interleaving it throughout the list) eliminated the suffix effect in visual lists, but not in auditory lists. However, factors other than auditory memory also have been demonstrated. Engle (1974) found that a suffix did not eliminate the modality effect. O. C. Watkins and M. J. Watkins (1980a) found that lists with alternating visual and auditory items were no easier to recall than single-modality lists. They suggested that substantial intermodal as well as intramodal interference between items may occur when subjects are required to attend to each item, as they were in this experiment. Finally, Gardiner and Gregg (1979) found a modality effect even when each list item was preceded and followed by 18 s of counting aloud, which presumably eliminated auditory sensory storage.

Recent literature has further suggested that auditory sensory memory may be confounded with some kind of articulatory memory or memory for sequence within the modality and suffix paradigms. A silently articulated suffix interferes with recall of a spoken list (Nairne & Crowder, 1982; Spoehr & Corin, 1978). The recency effect for a spoken list (but not for a written list) is greatly decreased by subsequent lip-read interference (Gardiner, Gathercole, & Gregg, 1983). The typical pattern of modality and suffix effects is produced

when, in place of auditory presentation, the list is lip-read and the suffix either lip-read or auditorily presented (Campbell & Dodd, 1980) or the list is silently mouthed by the subject and the suffix is either silently mouthed or spoken aloud (Nairne & Walters, 1983). The findings in this area led Crowder (1983) to suggest that the visual and articulatory counterparts to speech may be included within precategorical "acoustic" storage. Modality and suffix effects also have been obtained with deaf subjects, using written or signed instead of spoken lists and suffixes (Shand, in press; Shand & Klima, 1981), so it is not clear exactly what stimulus features are necessary to produce these effects. It is possible that any sequential (rather than instantaneous) stimuli will produce modality and suffix effects, or alternatively, that the stimuli must come from the subject's primary linguistic system. In any case, many of the nonauditory effects described above appear to be localized at the final one or two list positions, so the hypothesis that the final position in list recall indexes only auditory storage may have to be modified.

The demonstration of articulatory coding places in question the initial assumption that the auditory suffix must work by erasing an auditory store. This assumption has been challenged in another way by Crowder (1978, 1982a) on the basis of experiments in which multiple suffixes were used. These sometimes resulted in a disinhibition of recall, so the best explanation of suffix effects appears to be inhibition rather than erasure of storage. The assumption that the suffix effect ordinarily can index memory decay also has been challenged. O. C. Watkins and M. J. Watkins (1980b) used a silent copying task to prevent rehearsal after visual or auditory lists, and they found that the modality effect was still present 20 s after the list terminated. Further, using this distractor task, auditory suffix effects were obtained up to 20 s after the auditory list (M. J. Watkins & Todres, 1980). The interpretation was that ordinarily, a suffix-resistant trace of the last sound forms within several seconds of its presentation. In the presence of the distractor task, a suffix-resistant trace cannot form, and a modality-specific, vulnerable memory seems to last up to 20 s. An important problem for future



research will be to determine whether the 20-s trace found with a distractor task, and the disinhibition of masking found with multiple suffixes, are based upon auditory or articulatory storage. As yet, the vast literature on modality and suffix effects does not solve the question of the duration of auditory storage. It suggests instead caution in interpretation.

*Dichotic listening.* A classic paradigm for the study of auditory memory involves the recall of information on an unattended channel in dichotic listening. For example, A. M. Treisman (1964) presented subjects with the identical speech message in both ears, but with a channel to be shadowed leading or lagging behind the message on the other, unattended channel. Of relevance was the condition in which the shadowed channel lagged behind the unattended channel. Subjects noticed that the messages were identical only if the unattended message preceded the shadowed message by 1.5 s or less. Presumably, the unattended material went into an auditory store that could be compared with the shadowed message, but not after the store became too degraded. Glucksberg and Cowen (1970) presented evidence that information in the unattended channel during dichotic listening is (as Treisman assumed) in a pre-categorical form. Subjects were to shadow one channel, but following a cue were to recall digits embedded in discourse on the other, unattended channel. They were permitted to record check marks if they remembered that a digit had occurred but could not specify the digit. The cued recall rate was better at short cue delays (<5.3 s), and whenever subjects reported that a digit had occurred, they also correctly specified the identity of that digit. This (and additional supporting findings) suggested that subjects could not identify digits "on line" as they occurred, which would lead to a postcategorical form of storage susceptible to partial forgetting. Instead, a pre-categorical, auditory representation presumably was saved for several seconds and allowed delayed identification of material in the unattended channel.

Dichotic listening studies are relevant to issues raised by the literature on modality and suffix effects. When the subject is shadowing, articulatory coding of the unattended

channel is prevented. Further, because the subject is wearing headphones, auditory interference is reduced. (One's own voice is heard through bone conduction, but it is spectrally different from airborne speech.) How long can auditory information be retained in these circumstances if the unattended channel is silent following the target items? Norman (1969) probed recall of six items on the nonshadowed channel either immediately or 20 s after they were presented. There was substantial recall in the immediate probe condition (with a large recency effect), but only chance performance after a 20-s delay. However, one could argue that there might have been some auditory interference, either across channels or vocally produced. In two other studies, the possibility of vocally produced auditory interference was eliminated. Murray and Hitchcock (1969) presented dichotic digit sequences, and required (a) silent mouthing of one channel, (b) that the subject just imagine saying the words on one channel, or (c) silent mouthing of the word *the* to prevent any coding (control condition). Following the list, recall of the attended or unattended channel was required. Recall of uncoded messages was always poorer than was recall of coded messages, and was at about the same level regardless of the type of coding of the other channel, even in a condition in which the subjects knew in advance which ear would be cued for recall. Of particular interest, performance on lists in the control condition was at about the same level as uncoded messages in other conditions, even though there were two simultaneous uncoded messages in the control condition. Inasmuch as coding one channel (in the experimental conditions) did not detract from recall of an uncoded channel, this result confirms that the uncoded messages were held in a form that did not depend upon attentive processing. Finally, Bryden (1971) modified this paradigm in an important way. Subjects received dichotic lists with four digits on each channel and were instructed to say to themselves the numbers presented in one ear. They were then required to report items from both ears, but the required order of report of the two channels was varied. Surprisingly, the unattended ma-

terial was recalled almost equally well when it was reported first rather than second. Apparently, auditory memory did not decay substantially in the time it took to report the four digits from the attended channel. There was no estimate of this duration, but because the subjects usually reported three or four items in the interval, a reasonable estimate is 2 to 4 s. Thus, to summarize, the dichotic listening literature suggests that if the target items are followed by silence on the unattended channel, the auditory trace will last at least 2 to 4 s but not 20 s.

*Partial report.* Another paradigm in support of a long auditory store is modeled after the visual partial report paradigm of Sperling (1960) and a similar auditory task (Moray, Bates, & Barnett, 1965). Darwin, Turvey, and Crowder (1972) presented subjects with three consecutive auditory items simultaneously (digits and letters) at each of three spatial locations, followed by a visual cue indicating which one of three locations the subject was to report. There also was a whole-report condition. It was assumed that auditory information was retained in a spatiotemporal array, and that the visual cue permitted a more complete readout of auditory memory. It was found that the visual cue aided recall in comparison with the whole-report condition with cue delays of 0, 1, and 2 s, but not 4 s, and it therefore was suggested that sensory storage of information in the array decayed by 4 s. Massaro (1975, 1976b) and Kallman and Massaro (1983) cast doubt on this conclusion in a reanalysis of the Darwin et al. data. Massaro (1976b) conducted similar partial report experiments and found that a category cue was more effective than a location cue, arguing against an auditory form of storage. However, there were modifications in the procedure that could have increased the availability of categorical information (the use of two spatial locations rather than three) or decreased the availability of auditory information (the use of a tone rather than a light cue).

Two additional studies using the partial report method seem easier to interpret, because intrinsically noncategorical stimuli were used. M. Treisman and Rostron (1972) presented subjects with two-tone series simultaneously at three spatial locations. Using a

recognition probe at various delays, it was found that recognition of tones declined relative to a whole report condition, to a near-chance level as the probe delay increased from 0 to 1.6 s. Rostron (1974) used longer delays and found some decay up to 5 s, although the rate of decay was much lower than in the first 2 s.

*Periodicity detection.* A classic experiment by Guttman and Julesz (1963) has been interpreted by various theorists as illustrating either a short or long auditory storage. This range of interpretation is possible because of the complexity of the results. In the experiment, stimuli were produced from connected repetitions of a single segment of white noise, and the duration of the segment varied between stimuli. For segments presented with 4 to 19 repetitions per second, a repetitive knocking sound called *motorboating* was heard, whereas a continuous *whooshing* sound was heard with 1 to 4 repetitions per second. Nevertheless, one could still easily detect the individual iterations of noise. At one half to one repetition per second, iterations of the segment could be discerned with effort, and at slower rates this periodicity could not be detected. If one looks for evidence of a distinct qualitative change in perception, one finds the disappearance of motorboating when the length of each iteration exceeded 250 ms. Massaro (1972) maintained that the results were consistent with a 250-ms auditory store, although the exact mechanism of motorboating is unclear. On the other hand, subjects retained information about the spectral makeup of each iteration of white noise for up to 2 s, and this provides evidence for a longer auditory storage.

Although the Guttman and Julesz (1963) experiment was carried out only informally, there have been several interesting replications and modifications of their procedure. Pollack (1969) considered that listeners might perform the task simply by remembering points of extreme high or low amplitude, and therefore reconstructed the stimuli within a narrow amplitude range ( $\pm 5\%$  of the mean). He reported (without giving details) that the periodicities were still easy to detect. He also examined the detection of periodicity with pulse trains and with autocorrelated tone patterns (Pollack, 1968, 1969, 1972). More-

over, Warren and Bashford (1981) were able to replicate the main findings of Guttman and Julesz in a carefully controlled study. However, they noted that a highly trained observer was able to detect periodicities as slow as 0.1 Hz (see their Note 2). Thus, the findings on periodicity detection suggest that there are two forms of auditory sensory storage: one lasting roughly 200 to 300 ms and one lasting 2 to 10 s or more.

The finding (Warren & Bashford, 1981) that practice can result in a much longer use of auditory memory has an interesting parallel in a technique devised by Kubovy and Howard (1976). They presented tone bursts with a deviant note in each burst. The deviant note was perceptually segregated from the complex only if adjacent bursts provided continuity for that note (presumably through auditory memory). Most subjects perceived the segregated note only if the interburst interval was less than 1 s, but one musically trained subject did so with a 9.7 s interval.

*Two-stimulus comparison.* One of the simplest types of experimental task, the two-stimulus comparison task (Crowder, 1982b; Kinchla & Smyzer, 1967; Massaro, 1970a; Moss, Myers, & Filmore, 1970; Pisoni, 1973), provides one of the clearest measures of auditory storage. In this type of task, subjects are required to compare two sounds on each trial with a variable delay between sounds; they are asked to judge whether the sounds are the same or different. Crowder (1982b) presented pairs of vowel sounds to be judged same or different by the subject (termed an "AX task") in order to determine the interval between vowels that would result in an asymptotically low level of performance. It was argued that auditory storage aids in the discrimination, and that the point of asymptote reflects the total decay of auditory storage. The resulting estimate of its duration was 3 s. In another, slightly different task (Pollack, 1972), subjects listened to series of pulse bursts that could differ in timing from one burst to the next. A variable silent gap was inserted between bursts. Subjects were able to correctly compare bursts most often with a 500-ms gap and progressively less often with gaps of 2 s and 8 s.

A final study examining auditory memory across several seconds (Egan, Schulman, &

Greenberg, 1961) used a detection task. Subjects could receive a near-threshold tone at any time during a 2-min trial period, but a visual cue was presented up to 2 s after the tone. Subjects were poorer at tone detection with longer cue delays.

It may seem problematic that the various types of research on long auditory storage provide discrepant estimates of the duration of storage. Two additional studies illustrate the range of this discrepancy. Howell (1978) examined reaction time in an AX task requiring consonant discrimination between pairs of synthetic, consonant-diphthong syllables. With silent interstimulus delays of 500 ms or less, subjects made faster "same-consonant" judgments when the consonants and vowels were identical acoustically as well as phonemically. This "same-matching advantage" was eliminated by an interfering vowel, but not by a tone, placed between "A" and "X" (Howell & Darwin, 1977). Apparently, subjects could benefit from within-category consonant information only for about 500 ms. The results of Eriksen and Johnson (1964) are at the other end of the range. They examined subjects' ability to detect near-threshold-level tones while reading. A light cue was given at a variable interval after the tone (or absence of a tone on catch trials), and the subject was required to say whether or not a tone had been presented. In another condition, subjects were asked to report any tone as soon as it occurred. It was found that the percentage of cued detection surpassed immediate uncued detection even with cue delays of 10 s, suggesting that subjects were not detecting the tones as they were presented, but were dependent upon a decaying trace of the tone that lasted at least 10 s. However, a note of caution is in order (cf. Neisser, 1967), because subjects might subvocally encode the tone and later recall this encoding.

The fact that paradigms differ in the observed duration of storage does not necessarily imply that memory decays differently in these paradigms. The paradigms could draw upon a common type of storage, and task demands could modulate the decay that is observed. Specifically, the paradigms probably differ in the completeness of auditory information necessary to succeed in the task and in the complexity of stimuli. In the most demanding

example (Howell, 1978; Howell & Darwin, 1977), subjects would need to retain within-category auditory information from a consonant vividly enough to cause a facilitation of reaction time, and it is not surprising that a relatively short decay time is observed.

Although task differences could account for discrepancies between paradigms, a more careful comparison of the various results is in order. Figure 1 reflects an attempt to carry out that comparison by plotting various data sets on common axes. When a sensitivity measure free of response bias was available or could be calculated, the data were plotted in Figure 1A. Figure 1B depicts data sets that could not be converted to a sensitivity measure but were reported in percentage correct, and Figure 1C shows other measures (see legend) used to examine the suffix effect. Although the data come from many paradigms, one characteristic is shared. In each case, the stimulus to be remembered was followed by a variable interval, and then by another stimulus used to index auditory memory of the target (a response cue or list, a comparison sound, or a suffix item). Other data were excluded because fewer than three delays were used. Further, data generally were included only if the delay was silent. An exception was made for the detection studies (Egan et al., 1961; Eriksen & Johnson, 1964), in which the delay was filled with white noise, because detection tasks would otherwise be unrepresented. Certain data adequately typified by some part of Figure 1 were excluded. For example, two of four vowel discrimination conditions of Pisoni (1973) were selected for the figure (the highest and lowest functions). The consonant discrimination data were excluded because of ceiling effects (for between-category conditions) and floor effects (for within-category conditions). The suffix effect data of Crowder (1978) were included, but earlier data (e.g., Crowder, 1973) were excluded because the results are similar. What is presented is intended as a concise but representative view of results that could index the decay function of long auditory storage. Interestingly, some of the data may appear unfamiliar, because the present vertical versus horizontal scale factors are sometimes quite different from authors' own figures (e.g., cf. Rostron, 1974).

Much could be said about the differences in task demands in these various paradigms, but the figure suggests that this might be unnecessary, and that differences in results are more apparent than real. The functions suggest that after the stimuli are identified (and performance reaches a maximum), there is a rapid decay of storage in roughly the first 2 s. There is a slower decay thereafter, for at least 10 s. There is a general similarity among studies in the rate of decay in a given range of poststimulus delays, especially in the data expressed in  $d'$  (sensitivity) scores. Discrepancies in the estimates of auditory sensory memory persistence may be due largely to the differential sensitivity and range of delays examined in various experiments.

In summary, the data suggest that there are at least two distinct types of auditory memory: a short auditory store that lasts about 200 to 300 ms and a long auditory store that may last 10 s or more (its potential lifetime in the absence of interference is uncertain). However, one need not rely upon the observed differences in decay time in order to conclude that two auditory stores exist. There are also differences in the characteristics of these two stores. The most important differences are as follows. First, until short auditory storage has decayed, the subject believes that the stimulus is still present. Therefore, the durations of brief stimuli are overestimated (Efron, 1970a, 1970b, 1970c). In contrast, none of the paradigms described in support of long auditory storage seem to result in an extension of the experience of actual stimulation. Thus, short auditory storage is more literal. Second, short auditory storage seems capable of containing, at most, information about one brief sound segment. Each sound seems to overwrite the literal, short auditory storage of the previous sound, as observed in studies of the backward masking of recognition (e.g., Massaro, 1972). In contrast, long auditory storage seems capable of containing information from a sound sequence, inasmuch as each sound may only partially interfere with long auditory storage of previous sounds (Kallman & Massaro, 1979; A. M. Treisman, 1964; M. Treisman & Rostron, 1972). Other possible differences between short and long auditory storage are considered within the discussion of research

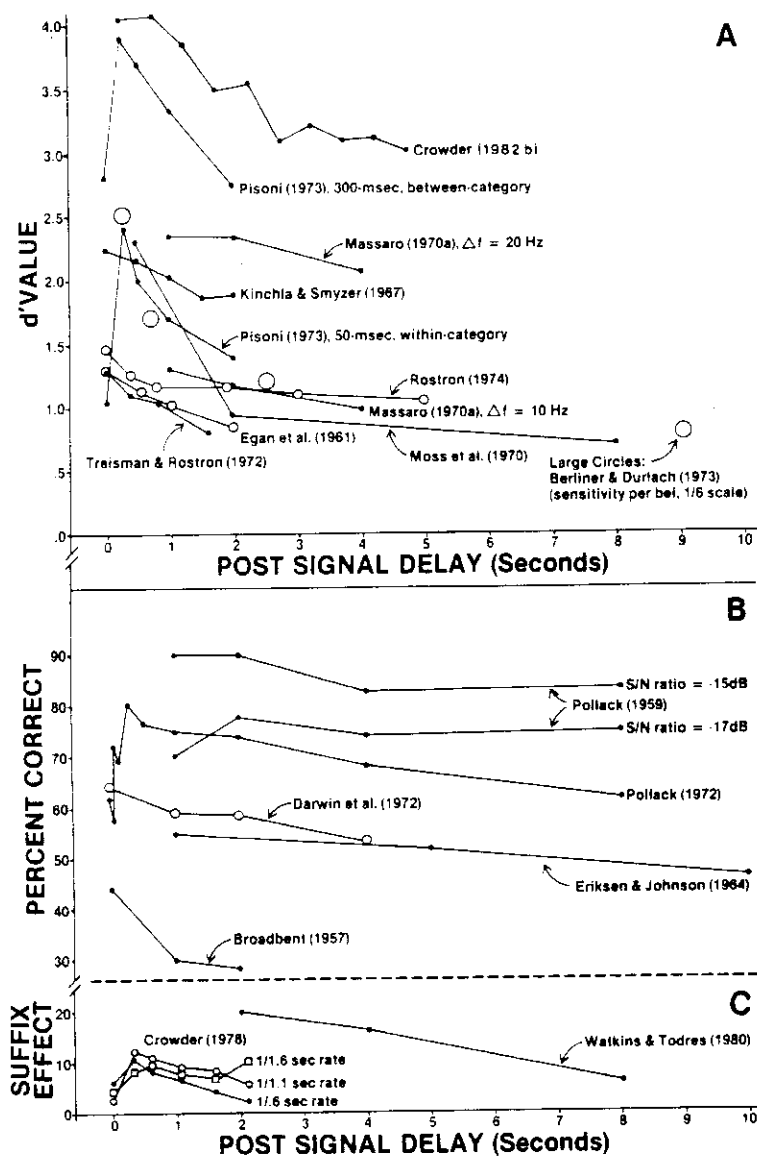


Figure 1. Measures of the decay of long auditory memory during a postsignal period. (A: studies in which a bias-free sensitivity measure was used. From top to bottom: Crowder [1982b, Expt. 2], AX vowel comparison; Pisoni [1973, Fig. 3], AX vowel comparison, 300-ms, between-category condition; Massaro [1970a, Table 2], AX tone comparison, blank interval,  $\Delta f = 20$  Hz; Kinchla & Smyzer [1967, Table 3], AX tone loudness comparison; Pisoni [1973], 50-ms, within-category condition; Rostron [1974], partial report of tones; Massaro [1970a],  $\Delta f = 10$  Hz; Moss et al. [1970, Fig. 2], AX tone comparison; Egan et al. [1961, Fig. 2], tone detection with a visual postsignal cue; M. Treisman & Rostron [1972, Fig. 1], partial report of tones; Berliner & Durlach [1973, Expt. 1, Fig. 8], loudness comparison in  $d'$  = sensitivity per bel. B: studies in which a percent correct measure was used. Pollack [1959], recognition of words in noise, postsignal list; Pollack [1972, Fig. 3, top left], comparison of pulse trains; Darwin et al. [1972, Fig. 1], partial report of alphanumeric items, third item of three; Eriksen & Johnson [1964, Fig. 1], tone detection while reading, postsignal cue, "certain + guess"; Broadbent [1957, Table 2, top row], recall of digits in an unattended channel. C: measures of the stimulus suffix effect across suffix delays. Watkins & Todres [1980, Fig. 1], difference in recall of final list item with vs. without a suffix; Crowder [1978, Expt. 5], proportion of errors on the last item, 3 list rates.)

issues below, but first a consideration of truly long-term auditory store is presented.

*Permanent or long-term auditory memory.* Despite the abundance of studies demonstrating that auditory properties are quickly forgotten (i.e., within seconds), it is obvious from phenomena such as voice recognition that some aspects of audition are retained for a longer period of time. Relevant to this, Wickelgren (1969) presented on each trial a 3-s standard tone, an interference tone that varied from 1 to 180 s in length, and a 1-s comparison tone. Subjects were required to judge whether the standard and comparison tones were the same or different, and the results indicated a decline in memory performance even between 24 and 180 s. Gardiner and Gregg (1979) presented word lists auditorily or visually, with an oral distractor task up to 18 s long between every two words. Despite the auditory (oral) interference, a large auditory modality superiority effect was obtained, suggesting that some aspect of auditory coding is resistant to both decay and interference. Craik and Kirsner (1974) found that words are recognized faster when spoken in the same voice as in the initial presentation, even at delays of 2 min (with 31 intervening items). Presumably, the auditory memory observed in these studies is encoded in a durable form, which may even be permanent. These studies have not established the upper bounds of auditory memory.

Although little is known about the exact form or content of this long-term auditory storage, it may have important implications for a theory of memory. One would not want to postulate a separate auditory store lasting only several seconds if precisely the same type of information were stored much longer. What is most relevant to perception, however, is that the information needed to make fine auditory discriminations decays rapidly across the first few poststimulus seconds. It is important to find out in future research what features distinguish this information from the type that can be saved longer. One possibility is that as the subject becomes familiar with a type of auditory stimulus, a distinctive feature set is developed, and that familiar features can be stored permanently. Although it sometimes would be difficult to delineate such feature sets, several experiments inves-

tigating the use of features in memory for speech are discussed later in this article.

#### Short and Long Auditory Stores: Research Strategies

Although it is now clear that two different types of auditory store exist, many important issues have not yet been resolved. Some of these are addressed below. The discussion of short auditory storage covers quantitative modeling, possible theoretical treatment of stimuli more complex than has been considered so far, and clues to the neural bases of storage. Within the discussion of long auditory storage, quantitative modeling is again addressed. The interrelated issues of the capacity and content of long auditory storage are also discussed. A final section discusses the implications of short and long auditory storage for the information processing system and explores the similarities and differences between auditory and visual processing.

#### *Short Auditory Storage*

*Quantitative modeling.* Zwillocki (1960, 1969) has offered a model for the temporal summation of loudness that seems capable of handling well the data on short auditory storage. In fact, it is consistent with the hypothetical account of storage described above. In Zwillocki's model, neural activity is integrated across time with a weighting function that counts the most recent neural activity most heavily. The neural activity is largest at the stimulus onset and decreases to an asymptotic level by about 200 ms. (The exact time depends on stimulus intensity.) The asymptotic level is nonzero if the stimulus continues beyond 200 ms. If the stimulus then ends, the activity level decays to zero.

The weighting function for integration in Zwillocki's model may be another way to describe auditory storage. This concept is illustrated in Figure 2. An 850-ms stimulus is shown in Figure 2A. The neural response that originates at any point of stimulation continues to sum with ongoing stimulation until it decays (Figure 2B). Note that the neural response is largest at the stimulus onset and equilibrates in about 200 ms. The sum of sensation at any point in time would

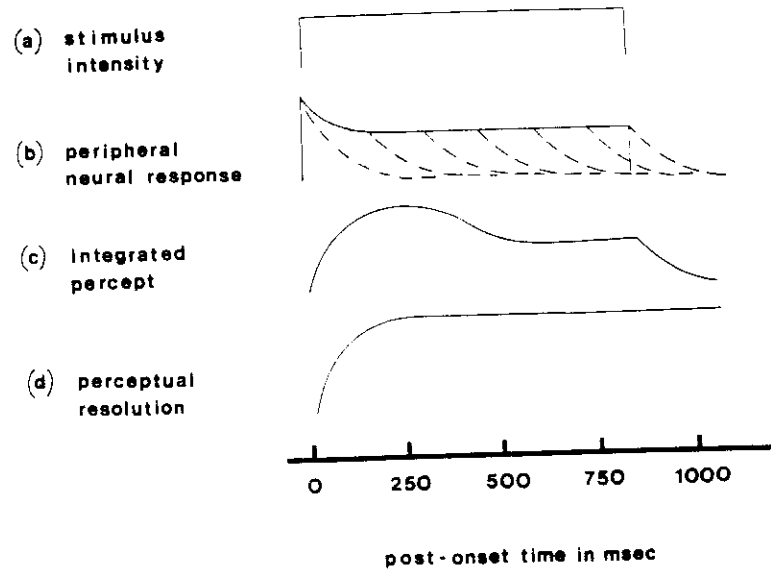


Figure 2. An illustration of Zwislöcki's (1969) model of sensory integration restated in terms of short auditory storage. (The stimulus information is conveyed through the peripheral auditory system with a neural response [b] that is largest shortly after the stimulus onset and equilibrates in 200–300 msec. Broken lines depict auditory persistence originating from the stimulus at various points in time. Persistence is integrated across time to form a percept [c]. The maximum reached by this percept [d] affects subjects' performance on a variety of detection, discrimination, and recognition tasks.)

produce the integrated percept described by Figure 2C, and the maximum reached by this percept (Figure 2D) would determine the degree of perceptual resolution the subject could display in various perceptual tasks (e.g., in perceived pitch).

Zwislöcki (1960, 1969, 1972) supported his model with a variety of neurophysiological and psychophysical data. Specifically, the general function of the peripheral neural response depicted in Figure 2B was confirmed with cellular recordings at various levels of the auditory system. The integrated percept, depicted in Figure 2C, was supported by studies of detection masking and loudness judgment and discrimination. Further, Irwin and Kemp (1976) used a gap detection task to test Zwislöcki's model and found that its quantitative predictions were more accurate than those of another model in which the peripheral neural response and weighting functions were assumed to be flat rather than peaked. Although not mentioned in these articles, it can also be shown that direct loudness estimation data (J. C. Stevens & Hall, 1966) closely match Zwislöcki's model (1969, Figure

7). This is an important consequence of the model. Even if the stimulus remains on, Zwislöcki's weighting parameter ensures that loudness summation or integration does not continue beyond about 200 ms, and this is borne out by the data. This model seems to be at least a reasonable candidate for a description of short auditory storage.

Zwislöcki's model can be compared with a variety of auditory storage phenomena. For this comparison, a simplification of the model is made. In the model there is an initial spike of neural activity lasting only a few ms, represented by  $\epsilon_b$ . Following this spike, the level of activity is roughly  $\epsilon_c + \epsilon_\infty$ , where  $\epsilon_c$  is time-dependent and  $\epsilon_\infty$  represents the asymptotic response level. With  $\beta$  and  $\gamma$  as constants, the total neural response to a stimulus at time  $t$  was said to be:

$$\epsilon(t) = \epsilon_b e^{-\beta t} + \epsilon_c e^{-\gamma t} + \epsilon_\infty. \quad (1)$$

Because  $\beta$  is very large ( $300 \text{ s}^{-1}$ ), the representation of spike activity ( $\epsilon_b e^{-\beta t}$ ) becomes negligible after the first few milliseconds, and it does not seem to be detectable in psychophysical data (Irwin & Kemp, 1976; Irwin &

Purdy, 1982). For the present purposes, the equation will therefore be simplified to:

$$\epsilon(t) = \epsilon_c e^{-\gamma t} + \epsilon_\infty. \quad (2)$$

On the basis of psychophysical data, Zwislowski offered the parameter values  $\gamma = 20$  and  $\epsilon_c/\epsilon_\infty = 2.5$ .

There are two classes of experimental paradigm for which this equation is directly or indirectly relevant. First, there are paradigms in which performance depends upon the decaying trace that follows a sound. An example is forward masking of detection, because detection of the target depends upon the extent to which the trace of the prior mask has decayed. For these phenomena, Equation 2 applies directly. Second, there are paradigms in which performance depends upon the ability of the trace to integrate sound across time. An example is loudness judgment, be-

cause loudness covaries with the duration of the sound. For this type of phenomenon, Zwislowski (1969, Equation 13) provided another formula, derived by integrating the decay equation multiplied by  $e^{-\alpha(t-T)}$ , where  $1/\alpha$  is a time constant of integration equal to 200 ms,  $t$  is the sound's duration, and  $T$  is time.

In the present article, Zwislowski's equations are compared with many of the phenomena that might reflect the decay of short auditory storage, either directly or through integration. In Figure 3 (A-H), various data of both types are plotted. For the phenomena that reflect the decay directly, the theoretical expectation is that the data should match Equation 2 above. The expectation is theoretically different for phenomena that involve integration. The appropriate curve for them would be some transformation of Zwislowski's Equation

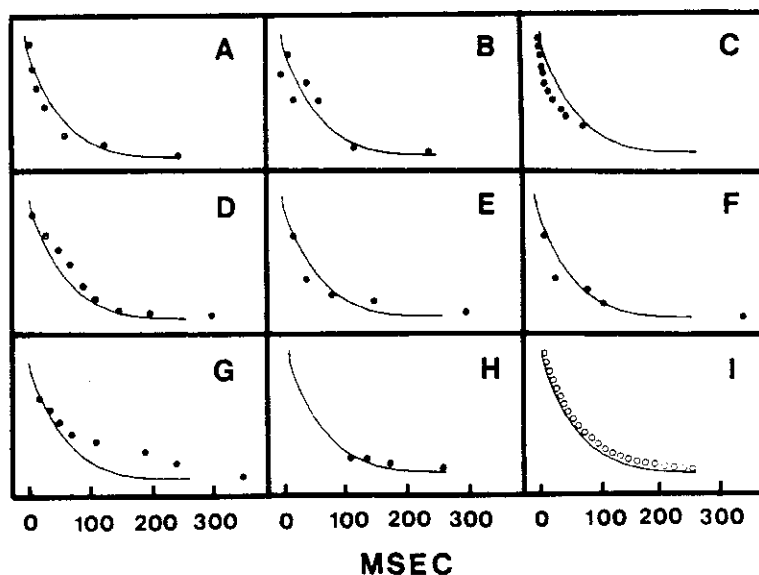


Figure 3. Phenomena that may reflect the decay of short auditory storage. (A: threshold shifts in forward masking of detection across signal delays [Jesteadt, Bacon, & Lehman, 1982]. B: shift in backward masking of detection across masker delays, mask at 70 dB [Zwislowski, 1971]. C: intensity of the second noise burst in a pair needed to detect the gap as a function of gap duration, first burst at 65 dB [Plomp, 1964]. D: overestimation in ms of the duration of a noise burst as a function of burst duration [Efron, 1970a]. E: audible tone loudness increment as a function of the duration of increment, 5 dB base level [Harris, 1963]. F: audible frequency difference as a function of tone duration, 1 kHz standard [Chih-an & Chistovich, 1960/1979], data for stimuli <20 ms omitted because they sound atonal [cf. Garner & Miller, 1947]. G: just-audible level of noise during a period of decreased intensity of variable duration, 60-dB base level [Miller, 1948]. H: constant error in tone counting at various presentation rates [Garner, 1951]. I: the function resulting from differentiation of Massaro's [1975] equation for  $d'$  in backward recognition masking, using his parameter values. Data are superimposed on a simplified version of Zwislowski's [1969] sensory decay function:  $\epsilon(t) = \epsilon_c e^{-20t}$  [solid lines]. In some panels, typical conditions were selected for display; similar functions obtain in other test conditions).



13, which gives the integrated trace. In the integration phenomena depicted in Figure 3, the dependent measures decrease across time (e.g., the detection threshold level decreases as duration increases). For simplicity, assume that for the integration phenomena in Figure 3, the dependent measure theoretically should equal  $(1 - I)$ , where  $I$  is the integrated trace from Zwillocki's equation expressed as a proportion of the asymptotic value. With this simple assumption and Zwillocki's parameter values, it happens that the theoretical curve for the integration phenomena is almost identical to the curve for trace decay (Equation 2), except for the first 10 ms, when the integration-based curve is slightly higher. In Figure 3, it can be seen that Equation 2 or Zwillocki's Equation 13 provide a good first approximation to most of these data sets, despite the variety of dependent measures. A further rationale for the figure is as follows.

In forward masking of detection (Figure 3A), signal detection depends on the amount of trace decay of the mask. In backward masking of detection (Figure 3B), it depends on integration of the decaying target trace. The gap detection paradigm of Plomp (Figure 3C) and the judgments of simultaneity of Efron (Figure 3D) may provide direct indexes of trace decay (see above). Sensitivity to a tone loudness increment depends upon the duration of the increment (Figure 3E), presumably because the integrated trace for the increment period increases as the increment period continues, up to about 200 ms. Similarly, tone frequency discrimination depends upon the tonal duration up to about 200 ms (Figure 3F). However, in the case of a loudness decrement (Figure 3G), the mechanism may be more complex. Does detection depend upon sufficient decay of the trace from the predecrement period or upon the magnitude reached by the integrated trace within the decrement period? Perhaps not surprisingly, the fit is poorest for this phenomenon. Finally, in a tone-counting task (Figure 3H), the magnitude of the integrated trace of each tone may importantly affect the ability to identify the tone and therefore to count tones, although other processes are relevant as well (Massaro, 1976a).

These fits were obtained with the original dependent measures in each case, using a

linear transformation of the dependent measure to standardize the level and range of scores. The level and range were adjusted in such a way that for each plot, two data points (the highest and lowest points, which are termed  $p_{\max}$  and  $p_{\min}$ ) fall on the theoretically predicted curve. However, the shape of the empirical functions and the time scale were not altered. Formally stated, each data point  $p_n$  observed at time  $t_n$  was transformed according to the equation  $p'_n = ap_n + b$ , where

$$a = [e^{-20(t_{\max})} - e^{-20(t_{\min})}] / (p_{\max} - p_{\min}) \quad (3)$$

and

$$b = [(p_{\max})e^{-20(t_{\min})} - (p_{\min})e^{-20(t_{\max})}] / (p_{\max} - p_{\min}). \quad (4)$$

Another asset of Zwillocki's model is that it can be used successfully to model performance in a backward recognition masking paradigm. Assume that the level of discriminability ( $d'$ ) is directly related to the total amount of neural activity in response to the stimulus. For the very brief (and constant-duration) stimuli used in this paradigm,  $\epsilon_{\infty}$  can be neglected, and the neural activity at time  $t$  following the stimulus is:

$$\epsilon(t) = \epsilon_0 e^{-\gamma t}, \quad (5)$$

where  $\epsilon_0$  is the peak activity following the onset of the sound. The subject's  $d'$  would be proportional to the finite integral of this equation from time 0 to  $t$ :

$$k_1 d' + k_2 = \int_0^t \epsilon_0 e^{-\gamma T} dT \quad (6)$$

or

$$d' = \left( \frac{\epsilon_0}{k_1 \gamma} \right) (1 - e^{-\gamma t}) - \left( \frac{k_2}{k_1} \right). \quad (7)$$

Interestingly, Equation 7 is in the same form as a model of performance in backward recognition masking offered by Massaro (1975). Moreover, the parameter values of Zwillocki and Massaro agree fairly well. Using parameter estimates from Massaro,  $(\epsilon_0/k_1\gamma) = 2$  and  $(e^{-\gamma}) = 0.3 \times 10^{-6}$ . (Thus,  $\gamma = 15$  for Massaro vs. 20 for Zwillocki, reasonably similar values.) Differentiating the resulting  $d'$  equation, one obtains  $\epsilon(t) = 30.04 (0.3 \times 10^{-6})^t$ . The last panel of Figure 3 shows that this estimate of sensory decay obtained by

differentiating an equation with Massaro's parameter values provides a good match to the estimate using parameter values of Zwislowski (1969). (The exact decay function according to Zwislowski's model would be  $\epsilon(t) = \epsilon_c e^{-\gamma t} + \epsilon_\infty$  for the duration of the sound, but  $\epsilon_{(t \text{ offset})} e^{-\gamma t}$  after the offset of the sound. One can approximate this two-stage decay with a single exponential function by omitting the term  $\epsilon_\infty$  when the stimulus is brief. However, the exact functions approximate the recognition masking data even better than a single exponential function, because an exponential function underestimates actual performance in the first 50 ms.) Thus, to summarize this work on quantitative modeling, Zwislowski's model seems quite accurate considering the variety of tasks and dependent measures to which the model was directly applied.

Finally, it is instructive to compare backward versus forward recognition masking within the framework of the present model. Both types of recognition masking have been obtained with the sounds in a pair approximately equal in intensity (in contrast to detection masking, in which the mask is more intense). Although there is some forward recognition masking at short interstimulus intervals, it is much less than the amount of backward masking (Massaro, 1973). The best explanation seems to be that in backward masking a decaying trace of the target is interfered with by the mask, which has not yet decayed at all. In contrast, in forward masking it is the trace of the mask that has decayed by the time the target is presented. This degraded trace of the mask still interferes with the perception of the target and lowers performance to some degree, but not to an extent approaching the performance decrement in backward recognition masking. Thus, both types of result can be understood in terms of the conflict between a decaying trace and a subsequent stimulus at its full perceptual strength.

*Complex stimuli.* Our understanding of short auditory storage obviously is incomplete as long as the conclusions are based upon the perception of short, steady-state sounds. One eventually needs to know how the concepts apply to complex stimuli such as ongoing speech or music. Watson, Wroton, Kelly, and

Benbassat (1975) provided a discussion of rationale and strategies for the use of simple versus complex stimuli in experimentation. The traditional psychoacoustic research emphasis on tones, noise bursts, and clicks was based not only on a desire for simplicity, but also on the structuralist belief that responses to more complex stimuli could be derived in an additive fashion. Other more recent research has focused on special properties of speech, but Watson et al. argued that, in order to avoid premature conclusions about the uniqueness of speech, types of complexity built up from simple stimuli also should be investigated. In accordance with this argument, their research yielding clues to the functioning of short auditory storage has involved stimuli only slightly more complex than steady-state sounds (e.g., stimuli composed of steady-state sound sequences or isolated sounds containing a single transition in frequency or spectral composition).

One such paradigm examines perception of temporal order. Determination of the temporal order of sounds in a pair seems to require only about 20 ms between sounds (Hirsh, 1959; Patterson & Green, 1970; Pisoni, 1977; Warren, 1974; Warren & Ackroff, 1976) and discrimination between two pairs differing in temporal order can occur with considerably smaller intrapair intervals (Patterson & Green, 1970; Yund & Efron, 1974). In contrast, if subjects are required to identify each component of the sequence separately, several hundred milliseconds may be required (Broadbent & Ladefoged, 1959; Warren & Obusek, 1972; Warren & Warren, 1970). This discrepancy seems to occur because listeners can perceive brief, rapid patterns holistically without having to identify each component, and a change in the temporal order of components alters the holistic impression (Broadbent & Ladefoged, 1959; Warren & Ackroff, 1976).

The holistic impression of a sound should be influenced by short auditory storage, which is interrupted in rapid sequences for all except the last sound. For example, a rapid, high-low tone sequence sounds lower than does a low-high sequence (Massaro, 1975; Yund & Efron, 1974), probably due to the contribution of the second sound's auditory trace to the total percept. One can interfere with

temporal-order perception by overwriting the auditory trace. In three-tone sequences that include a high, medium, and low tone in any order, subjects often can identify the final tone without being able to determine the order of the first two tones (Divenyi & Hirsh, 1974; Nabelek, Nabelek, & Hirsh, 1973). Moreover, a fourth tone interferes with three-tone order perception (Divenyi & Hirsh, 1975). Finally, a similar process may occur in judgments of relative offset time for two-component stimuli. Tone-offset time judgments can be impaired by backward masking (Pastore, 1983). An alternative model of simultaneity and order judgments based on attention switching was proposed by Allan (1975), but that model was supported using stimuli with one visual and one auditory component.

In addition to temporal-order perception, frequency resolution within tone sequences also depends upon short auditory storage. Watson et al. (1975) examined frequency discriminations within 10-tone sequences lasting 500 ms. It was found that subjects were much more able to detect changes in tones that occurred later in the sequence. The improvement steadily increased throughout the pattern, but there was more improvement in the last 150 ms than in the earlier portion of the complex stimulus. This suggests that in complex sequences, as in simpler sequences, discrimination is easier when the short auditory trace is followed by silence.

Another type of complexity that might be understood in the present framework is spectral transition within a sound. The present theoretical stance leads to two hypotheses. First, when listeners encounter a transient sound, they automatically compute a subjective, integrated frequency spectrum that changes across time. At any moment, the integrated spectrum would be weighted more heavily toward frequencies presented more recently. Second, it is expected that a smooth transition would prevent the backward masking of recognition of segments within the sound. These hypotheses are easily applied to speech perception. The integrated spectrum would represent one important type of information that listeners could obtain from rapidly changing stimuli such as consonants. Recent evidence supports the claim that run-

ning spectra distinguish between consonants (Blumstein & Stevens, 1980; Kewley-Port, 1983; K. N. Stevens & Blumstein, 1981). It is also interesting to note as an informal observation that when synthetic, consonant-vowel syllables are constructed with the second formant transitions reversed across the consonant, but with the vowels intact, the syllables often seem difficult to distinguish from the original stimuli. Temporal summation also may explain why subjects are able to group together mirror-image speech and nonspeech stimuli more easily than stimuli with a common direction of transition (Grunke & Pisoni, 1982). The second assumption about changing stimuli—that smooth transitions prevent masking of recognition—would allow the perception of connected speech without self-masking. To examine this type of assumption, it would be useful to examine perception of temporal order in two-component sequences versus the perception of the direction of change in a sound that moved in continuous fashion from one frequency composition to another.

*Neural basis.* Last, a useful aspect of short auditory storage is that an investigation of its neural basis seems more feasible than for other types of memory. It is quite possible that auditory memory is related to peaks in the evoked response waveform that systematically occur 200 to 500 ms after the onset of a sound when it is processed (e.g., Squires, Squires, & Hillyard, 1975). Abeles, Assaff, Gottlieb, Hodis, and Vaadia (1975) discussed neural circuitry that might subserve this type of memory. They presented relevant 200-ms patterns of single-cell activity within the auditory cortex of cats. A direct comparison between behavioral measures of sensory storage and physiological measures unfortunately has not been made. Contributing to this end, Ross, Ferreira, and Ross (1974) found that when different auditory stimuli were used as the CS+ and CS- in eyelid conditioning, differential conditioned responding could be eliminated with backward recognition masking. This suggests that short auditory storage is incorporated into processing even at an involuntary level. Another way to examine the neural basis of short auditory storage is to employ subjects in infancy, before the cortex is mature. Cowan, Suomi, and Morse

(1982) conducted a backward recognition masking study with 8- to 9-week-old infants, using a nonnutritive sucking procedure. Infants received repeating pairs of vowels in which the first vowel in a pair changed. It was found that a release from backward masking occurred with a 400-ms stimulus onset asynchrony between vowels in a pair, but not with a 250-ms asynchrony, in contrast to the asymptote by 250 ms in adult work. The possibility that the short auditory store lasts longer in young infants than in adults and matures with cortical development is consistent with a variety of behavioral and psychophysiological evidence (see Aslin, Pisoni, & Jusczyk, in press, and Morse & Cowan, 1982, for reviews).

#### Long Auditory Storage

*Quantitative models.* There have been at least four attempts to model the decay of auditory memory across several seconds (Durlach & Braida, 1969; Kinchla & Smyzer, 1967; Massaro, 1970b, 1975; Wickelgren, 1969). Each took a slightly different approach, and the models were based in part upon different types of data. The models used an analogy to signal-detection theory and derived a bias-free measure of subjects' memory strength (analogous to sensitivity). However, the models differ in their conceptualization of the way in which forgetting occurs. Kinchla and Smyzer (1967) and Durlach and Braida (1969), basing their model on the perception and discrimination of tone intensity, proposed that the subject retains a value for loudness of the target, but that random variation is added to this value during the delay interval. The presumed effect is that the variance of this stimulus value increases linearly across the delay interval. Wickelgren (1969) and Massaro (1970b, 1975), who focused on tone frequency rather than intensity discrimination, had a different conceptualization. The memory of each perceived tone was said to have an associated strength or distinctiveness, which would allow discrimination of the tone from other sounds. In Wickelgren's (1969) model, memory strength decays according to the sum of two exponential components. In a tone-comparison experiment, it was found that there was rapid forgetting in the first few

seconds and additional forgetting at a much more gradual rate across 180 s. One exponential component represented the relatively rapid decay, and the other component represented the slow decay. However, Wickelgren's experiments included an interference tone between the standard and comparison. Interference rather than decay could have caused forgetting. Massaro (1970b) proposed an alternative model based on the effects of the interfering stimulus presented for  $t_i$  seconds. As the interfering stimulus is processed at the rate  $(1 - e^{-\alpha t_i})$ , the target memory strength decreases accordingly. If the strength of the target immediately after encoding is  $s(t_s)$ , the remaining strength following the interfering stimulus was said to be

$$s(t_s, t_i) = s(t_s)e^{-\lambda(1-e^{-\alpha t_i})}. \quad (8)$$

An attempt was made to compare Durlach and Braida's model with Massaro's model (i.e., one model of each type) in a simple fashion. Using parameter estimates of the respective authors, the models were simplified to the point that only one parameter remained to be estimated within each equation. Both were assessed with data from Crowder (1982b; personal communication, September 1983), because this data set includes values for a larger number of delay intervals than does any other experiment. Durlach and Braida proposed that sensitivity to a stimulus difference based on sensory memory was

$$d' = [\sqrt{2}K \log(I/I^*)]/(\beta^2 + AT)^{1/2}, \quad (9)$$

where  $A$ ,  $\beta$ , and  $K$  are constants and  $T$  is the interstimulus delay. Because their model was intended to deal with intensity differences, it included the quantity  $\log(I/I^*)$ , the logarithm of the intensity ratio of the two sounds. However, the model can be applied to pairs differing in other stimulus dimensions simply by replacing  $\log(I/I^*)$  with some other quantity (which will not affect the following analysis). With simple arithmetic, the model can be restated as  $d' = [P1 \cdot T + P2]^{-1/2}$ , for  $P1 = A/2 K^2 [\log(I/I^*)]^2$  and  $P2 = \beta^2/2K^2 \cdot [\log(I/I^*)]^2$ . Using the parameter values obtained by Durlach and Braida,  $K/\beta = 16$  and  $K/\sqrt{A} = 30$ , it happens that  $P2 = 3.51 P1$ , so  $d' = [P1(T + 3.51)]^{-1/2}$ . This equation was used to fit the data of Crowder (1982b; personal communication, September 1983)

using the nonlinear regression program BMDPAR (Dixon, 1983).

In Massaro's (1970b) model, the sensory information necessary for discrimination decays from its full strength ( $\alpha$ ) at the rate of  $e^{-\lambda(1-e^{-\nu t})}$ . For an experiment with interstimulus delays up to 8 s, it was observed that  $\nu = 0.103$ . The amount of decay reaches an asymptotic level  $e^{-\lambda}$ . To model decay in Crowder's (1982b) data,  $e^{-\lambda}$  probably must differ from the value observed by Massaro, because the interstimulus delay was silent in Crowder's experiment and filled with a tone in the data Massaro used to examine decay. BMDPAR was used to find the best fit of the equation  $d' = 4.2 e^{-\lambda(1-e^{-0.103T})}$  to Crowder's data. The value 4.2 is a rough estimate of  $\alpha$ , performance free of any decay, for Crowder's stimuli. The best fit occurred with  $\lambda = 0.916$ .

Interestingly, both models fit the data almost equally well, as shown in Figure 4

( $MS_{error} = 0.016$  for Durlach and Braida's model as opposed to 0.021 for Massaro's model). In comparing the models, it is also important to consider how the models should change for stimulus sets at other levels of difficulty. Presumably, within Massaro's model,  $\alpha$  (rather than  $\lambda$ ) should change, because  $\alpha$  represents sensitivity free of decay. Within the Durlach and Braida model,  $P1$  must change, because it is the only parameter not set by their (static) parameter estimates, and because it is sensitive to changes in  $\log(I/I^*)$  or whatever measure of stimulus discrepancy is used instead. However, the family of decay curves resulting from the equation  $d' = P1(T + 3.51)]^{-1/2}$  across 10 s is similar to the family of curves resulting from  $d' = \alpha e^{-0.916(1-e^{-0.103T})}$ . This equivalence is illustrated in Figure 4. This comparison suggests that it may be difficult to decide between the two models. The similar outcomes of the

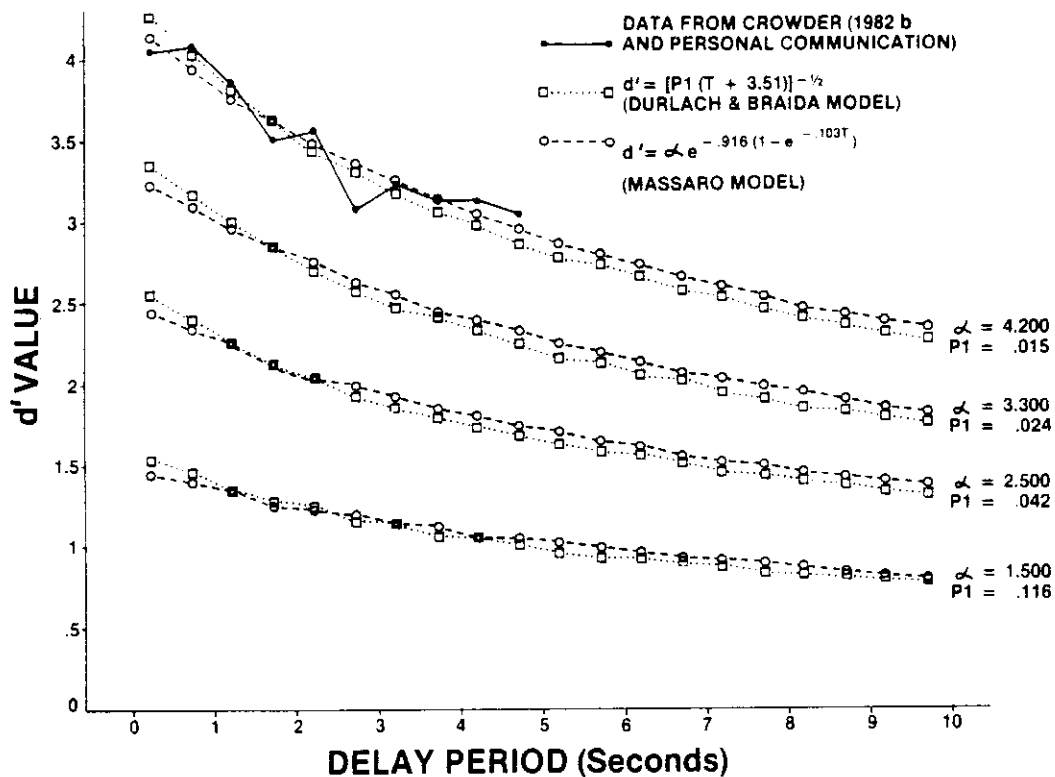


Figure 4. Temporal decay functions produced by two models of long auditory storage, with one variable parameter in each model. (Also shown: vowel discrimination data from Crowder (1982b; personal communication, September 1983).

models are not due to a logical equivalence: At much longer delay intervals (e.g., >100 s), the Durlach and Braida model produces a value approaching 0, whereas Massaro's model approaches a positive asymptote of  $\alpha e^{-\lambda}$ . Durlach and Braida did propose that when sensory memory is unavailable, subjects can rely on a second, "context" mode of perception that is independent of the inter-stimulus delay, and this makes the models' predictions even more similar. Thus, it is a challenge not yet met to derive differential predictions of the models in a common test situation.

Rather than attempting to choose between these models, a more salient point is that either model, when fit to Crowder's data, suggests that auditory memory decay may continue for 10 s or more, as Figure 4 illustrates. Crowder found no decay after about 3 s, but this may be because the range of delays used was too small to pick up additional decay. In any case, these two models, in combination with data from Eriksen and Johnson (1964), Moss et al. (1970), Pollack (1972), and M. J. Watkins and Todres (1980), all shown in Figure 1, suggest that the duration of auditory memory may be longer than 3 s.

*Capacity and content of storage.* Past research has not made clear how much material can be placed in long auditory storage at once or in what form it is held. One hypothesis about the capacity of storage is that any amount of material can be entered, although it is subject to the properties of decay that already have been discussed. However, this possibility can be ruled out inasmuch as different types of stimuli following a target create different amounts of interference (e.g., Massaro, 1970a; Morton et al., 1971). Another possibility is that, subject to decay, auditory input will enter into a single store without interference as long as it continues to be perceived as coming from a single source or stream. This depends upon the similarity of successive sound segments, but also upon the presence of smooth transitions (Bregman, 1978; Bregman & Dannenbring, 1973; Cole & Scott, 1973; Dorman, Cutting, & Raphael, 1975). Whether perception of a single stream improves auditory memory of a target within the stream has not yet been

examined. It is clear, at least, that temporal-order perception is hindered by stream splitting (Bregman & Campbell, 1971; Dorman et al., 1975). In a continuous signal with smooth transitions to prevent stream splitting, such as speech, it is possible that long auditory storage is primarily decay limited.

There is an apparent paradox concerning sound similarity and the capacity of long auditory storage. When two sounds are presented in succession, the amount of interference with the first sound is directly related to their similarity (Kallman & Massaro, 1979). However, if the sounds are too dissimilar, they are likely to be perceived as coming from separate origins (streams). This would almost certainly be detrimental in the perception of ongoing speech, if only because one must perceive the correct temporal order of speech sounds (and possibly because identification of the individual sounds also would be impaired). Smooth transitions between sounds therefore may aid in speech perception, by allowing adjacent segments to be maximally dissimilar without the occurrence of stream splitting.

Another important issue that has not been resolved is the form of items in long auditory storage. It is certain that long auditory storage is not completely literal, or else it should be experienced as a continuation of the stimulus. On the other hand, one should bear in mind that some of the experiments used to index long auditory storage employed tones, noise segments, or speech stimuli within a phonetic category, so one cannot easily take the position that nothing is stored except conventional category labels. It is theoretically possible, however, that subjects generate private categories and store labels for those categories. However, for a novel sound it is difficult to envision this type of categorization. Another possibility is that subjects automatically assign values to stimuli on various continuous dimensions (e.g., pitch or vowel height). Deutsch (1975) suggested that the perceptual process of assigning a sound a value on a dimension is one type of categorization and argued for the existence of this process on the basis of pitch perception phenomena.

An assessment of the alternative positions might be aided by analysis of error patterns in recall experiments with multiple speech

stimuli, in which the delay between the stimulus and recall cue is manipulated. If subjects retain speech sounds as collections of features, then they might tend to forget these sounds one feature at a time. On the other hand, if speech sounds are retained in a more integral form, there might be no consistent temporal pattern in the errors, other than a decline in the percentage of correct recall, as the delay increases. The available evidence favors the feature theory of auditory memory. Wickelgren (1965, 1966) presented sets of consonants or vowels and required written recall of each set. Patterns of errors confirmed that for both consonants and vowels, features are forgotten separately. Further, Cole (1973) found that a feature analysis predicted errors fairly well in a list of consonants or vowels to be recalled, although somewhat more successfully for vowels than for consonants.

#### *Auditory Memory and Models of Processing*

The two types of memory discussed in the present article have been described as separate stores. However, this does not mean that the conclusions require complete acceptance of a storage model of processing (e.g., Atkinson & Shiffrin, 1968). The storage metaphor has been used here only for clarity. The data indicate that the short auditory store and long auditory store function differently and on very different time scales, but the two stores could refer either to separate structures or to separate processes.

In its simplest form, the storage metaphor would suggest that information flows from one store or compartment to another (i.e., through sensory store, short-term store, and long-term store), although the compartments do not necessarily refer to separate areas of the brain. In each transition from store to store, the information is encoded or processed in some way. To account for the data discussed above within this model, there would have to be a sensory store prior to initial stimulus recognition, and another, longer store following recognition (cf. Massaro, 1975). Two alternatives to this approach must be considered. First, in the levels of processing approach ( Craik & Lockhart, 1972), the characteristics of memory are not caused by separate stores, but by a continuum of en-

coding of the stimulus. The more elaborate the encoding of a stimulus, the better it will be remembered (that is, memory is a byproduct of processing). The continuum of coding would shift from acoustic and phonetic (shallow forms of processing) to semantic (a deep form of processing). In this view, then, the raw sensory properties retained in short storage are lost because they represent the most shallow form of processing. Retention for several seconds of auditory qualities that distinguish one sound from another presumably indicates that a more elaborate type of processing has taken place. Di Lollo (1980) put forth a similar account of sensory storage in the visual modality and suggested that the expressions of sensory persistence are limited only by the discernable phases of information processing. A second, slightly different view would be that the two auditory stores differ quantitatively rather than qualitatively. In this view, the large trace magnitude needed for sensory persistence is lost in about 300 ms, but the residual trace magnitude is sufficient to retain useful auditory memory for some seconds. The two alternatives to the two-store model are addressed in turn.

Baddeley (1978) warned of a circularity in the levels of processing approach inasmuch as there are insufficient independent measures of the depth of processing apart from the mnemonic effects. He also reviewed evidence that rather than a processing continuum, memory seems to be influenced by acoustic/phonetic and semantic processing as discrete modes. If there are only a few discrete types of processing rather than a continuum, there is insufficient cause to rule out the concept of separate memory stores, because there might be no empirical difference between discrete stores and discrete processes. Interestingly, though, the auditory memory literature includes a type of evidence that may reduce the circularity of the levels of processing argument. It is reasonable that a process should begin at the onset of stimulation and should last for a relatively constant period independent of the stimulus duration, but this may be less reasonable for a store. One type of evidence suggests that short auditory storage represents a process: the direct sensory persistence measure of Efron (1970a, 1970b, 1970c). Regardless of the

stimulus duration, the duration of sensation was found to be 130 to 180 ms from the stimulus onset. Moreover, this finding was duplicated by Efron in the visual modality (see also Di Lollo, 1980), suggesting that there might be a common process following either auditory or visual stimulation. The process is probably initial recognition of the stimulus, which operates on a time scale similar to sensory persistence (Massaro, 1972, 1975).

A qualification of this approach is necessary, however. Stimulus recognition requires a much shorter time for highly discriminable stimulus sets, and there is no evidence that a resolution process of some type continues for 200 ms even after the stimulus is adequately identified. Should it be expected that a quickly recognized sound would produce less sensory persistence than would a sound that had to be processed longer? This seems unlikely. Thus, if brief sensory storage is actually a process, then the process is probably not recognition per se; the process may be a neural pattern underlying both sensory persistence and recognition.

It is not clear how to apply the levels-of-processing approach to long auditory storage. Evidence has been reviewed that auditory memory can last for many seconds, but it is not clear what central process would last that long. Semantic processing may not be a good candidate, because the memory under consideration occurs even for meaningless acoustic patterns. Rehearsal is not a reasonable candidate because long auditory memory exists for unattended material. If the only process is neural activity that preserves auditory information, then the levels of processing approach again becomes circular. A similar caution was issued by Haber (1983, p. 49) in his response to Di Lollo's suggestions that iconic storage is actually a process.

Perhaps it is necessary to formulate a new theoretical position, taking the best from the stores and levels approaches. In this position, stores are not necessarily compartments separate from other aspects of processing. A pattern of neural activity can result in storage and processing of auditory information at the same time. However, the modes and durations of relevant neural activity patterns do not form a continuum. One type of activity results

in short auditory storage, and another type results in storage for a time an order of magnitude larger (although, the second type of activity may require that the first type has occurred). Moreover, not every pattern of neural activity must be simultaneous with an encoding process of some type. Short auditory storage generally seems to coincide temporally with an encoding process, but long auditory storage may not.

The other alternative to the multiple-stores approach maintains that short auditory storage is just long auditory storage in its more intense, initial form. According to this view, one could write a single equation to describe auditory memory decay 100 ms or 100 s after the stimulus. Arguing against this solution, the parameters describing decay for the two stores are vastly different. If decay continued at the rate for short auditory storage shown in Figure 3,  $e^{-20t}$ , then by 1 s the memory trace would be only  $2.1 \times 10^{-9}$  its original magnitude! Thus, this possibility is unlikely, although it cannot be ruled out entirely.

*Auditory versus visual storage.* Recently, there has been serious debate about visual sensory storage, well summarized by Haber (1983) and 32 following commentaries. Haber maintained that research on iconic storage is irrelevant to ordinary perception, because visual stimuli generally remain present, making the icon unnecessary. This argument obviously does not apply to audition, in which most stimuli are transient. However, many of the commentators felt that iconic storage is useful if one allows a sophisticated model of storage rather than the naive concept of a static picture in the head, and the more sophisticated model does apply to audition. One point is that sensory storage probably is a buffer within the flow of information rather than a transmitter of sensory input at discrete intervals (e.g., see Meyer, 1983). Second, an important function of sensory storage, even while the stimulus remains present, is integration across time (see the Coltheart, 1983, and Navon, 1983, commentaries). Third, brief sensory storage may be best viewed as a process (Di Lollo, 1980, and 1983 commentary), and this process might be initiated whenever there is a rapid change in the stimulus (see Phillips, 1983, commentary). Given the modifications, which agree with



the view of auditory storage offered in the present article, there is little clear theoretical difficulty with the sensory storage concept in either modality. Coltheart (1980) argued that visual information storage and visible persistence cannot be due to the same source because they respond differently to the manipulation of stimulus intensity, but Long (1983 commentary) offered disconfirming evidence. Additionally, many of the commentaries referred to an experiment of Davidson, Fox, and Dick (1973), in which a letter string was presented and then masked in one location following an eye movement. The result was that information was lost at the retinotopic locus masked, but paradoxically, the mask appeared to the subject to cover a different letter, which had been presented at the spatiotopic locus of the mask. This experiment does not nullify the concept of iconic storage. It suggests that information is organized differently at various levels of the nervous system, and that there may be a retinotopic locus of masking that does not fully correspond to the subject's subsequent integrated memory (see Navon, 1983, commentary). Similarly, in audition, there might be both ear-specific and central loci for storage or persistence, and this is an important topic for future research.

The properties of sensory storage are generally similar in vision and audition. Many studies in both modalities contribute to the conclusion that there is a brief, unanalyzed sensory trace and a secondary, more processed sensory memory that lasts at least several seconds (Massaro, 1975, chaps. 17, 21, 24, 25, and 26; 1979). Additionally, Broadbent (1956) administered simultaneous visual and auditory stimuli and found that subjects were equally likely to report either modality first (suggesting that they were not making use of a longer lasting auditory store). However, the similarity between modalities is not total, because there is a well-established auditory modality superiority. The underlying difference between modalities may be that the visual mode is more sensitive spatially, whereas the auditory mode is more sensitive temporally (Penney, 1975). The modality effect generally has been obtained in experiments with lists of words presented one at a time, a method that stresses temporal se-

quential coding. If one compares visual with auditory memory for words presented in an instantaneous array, there is little question that the modality effect would reverse, due to the stress on spatial coding.

### Conclusions

This article has attempted to bring together many types of evidence about the nature of auditory memory. The evidence clearly indicates that there are at least two separate types of auditory storage. The first is short auditory storage that lasts only about 200 to 300 ms; it is used in the initial stimulus recognition and carries with it the experience that the sound is continuing. A second type, long auditory storage lasting at least several seconds, is distinguished from the shorter type by its storage properties.

A summary of the proposed differences between short and long auditory storage and the most relevant supporting evidence is found in Table 1. The duration of the short store was determined with gap detection and judgments of simultaneity, whereas the much greater duration of the long store can be observed in the data included in Figure 1. The gap detection and simultaneity judgment paradigms indicate also that short auditory storage is experienced as a continuation of sensation, but the longer storage clearly is not experienced in this way. The task in which the non-sensory status of long auditory storage can be most clearly demonstrated is shadowing of one channel in dichotic listening. Because attention is kept away from the nonshadowed channel and then switched to it, the subject gets the impression of an after-ringing or echo of the nonshadowed input. The stores differ also in the events to which they are time locked. Efron's (1970a, 1970b, 1970c) results suggest that short auditory storage progresses from the stimulus onset and continues for a relatively constant time regardless of the stimulus duration (although this description may apply only to a steady-state signal; a stimulus change might reinitiate the process). In contrast, Massaro's (1970b) model of auditory recognition and forgetting suggests that long auditory storage does not begin to decay until the stimulus resolution process reaches a certain degree of

Table 1  
*Proposed Differences Between Short and Long Auditory Stores and Selected Relevant Evidence*

| Short auditory storage   | Long auditory storage   | Selected evidence   |
|--|---|---|
| Duration: 150-350 ms   | Duration: 2-20 s  | Crowder, 1982b; Efron, 1970a, 1970b, 1970c; Plomp, 1964; data in Figure 1                               |
| Experienced as sensation   | Experienced as memory   | Broadbent, 1958; Efron, 1970a, 1970b, 1970c; Neisser, 1967; Plomp, 1964                                 |
| Constant duration measured from event onset                        | Decay beginning at point of full stimulus resolution                                  | Efron, 1970a, 1970b, 1970c; Massaro, 1970b, 1975; Pisoni, 1973  |
| Storage of unanalyzed input?                                       | Storage of partly analyzed input?   | Cole, 1973; Massaro, 1972, 1976b; Wickelgren, 1965, 1966  |
| Contains a spectral average weighted in favor of most recent input | Can contain a temporally structured sequence if one continuous stream                 | Blumstein & Stevens, 1980; Bregman, 1978; Dorman et al., 1975; Kewley-Port, 1983; Zwislocki, 1960, 1969 |
| Interference is due to overwriting, and cannot be disinhibited     | Interference is partial, depends upon target-mask similarity, and can be disinhibited | Crowder, 1978, 1982a; Kallman & Massaro, 1979, 1983   |

completion. This analysis is supported by AX discrimination data showing an initial improvement and a subsequent decline in performance as the interstimulus interval is increased (Pisoni, 1973). What type of information is held in short versus long storage? Short storage contains relatively unanalyzed information (Massaro, 1972, 1975), whereas long storage may contain a feature composite (Cole, 1973; Wickelgren, 1965, 1966). A reasonable model for the content of short storage is based on Zwislocki (1960, 1969): A recency-weighted, integrated spectrum is stored. The integrated spectrum changes dynamically as the input changes. There is no evidence that short auditory storage is capable of containing at any one moment a temporal sequence of input, although the changing stimulation alters the integrated spectrum. In contrast, long auditory storage can contain a sequence of speech or nonspeech segments, provided that there are smooth transitions from one segment to the next. Finally, because short auditory storage cannot contain a temporal sequence of stimuli, a new stimulus can overwrite a previous stimulus (Kallman & Massaro, 1979, 1983). In contrast, within long auditory storage each stimulus may only partially interfere with memory for previous stimuli (Crowder, 1982a), depending upon the similarity between stimuli (Kallman & Massaro, 1979).

These differences between the two types of auditory storage appear to be important for theories of information processing.

#### References

- Aaronson, D. (1974). Stimulus factors and listening strategies in auditory memory: A theoretical analysis. *Cognitive Psychology*, 6, 108-132.
- Abeles, M., Assaff, V., Gottlieb, Y., Hodis, Y., & Vaadia, E. (1975). Speculations on a neural substrate for immediate memory. In R. Galun, P. Hillman, I. Parnas, & R. Werman (Eds.), *Sensory physiology and behavior* (pp. 117-125). New York: Plenum Press.
- Allan, L. G. (1975). The relationship between judgments of successiveness and judgments of order. *Perception and Psychophysics*, 18, 29-36.
- Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (in press). Auditory development and speech perception in infancy. In M. Haith & J. Campos (Eds.), *Carmichael's manual of child psychology: Vol. 2. Infancy and the biology of development* (pp. 573-687). New York: Wiley.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.
- Ayres, T., Jonides, J., Reitman, J., Egan, J. C., & Howard, D. A. (1979). Differing suffix effects for the same physical suffix. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 315-321.
- Baddeley, A., & Hull, A. (1979). Prefix and suffix effects: Do they have a common basis? *Journal of Verbal Learning and Verbal Behavior*, 18, 129-140.
- Baddeley, A. D. (1978). The trouble with levels: A

- reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, 85, 139-152.
- Balota, D. A., & Engle, R. W. (1981). Structural and strategic factors in the stimulus suffix effect. *Journal of Verbal Learning and Verbal Behavior*, 20, 346-357.
- Beasley, D. S., & Beasley, D. C. (1973). Auditory reassembly abilities of black and white first- and third-grade children. *Journal of Speech and Hearing Research*, 16, 213-221.
- Békésy, G. von. (1933). Über die horsamkeit der ein- und ausschwingrorgange mit berücksichtigung der raumakustik. *Annalen der Physik*, 16, 844-860.
- Berliner, J. E., & Durlach, N. I. (1973). Intensity perception: IV. Resolution in roving-level discrimination. *Journal of the Acoustical Society of America*, 53, 1270-1287.
- Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, 67, 648-662.
- Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 380-387.
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 89, 244-249.
- Bregman, A. S., & Dannenbring, G. L. (1973). The effect of continuity on auditory stream segregation. *Perception and Psychophysics*, 13, 308-312.
- Broadbent, D. E. (1956). Successive responses to simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, 8, 145-152.
- Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press.
- Broadbent, D. E., & Ladefoged, P. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 11, 1539.
- Bryden, M. P. (1971). Attentional strategies and short-term memory in dichotic listening. *Cognitive Psychology*, 2, 99-116.
- Burrows, D. (1972). Modality effects in retrieval of information from short-term memory. *Perception and Psychophysics*, 11, 365-372.
- Campbell, R., & Dodd, B. (1980). Hearing by eye. *Quarterly Journal of Experimental Psychology*, 32, 85-99.
- Chih-an, L., & Chistovich, L. A. (1979). Frequency difference limens as a function of tonal duration. In D. E. Schubert (Ed.), *Psychological acoustics*. Stroudsburg, PA: Dowden, Hutchinson & Ross. (Reprinted from *Soviet Physics-Acoustics*, 1960, 6, 81-86).
- Cole, R. A. (1973). Different memory functions for consonants and vowels. *Cognitive Psychology*, 4, 39-54.
- Cole, R. A., & Scott, B. (1973). Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology*, 27, 441-449.
- Coltheart, M. (1980). Iconic memory and visible persistence. *Perception and Psychophysics*, 27, 183-228.
- Coltheart, M. (1983). Ecological necessity of iconic memory. *Behavioral and Brain Sciences*, 6, 17-18.
- Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science*, 10, 135-136.
- Cowan, N., Suomi, K., & Morse, P. A. (1982). Echoic storage in infant perception. *Child Development*, 53, 984-990.
- Craik, F. I. M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, 26, 274-284.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Crowder, R. G. (1971). The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, 10, 587-596.
- Crowder, R. G. (1973). The delayed stimulus suffix effect following arrhythmic stimulus presentation. *Quarterly Journal of Experimental Psychology*, 25, 433-439.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
- Crowder, R. G. (1978). Mechanisms of auditory backward masking in the stimulus suffix effect. *Psychological Review*, 85, 502-524.
- Crowder, R. G. (1982a). Disinhibition of masking in auditory sensory memory. *Memory and Cognition*, 10, 424-433.
- Crowder, R. G. (1982b). Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 153-162.
- Crowder, R. G. (1983). The purity of auditory memory. *Philosophical Transactions of the Royal Society of London, B* 302, 251-265.
- Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics*, 5, 365-373.
- Dallet, K. (1965). "Primary memory": The effects of redundancy upon digit repetition. *Psychonomic Science*, 3, 237-238.
- Darwin, C. J., & Baddeley, A. D. (1974). Acoustic memory and the perception of speech. *Cognitive Psychology*, 6, 41-60.
- Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, 3, 255-267.
- Davidson, M. L., Fox, M. J., & Dick, A. O. (1973). Effects of eye movements on backward masking and perceived location. *Perception and Psychophysics*, 14, 110-116.
- Deatherage, B. H., & Evans, T. R. (1969). Binaural masking: Backward, forward, and simultaneous effects. *Journal of the Acoustical Society of America*, 46, 362-371.
- Deutsch, D. (1975). Auditory memory. *Canadian Journal of Psychology*, 29, 87-105.
- Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General*, 109, 75-97.
- Di Lollo, V. (1983). Icons no, iconic memory yes. *Behavioral and Brain Sciences*, 6, 19-20.
- Divenyi, P. L., & Hirsh, I. J. (1974). Identification of temporal order in three-tone sequences. *Journal of the Acoustical Society of America*, 56, 144-151.
- Divenyi, P. L., & Hirsh, I. J. (1975). The effect of blanking on the identification of temporal order in three-tone sequences. *Perception and Psychophysics*, 17, 246-252.

- Dixon, W. J. (Ed.). (1983). *BMDP statistical software: 1983 printing with additions*. Berkeley, CA: University of California Press.
- Dorman, M. F., Cutting, J. E., & Raphael, L. J. (1975). Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 121-129.
- Doughty, J. M., & Garner, W. R. (1948). Pitch characteristics of short tones: II. Pitch as a function of tonal duration. *Journal of Experimental Psychology*, 38, 478-494.
- Durlach, N. I., & Braida, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, 46, 373-383.
- Efron, R. (1970a). Effects of stimulus duration on perceptual onset and offset latencies. *Perception and Psychophysics*, 8, 231-234.
- Efron, R. (1970b). The minimum duration of a perception. *Neuropsychologia*, 8, 57-63.
- Efron, R. (1970c). The relationship between the duration of a stimulus and the duration of a perception. *Neuropsychologia*, 8, 37-55.
- Egan, J. P., Schulman, A. I., & Greenberg, G. Z. (1961). Memory for waveform and time uncertainty in auditory detection. *Journal of the Acoustical Society of America*, 33, 779-781.
- Elliott, L. L. (1971). Backward and forward masking. *Audiology*, 10, 65-76.
- Engle, R. W. (1974). The modality effect: Is precategorical acoustic storage responsible? *Journal of Experimental Psychology*, 102, 824-829.
- Eriksen, C. W., & Johnson, H. J. (1964). Storage and decay characteristics of nonattended auditory stimuli. *Journal of Experimental Psychology*, 68, 28-36.
- Gardiner, J. M., Gathercole, S. E., & Gregg, V. H. (1983). Further evidence of interference between lipreading and auditory recency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 328-333.
- Gardiner, J. M., & Gregg, V. H. (1979). When auditory memory is not overwritten. *Journal of Verbal Learning and Verbal Behavior*, 18, 705-719.
- Garner, W. R. (1947). Accuracy of binaural loudness matching with repeated short tones. *Journal of Experimental Psychology*, 37, 337-350.
- Garner, W. R. (1949). The loudness and loudness matching of short tones. *Journal of the Acoustical Society of America*, 21, 398-403.
- Garner, W. R. (1951). The accuracy of counting repeated short tones. *Journal of Experimental Psychology*, 41, 310-316.
- Garner, W. R., & Miller, G. A. (1944). Differential sensitivity to intensity as a function of the duration of the comparison tone. *Journal of Experimental Psychology*, 34, 450-463.
- Garner, W. R., & Miller, G. A. (1947). The masked threshold of pure tones as a function of duration. *Journal of Experimental Psychology*, 37, 293-303.
- Glucksberg, S., & Cowen, G. N., Jr. (1970). Memory for nonattended auditory material. *Cognitive Psychology*, 1, 149-156.
- Green, D. M., Birdsall, T. G., & Tanner, W. P. (1957). Signal detection as a function of signal intensity and duration. *Journal of the Acoustical Society of America*, 29, 523-531.
- Greenberg, S. N., & Engle, R. W. (1983). Voice changes in the stimulus suffix effect: Are the effects structural or strategic? *Memory and Cognition*, 11, 551-556.
- Grunke, M. E., & Pisoni, D. B. (1982). Some experiments on perceptual learning of mirror-image acoustic patterns. *Perception and Psychophysics*, 31, 210-218.
- Guttman, N., & Julesz, B. (1963). Lower limits of auditory periodicity analysis. *Journal of the Acoustical Society of America*, 35, 610.
- Haber, R. N. (1983). The impending demise of the icon: A critique of the concept of iconic storage in visual information processing. *The Behavioral and Brain Sciences*, 6, 1-54.
- Hamilton, P. M. (1957). Noise masked thresholds as a function of tonal duration and masking noise band width. *Journal of the Acoustical Society of America*, 29, 506-511.
- Harris, J. D. (1963). Loudness discrimination. *Journal of Speech and Hearing Disorders* (monograph suppl. 11).
- Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 31, 759-767.
- Hitch, G. J. (1975). The role of attention in visual and auditory suffix effects. *Memory and Cognition*, 3, 501-505.
- Howell, P. (1978). Syllabic and phonemic representations for short-term memory of speech stimuli. *Perception and Psychophysics*, 24, 496-500.
- Howell, P., & Darwin, C. J. (1977). Some properties of auditory memory for rapid formant transitions. *Memory and Cognition*, 5, 700-708.
- Irwin, R. J., & Kemp, S. (1976). Temporal summation and decay in hearing. *Journal of the Acoustical Society of America*, 59, 920-926.
- Irwin, R. J., & Purdy, J. C. (1982). The minimum detectable duration of auditory signals for normal and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 71, 967-974.
- Jesteadt, W., Bacon, S. P., & Lehman, J. R. (1982). Forward masking as a function of frequency, masker level, and signal delay. *Journal of the Acoustical Society of America*, 71, 950-962.
- Kallman, H. J., & Massaro, D. W. (1979). Similarity effects in backward recognition masking. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 110-128.
- Kallman, H. J., & Massaro, D. W. (1983). Backward masking, the suffix effect, and preperceptual storage. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 9, 312-327.
- Kewley-Port, D. (1983). Time-varying features and correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 73, 322-335.
- Kinchla, R. A., & Smyzer, F. (1967). A diffusion model of perceptual memory. *Perception and Psychophysics*, 2, 219-229.
- Klatzky, R. L. (1980). *Human memory* (2nd ed.). San Francisco: Freeman.
- Kubovy, M., & Howard, F. P. (1976). Persistence of a pitch-segregating echoic memory. *Journal of Experi-*

- mental Psychology: Human Perception and Performance*, 2, 531-537.
- Long, G. M. (1983). The icon as visual phenomenon and theoretical construct. *Behavioral and Brain Sciences*, 6, 28-29.
- Massaro, D. W. (1970a). Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, 83, 32-39.
- Massaro, D. W. (1970b). Perceptual processes and forgetting in memory tasks. *Psychological Review*, 77, 557-567.
- Massaro, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 79, 124-145.
- Massaro, D. W. (1973). A comparison of forward versus backward recognition masking. *Journal of Experimental Psychology*, 100, 434-436.
- Massaro, D. W. (1975). *Experimental psychology and information processing*. Chicago: Rand McNally.
- Massaro, D. W. (1976a). Perceiving and counting sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 337-346.
- Massaro, D. W. (1976b). Perceptual processing in dichotic listening. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 331-339.
- Massaro, D. W. (1979). Reading and listening (tutorial paper). In P. A. Kolers, M. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (Vol. 1, pp. 331-345). New York: Plenum Press.
- Meyer, G. E. (1983). Visual persistence: Just a flash in the scan? *Behavioral and Brain Sciences*, 6, 33-34.
- Miller, G. A. (1948). The perception of short bursts of noise. *Journal of the Acoustical Society of America*, 20, 160-170.
- Miskolczy-Fodor, F. (1959). Relation between loudness and duration of tonal pulse: I. Response of normal ears to pure tones longer than click-pitch threshold. *Journal of the Acoustical Society of America*, 31, 1128-1134.
- Moray, N., Bates, A., & Barnett, T. (1965). Experiments on the four-eared man. *Journal of the Acoustical Society of America*, 38, 196-201.
- Morse, P. A., & Cowan, N. (1982). Infant auditory and speech perception. In T. M. Field, A. Huston, H. C. Quay, L. Troll, & G. E. Finley (Eds.), *Review of human development* (pp. 32-61). New York: Wiley.
- Morton, J., Crowder, R. G., & Prussin, H. A. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology*, 91, 169-190.
- Moss, S. M., Myers, J. L., & Filmore, T. (1970). Short-term recognition memory of tones. *Perception and Psychophysics*, 7, 369-373.
- Munson, W. A. (1947). The growth of auditory sensation. *Journal of the Acoustical Society of America*, 19, 584-591.
- Murray, D. J., & Hitchcock, C. H. (1969). Attention and storage in dichotic listening. *Journal of Experimental Psychology*, 81, 164-169.
- Nabelek, I. V., Nabelek, A. K., & Hirsh, I. J. (1973). Pitch of sound bursts with continuous or discontinuous change of frequency. *Journal of Acoustical Society of America*, 53, 1305-1312.
- Nairne, J. S., & Crowder, R. G. (1982). On the locus of the stimulus suffix effect. *Memory and Cognition*, 10, 350-357.
- Nairne, J. S., & Walters, V. L. (1983). Silent mouthing produces modality- and suffix-life effects. *Journal of Verbal Learning and Verbal Behavior*, 22, 475-483.
- Navon, D. (1983). The demise of the icon or of the icon-as-a-picture metaphor? *Behavioral and Brain Sciences*, 6, 34-35.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Norman, D. A. (1969). Memory while shadowing. *Quarterly Journal of Experimental Psychology*, 21, 85-93.
- Pastore, R. E. (1983). Temporal order judgment of auditory stimulus offset. *Perception and Psychophysics*, 33, 54-62.
- Patterson, J. M., & Green, D. M. (1970). Discrimination of transient signals having identical energy spectra. *Journal of the Acoustical Society of America*, 48, 894-905.
- Penner, M. J. (1977). Detection of temporal gaps in noise as a measure of the decay of auditory sensation. *Journal of the Acoustical Society of America*, 61, 552-557.
- Penney, C. G. (1975). Modality effects in short-term verbal memory. *Psychological Bulletin*, 82, 68-84.
- Phillips, W. A. (1983). Change perception needs sensory storage. *Behavioral and Brain Sciences*, 6, 35-36.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 253-260.
- Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 61, 1352-1361.
- Plomp, R. (1964). Rate of decay of auditory sensation. *Journal of the Acoustical Society of America*, 36, 277-282.
- Pollack, I. (1959). Message uncertainty and message reception. *Journal of the Acoustical Society of America*, 31, 1500-1508.
- Pollack, I. (1968). Periodicity discrimination for auditory pulse trains. *Journal of the Acoustical Society of America*, 43, 1113-1119.
- Pollack, I. (1969). Depth of sequential auditory information processing. *Journal of the Acoustical Society of America*, 46, 952-964.
- Pollack, I. (1972). Memory for auditory waveform. *Journal of the Acoustical Society of America*, 52, 1209-1215.
- Ross, L. E., Ferreira, M. C., & Ross, S. M. (1974). Backward masking of conditioned stimuli: Effects on differential and single-cue classical conditioning performance. *Journal of Experimental Psychology*, 103, 603-613.
- Rostron, A. B. (1974). Brief auditory storage: Some further observations. *Acta Psychologica*, 38, 471-482.
- Scharf, B. (1978). Loudness. In E. C. Carterette & M. D. Friedman (Eds.), *Handbook of perception* (Vol. 4, pp. 187-242). New York: Academic Press.
- Shand, M. A. (in press). Sign-based short-term coding of ASL signs and printed English words by congenitally deaf signers. *Cognitive Psychology*.
- Shand, M. A., & Klima, E. S. (1981). Nonauditory suffix effects in congenitally deaf signers of American Sign Language. *Journal of Experimental Psychology: Human Learning and Memory*, 7, 464-474.
- Shriner, T. H., & Daniloff, R. G. (1970). Reassembly of

- segmented CVC syllables by children. *Journal of Speech and Hearing Research*, 13, 537-547.
- Small, A. M., Brandt, J. F., & Cox, P. G. (1962). Loudness as a function of signal duration. *Journal of the Acoustical Society of America*, 34, 513-514.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74 (11, Whole No. 498).
- Spoehr, K. T., & Corin, W. J. (1978). The stimulus suffix effect as a memory coding phenomenon. *Memory and Cognition*, 6, 583-589.
- Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, 38, 387-401.
- Stevens, J. C., & Hall, J. W. (1966). Brightness and loudness as functions of stimulus duration. *Perception and Psychophysics*, 1, 319-327.
- Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 1-38). Hillsdale, NJ: Erlbaum.
- Stevens, S. S., & Davis, H. (1938). *Hearing: Its psychology and physiology*. New York: Wiley.
- Treisman, A. M. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, 3, 449-459.
- Treisman, M., & Rostron, A. B. (1972). Brief auditory storage: A modification of Sperling's paradigm. *Acta Psychologica*, 36, 161-170.
- Turnbull, W. W. (1944). Pitch discrimination as a function of tonal duration. *Journal of Experimental Psychology*, 34, 302-316.
- Warren, R. M. (1974). Auditory temporal discrimination by trained listeners. *Cognitive Psychology*, 6, 237-256.
- Warren, R. M., & Ackroff, J. M. (1976). Two types of auditory sequence perception. *Perception and Psychophysics*, 5, 387-394.
- Warren, R. M., & Bashford, J. A., Jr. (1981). Perception of acoustic iterance: Pitch and infrapitch. *Perception and Psychophysics*, 29, 395-402.
- Warren, R. M., & Obusek, C. J. (1972). Identification of temporal order within auditory sequences. *Perception and Psychophysics*, 12, 86-90.
- Warren, R. M., & Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American*, 223, 30-36.
- Watkins, M. J., & Todres, A. K. (1980). Suffix effects manifest and concealed: Further evidence for a 20-second echo. *Journal of Verbal Learning and Verbal Behavior*, 19, 46-53.
- Watkins, O. C., & Watkins, M. J. (1980a). Echoic memory and voice quality: Recency recall is not enhanced by varying presentation voice. *Memory and Cognition*, 8, 26-30.
- Watkins, O. C., & Watkins, M. J. (1980b). The modality effect and echoic persistence. *Journal of Experimental Psychology: General*, 109, 251-278.
- Watson, C. S., & Kelly, W. J. (1981). The role of stimulus uncertainty in the discrimination of auditory patterns. In D. J. Getty & J. H. Howard (Eds.), *Auditory and visual pattern recognition* (pp. 37-59). Hillsdale, NJ: Erlbaum.
- Watson, C. S., Wroton, H. W., Kelly, W. J., & Benbassat, C. A. (1975). Factors in the discrimination of tonal patterns: I. Component frequency, temporal position, and silent intervals. *Journal of the Acoustical Society of America*, 57, 1175-1185.
- Weber, D. L., & Green, D. M. (1978). Temporal factors and suppression effects in backward and forward masking. *Journal of the Acoustical Society of America*, 64, 1392-1399.
- Wickelgren, W. A. (1965). Distinctive features and errors in short-term memory for English vowels. *Journal of the Acoustical Society of America*, 38, 583-588.
- Wickelgren, W. A. (1966). Distinctive features and errors in short-term memory for English consonants. *Journal of the Acoustical Society of America*, 39, 388-398.
- Wickelgren, W. A. (1969). Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology*, 6, 13-61.
- Wright, H. N. (1964). Temporal summation and backward masking. *Journal of the Acoustical Society of America*, 36, 927-932.
- Yund, E. W., & Efron, R. (1974). Dichoptic and dichotic micropattern discrimination. *Perception and Psychophysics*, 15, 383-390.
- Zwislocki, J. J. (1960). Theory of temporal auditory summation. *Journal of the Acoustical Society of America*, 32, 1046-1060.
- Zwislocki, J. J. (1969). Temporal summation of loudness: An analysis. *Journal of the Acoustical Society of America*, 46, 431-440.
- Zwislocki, J. J. (1971). Central masking and neural activity in the cochlear nucleus. *Audiology*, 10, 48-59.
- Zwislocki, J. J. (1972). A theory of central auditory masking and its partial validation. *Journal of the Acoustical Society of America*, 52, 644-659.
- Zwislocki, J. J., Pirodda, E., & Rubin, H. (1959). On some post-stimulatory effects at the threshold of audibility. *Journal of the Acoustical Society of America*, 31, 9-14.

Received December 6, 1983

Revision received March 26, 1984 ■