Measuring the difference limen for identification of order of onset for complex auditory stimuli

RICHARD E. PASTORE and SHANNON M. FARRINGTON
State University of New York, Binghamton, New York

The approximately 20-msec perceptual threshold for identifying order of onset for components of auditory stimuli has been considered both as a possible factor contributing to the perception of voicing contrasts in speech and as no more than a methodological artifact. In the present research, we investigate the identification of the temporal order of onset of spectral components in terms of the first of a sequence of thresholds for complex stimuli (modeled after consonant-vowel [CV] syllables) that vary in degree of onset. The results provide clear evidence that the difference limen (DL) for discriminating differences in onset time follows predictions based on a fixed perceptual threshold or limit at relatively short onset differences. Furthermore, the DL seems to be a function of context coding of stimulus information, with both the DL and absolute threshold probably reflecting limits on the effective perception and coding of the short-term stimulus spectrum.

Over three decades ago Hirsh reported some temporal limits of perception that have been widely cited and often misunderstood. In various reports, Hirsh (1959, 1967, 1974; Divenyi & Hirsh, 1974; Hirsh & Fraisse, 1964) reported that there is a hierarchy of temporal limits on perception of monaural (or diotic) stimuli, with each of the different temporal thresholds replicated a number of different times by different researchers using a variety of psychophysical procedures (Moore, 1993, summarizes several of the many possible temporal limits on auditory perception). The present research focuses on the order threshold, contrasting it with the simultaneity threshold with which it is most often confused. For two stimuli differing in onset, a difference of only approximately 2 msec is required for detecting the onset asynchrony, thus defining a sensory threshold for “simultaneity.” The simultaneity threshold probably reflects the detection of spectral changes associated with increasing onset asynchrony. Hirsh (1959) also found that an onset difference of approximately 17–20 msec was required to identify which of the stimuli had an earlier onset, thus defining a perceptual threshold for order identification. This perceptual threshold, the focus of the present research, probably represents a lower limit on defining identifiable pitch qualities.

Consistent with these two thresholds are results reported by Divenyi and Hirsh (1974) on the perception of a sequence of three successive tones varied in duration. With component tones of less than approximately 4 msec

in duration, the percept of the stimulus set was of a single, complex sound. When the component tones were somewhat longer, the percept was one of a succession of sounds whose temporal organization was “indeterminate.” It was only when the individual tones were more than approximately 20 msec in duration that they were perceived as a sequence or melody of tones. Rasch (1978) reported a change in perception of a pair of complex stimuli when one stimulus started 30 msec or more before the other. More recently, Sinex and McDonald (1988) demonstrated a neural coding of onset differences of approximately 20–40 msec. These findings have motivated additional efforts to evaluate a possible acoustic basis for the voicing boundary defined along a voice-onset time (VOT) continuum (see, e.g., Kluender, 1991; Kluender, Lotto, & Jenison, 1995; Sinex, McDonald, & Mott, 1991). Thus, there seems to be a perceptual threshold of approximately 15–30 msec for the identification of a pitch quality that can be conjoined with a temporal position.¹ This perceptual threshold is not the same as the sensory threshold of approximately 2–4 msec for simultaneity (cited above).

A hierarchy or ordering of relatively low-level (sensory) and high-level (perceptual) thresholds is not unique to the temporal dimension. Research in our laboratory has shown that when the tuning of a C major chord is altered by changing either the E or G component, musicians easily detect the presence of a mistuning, but require a larger change in frequency to identify which tone was altered (Acker, 1994). This brings us back to Hirsh’s important point that there are a number of different types of temporal limitations, with the short simultaneity threshold probably reflecting a sensory limitation, and the approximately 20-msec order identification threshold reflecting a perceptual limitation.

For a number of years, this approximately 20-msec threshold for temporal order threshold (TOT) identifica-

This research was supported in part by Grant F49069310327 from the Air Force Office of Scientific Research to R.E.P. Opinions expressed are those of the authors and do not necessarily reflect those of the granting agency. The authors would like to thank Barbara Acker for her helpful comments and criticisms of earlier drafts of this manuscript. Correspondence should be addressed to R. E. Pastore, Psychoacoustics and Auditory Cognition Laboratory, Department of Psychology, SUNY-Binghamton, Binghamton, NY 13902-6000 (e-mail: pastore@cc.binghamton.edu).

tion or judgment was of interest both because of its role as an important perceptual limitation and because of its potential relationship to the perception of syllable-initial stop consonants of English contrasted in voicing. As noted by Hirsh (1959), many researchers have observed that such stop consonants typically exhibit a labeling or categorization boundary when voicing onset is delayed by approximately 24–30 msec relative to the release or onset of the syllable (alveolar and velar stops usually exhibit longer voicing boundaries; see below). Hirsh (1959) made the logical conjecture that TOT may be an important underlying basis for cuing the voicing contrast. The existence of TOT implies that although subjects may be able to detect small differences in onset for various components of the complex signal (e.g., speech), an onset difference of at least 20 msec (in ideal laboratory situations) is required for subjects to reliably “perceive” (or identify) which component had the earlier onset. Thus, for example, an initial consonant with voicing lag would be voiceless if the higher frequency resonances (activated by aspiration noise) are perceived as being clearly present before the onset of voicing or activation of the lower frequency resonances.

Later research demonstrated categorical perception along a temporal onset continuum for noise buzz stimuli (Miller, Wier, Pastore, Kelly, & Dooling, 1976) and for pairs of tones (Pisoni, 1977). These stimuli were classified (labeled) in terms of relatively discrete categories with a boundary at approximately 20 msec. Furthermore, discrimination between pairs of stimuli exhibited troughs of chance performance within categories and heightened discrimination for comparisons across the category boundary. This pattern of categorical perception results was fairly similar to that reported for synthetic speech stimuli varying in VOT (see, e.g., Lisker & Abramson, 1970). Thus, the approximately 20-msec TOT was of interest both because it is one of a set of temporal limits on perception and because it has a potential role in differentiating speech stimuli contrasted in voicing; the latter application to speech thus defines what we will call the TOT hypothesis.

Criticisms of the TOT Hypothesis

The first major criticism of the TOT hypothesis was based on a careful, systematic study by Summerfield (1982). Hirsh (1959) found that TOT was independent of stimulus type for a wide range of laboratory stimuli (i.e., stationary tones and noise). In contrast to the apparent invariance of the TOT threshold, a number of studies demonstrated that the VOT boundary for American English syllables was a function of place of articulation, varying from 24–28 msec for labial stops to 28–52 msec (or more) for velar stops (for review, see Pastore, 1987b). Summerfield built on these observations, demonstrating that the temporal onset threshold for tones and noise-buzz stimuli was not only too short, but also relatively constant. These temporal onset, nonspeech continua thus seemed to be independent of many of the spectral parameters of speech important for changes in place of articu-

lation and the VOT voicing boundaries that are dependent on place of articulation. These criticisms appeared, at the time, to provide a strong basis for rejecting the TOT hypothesis.

Following the work of Summerfield (1982), several studies (Lisker, 1975; Parker, 1988; Summerfield & Haggard, 1977) demonstrated that the TOT can be a function of stimulus parameters, especially when those stimulus parameters are consistent with stimulus properties exhibited in natural and synthetic speech stimuli. Hillenbrand (1984) demonstrated some variation in TOT for stimuli with dynamic changes in frequency at onset, thus mimicking with tonal stimuli the frequency changes associated with *F1*, *F2*, and *F3* transitions. Parker (1988) demonstrated changes in the threshold as a function of changes in the lower (*F1*) onset frequency. Pastore and colleagues (Pastore, Harris, & Kaplan, 1982; Pastore, Layer, Morris, & Logan, 1988) demonstrated significant changes in the TOT for stimuli with dynamic frequency transitions at onset coupled with variation in rise time and the presence of an initial release burst. For stimuli with slow rise times (typical of speech stimuli) and strong initial release bursts, TOT was found to be as long as 40 msec for certain types of stimuli, especially those patterned after velar stimuli contrasted in voicing. These results would all seem to provide some support for the TOT hypothesis.

If one adopts an extreme view about the TOT hypothesis, positing that the voicing contrast is based solely on whether or not the listener can perceive the clear, earlier onset of specific components, the listing by Lisker, Liberman, Erickson, Dechovitz, and Mandler (1977) of many separate cues for the voicing contrast represents clear evidence against the hypothesis. A weaker version of the hypothesis is that a temporal limit on the perception of an earlier onset component is an important contributing factor or cue for the perception of voiceless quality, but not the only cue. If the voicing contrast, or any other phonetic contrast, were based on any single factor or threshold, the decades of research by outstanding scientists would have long ago succeeded in identifying the invariant cue determining a phonetic contrast. If the weaker version of the TOT hypothesis is reasonable, the one critical question concerns the specific components of the stimuli that must be perceived in a temporally ordered fashion.

It is not necessary for the voicing contrast for all categories of place of articulation to be based on the ordered perception of a single or invariant set of stimulus components. However, a logical conjecture for one possible, common set of stimulus components for voicing lag continua would be the early activation of the higher formants by aspiration noise within a very brief period following the release of the consonant and prior to the onset of the broader spectrum voicing source.

Criticisms of Order Threshold Findings

In the late 1980s there were two new criticisms of TOT findings, both arguing against even the existence of the

threshold. The argument by Rosen and Howell (1987) is a theoretical one based on methodological issues. The argument by Kewley-Port, Watson, and Foyle (1988) is based on empirical work looking at sensory limits under minimal uncertainty conditions. Although these arguments about the absence of a 20-msec temporal order threshold ignore the replication of the threshold using a number of different psychophysical procedures, we will address the heart of each criticism before summarizing the present research effort.

Hirsh Methodology

Rosen and Howell (1987) argued that the Hirsh (1959) procedure is really a labeling task. When Rosen and Howell plotted the original Hirsh results in terms of a two-category labeling task, they obtained shallow functions, rather than a relatively discrete change in perception. Furthermore, the 50% category boundary was at approximately onset synchrony (0 msec onset difference), with there being no indication of a threshold or limitation at approximately 20 msec. With an understanding of the assumptions underlying the measurement of a DL, and the procedural limitations with which Hirsh (1959) had to deal, one can see that the Rosen and Howell argument that the TOT threshold does not exist is based on the implicit and testable assumption that the relevant DL does not exist.

Hirsh (1959) used standard psychophysical methodology to measure a DL. Relevant notions about measuring the DL can be found in standard texts from the time of the Hirsh research (e.g., Osgood, 1953; Woodworth & Schlosberg, 1954) and are summarized in Figure 1 (panel A) in terms of arbitrary units along the abscissa. According to classic notions, there is a region around the stimulus standard, the point of objective or physical equality (POE), within which subjects perceive an equivalence to the standard; within this region subjects cannot reliably identify a difference from the POE. This region of perceptual equality is called the "interval of uncertainty" (IU), its midpoint is called the "point of subjective equality" (PSE), and the DL is half of the size of the IU. In the specific case of the research by Hirsh (1959), there really are two regions of uncertainty surrounding temporal onset synchrony. There is a very narrow region for the perception of synchrony (with a 2-msec DL, this IU is approximately 4 msec) and a broader region within which subjects can detect asynchrony, but cannot reliably identify the order of onset (with a 20-msec DL, this IU is approximately 40 msec). Beyond this larger IU, subjects can reliably identify the order or direction of a difference from the standard. In panel A of Figure 1, the distribution of three responses indicates the three underlying perceptual regions. As with any threshold, there is a statistical distribution of response probabilities, rather than a quantum or discrete change in the probability of response. Thus, the shape and steepness (slope) of the functions in panel A of Figure 1 are typical of threshold research.

Because of limits on methodology imposed by the nature of temporal differences (especially using the tech-

nology available in the 1950s), Hirsh was forced to use the method of constant stimuli to measure the DL. For methods of constant stimuli, Osgood (1953) described both three-category and two-category procedures for mapping the three response regions. Panel A in Figure 1 shows typical results for measuring the DL using a three-category version of the method of constant stimuli. The abscissa represents an arbitrary designation of a stimulus continuum, with zero being the stimulus standard or POE. In the case of a temporal order continuum, zero would represent onset synchrony, with the continuum representing the relative onset of the two components to the stimulus (e.g., the lower frequency component having an earlier, versus later, onset). The three curves, each representing an assumed underlying perceptual state, plot the relative probabilities for responses "less than" and "greater than" the standard (i.e., discriminable differences), and the "uncertain" category. In this hypothetical illustration, the three response probabilities are based on one normal (uncertain category) and two cumulative normal (less and greater categories) distributions, with the relative probabilities scaled to sum to unity for each stimulus. In classic psychophysical methodology, the IU is estimated from the stimulus values yielding 50% identification for the less and greater categories, with the DL being half this interval. This illustration (panel A of Figure 1) has plotted an ideal IU that is symmetric around the POE (i.e., $PSE = POE$).

The three-response procedure provides the basis for understanding the concepts underlying the measurement of DL, but the procedure cannot be used easily for measuring the DL for identifying order of onset. In measuring TOT threshold, the separate threshold for simultaneity represents a major problem in defining allowable responses for subjects. As noted, as the onset difference increases (in either direction) from onset synchrony, perception changes from (1) synchrony to (2) detection of asynchrony coupled with an inability to identify order, and finally to (3) a clear ordered onset; the three perceptual categories define two perceptual boundaries or thresholds (simultaneity and order). In the Hirsh temporal order study, the targeted middle category is not simultaneity ($IU = 4$ msec), but rather the broader ($IU = 40$ msec) range for inability to correctly perceive order of onset; the middle response category must indicate a failure to accurately identify order of onset rather than perceived equality of onset. The absence of an accurate, meaningful designation for the perceptual category corresponding to the IU means that the three-category procedure could not be effectively implemented. In order to measure the TOT, the simplest and most understandable instructions are to have the subjects use only the two responses that indicate the perceived order of onset. This procedure, used by Hirsh (1959), follows the standard two-category method described by Osgood (1953) for measuring the DL.

The two-response category psychometric functions, derived directly from the three distributions found in panel A of Figure 1, are plotted in panel B of Figure 1.

Figure 1. A summary of the hypothetical distributions of response categories in measuring the difference limen for some arbitrary stimulus value along a hypothetical continuum. Panel A illustrates the three expected perceptual categories: a center region (or interval) of uncertainty surrounding perceptual and physical equality, and regions of the perceived differences above and below the interval of uncertainty. The center curve is a Gaussian distribution; the other two curves are cumulative Gaussian distributions adjusted so that the sum of ordinates at every point equals a constant. The three-category method has responses corresponding to each of the three distributions. Panel B shows the distribution of responses for measuring the DL using the two-category procedure. For any given stimulus value, probability of a given high (or low) response is based on the probability derived from that distribution in panel A plus 50% (for guessing) of the probability from the uncertainty distribution. The 50% threshold values from panel A correspond with the 75% threshold values in panel B.

The “less” responses should occur either when the subject actually perceives a “less” event ($p[\text{less}]$ curve in panel A) or when the subject, in failing to perceive either category ($p[\text{uncertain}]$ curve in panel A), guesses “less.” Assuming an equal probability of guessing, the probability of responding “less” for any given stimulus value therefore should be equal to the $p(\text{less})$ plus one half of $p(\text{uncertain})$. The resulting two-category psychometric

functions have been plotted in panel B. These functions are much shallower than the original, underlying functions (Figure 1, panel A). Furthermore, the 50% point for the two-category psychometric functions, by definition, represents the PSE and should approximate the POE, which, for TOT, should be onset synchrony. Thus, the Rosen and Howell (1987) argument makes the implicit assumption that the middle category (the IU or the DL in

both directions) does not exist, with the resulting correspondence between the PSE and POE then taken as evidence supporting the assumption. A generalization of this logic would imply that no DL exists for any stimulus dimension.

Because any procedure with fewer responses than perceptual categories forces guessing for any undefined categories, the researcher must use a different criterion for threshold. Specifically, the stimulus difference that had produced 50% performance in the three-category procedure (panel A of Figure 1) now reflects the same 50% correct identification (actual perception of a difference) plus guessing correctly for (approximately 0.5) the trials where no difference is perceived ($0.5 * 0.5$). Thus, a 75% criterion for the two-category procedure theoretically yields the identical DL estimate as the 50% criterion with the three-category procedure (as illustrated in Figure 1).

The functions in panel B of Figure 1 provide us with an important, additional indicator of the category structure of the subjects. Although originally based on normal distributions (panel A), the psychometric functions for the two-category procedure (Figure 1, panel B) do not approximate a cumulative normal distribution, but rather, have distinctive deviations (e.g., are too shallow in slope) for stimuli falling within (and near) the IU. The deviation from a reasonable cumulative normal distribution found in panel A of Figure 1 can be interpreted as a strong indication that the subjects have an additional, intervening category that they are forced to distribute between the two allowable response categories; the results from the Hirsh (1959) two-category procedure, scaled in terms of *z* scores, exhibit this characteristic. In addition, it is fairly typical to find such deviations from a reasonable cumulative normal distribution in the labeling results for speech continua, implying the existence of an intermediate perceptual category that was not directly reflected in the set of available responses (Pastore, 1988).

Sensory or Perceptual Limit

From a different perspective, Kewley-Port et al. (1988) also argued that, at the limits of sensory capabilities, there is no temporal order threshold. Their study reports an absence of evidence for a temporal order threshold of approximately 15–20 msec for highly practiced subjects under minimal uncertainty conditions. Although Pastore (1988) questioned whether this study actually evaluated TOT (see also Watson & Kewley-Port, 1988), the important issue for the present research is that Hirsh (1959) and others (e.g., Miller et al., 1976) have explicitly argued that the temporal order limitation is perceptual, and does not represent a sensory limitation. Therefore, the issue of whether or not one finds evidence for a sensory threshold of approximately 15–20 msec, although interesting in its own right, is not relevant to research on TOT threshold and its possible role in the perception of voicing contrasts. Finally, since TOT measurements have been replicated a number of times under a wide variety of conditions relevant to speech stimuli, a single failure to replicate the threshold measurement under extreme lab-

oratory conditions cannot serve to invalidate the concept of the threshold.

More important for the present work is the effort by Kewley-Port et al. (1988) to empirically evaluate temporal order perception in terms of classic psychophysical relationships such as the Weber fraction. Earlier, Pastore (1987a) had addressed from a theoretical perspective the nature of TOT and voicing contrasts in terms of Weber's law, with judgments for short onset stimuli having the DL defined by a fixed TOT and with longer onset stimuli (greater than TOT) exhibiting a DL that follows the Weber fraction ($\Delta t/t = k$). This analysis thus indicates a two-stage function relating t_i , the discrimination standard, to $t_i + \Delta t$, the just discriminably longer onset stimulus. If t_0 is the location of the TOT threshold, then $t_i + \Delta t$, the limit on identification of order, should be at t_0 for all stimuli less than t_0 (e.g., $t_i < t_0$, since any subliminal stimulus should be just discriminable from the stimulus that first exceeds threshold). For stimuli with onsets longer than this threshold ($t_i > t_0$), the location of $t_i + \Delta t$, the identification threshold, should vary continuously and probably should follow Weber's law, with Δt proportional to t_i . If TOT does not exist, then the lower limit on this latter function relating Δt to t_i (and probably reflecting Weber's law) should be the 2-msec threshold for simultaneity.

Trace Versus Context Coding

An alternative form of evidence for TOT might be found in terms of the distinction between trace and context coding of information in a detection or discrimination task. These notions were originally developed by Durlach and Braida (1969) and more recently have been applied to speech (Macmillan, Braida, & Goldberg, 1987; Uchanski, Millier, Reed, & Braida, 1992). An evaluation of types of coding for TOT is of interest not only in its own right, but also because of recent work by Schouten and van Hoesen (1992) on the discrimination of phonemes contrasted in place of articulation, which seems to indicate a lack of trace coding (thus, only use of context coding). If the judgment of order of onset of stimuli with small onset differences is based on whether or not the comparison stimulus exceeds a fixed, relatively abrupt, quantitative change in perception, the availability of trace information should not have a significant effect on judgments based on the presence/absence of a qualitative change in perception. If subjects are using trace information to perform the discrimination task, discrimination performance should be best at short values of the interstimulus interval (ISI) and should decrease with increasing ISI, since trace information decays as a function of time. However, if trace information is irrelevant to task performance (as expected for discrimination standard with a small onset difference), the TOT threshold should be independent of ISI.

If the DL is independent of ISI, discrimination might be based on a type of context coding. Context coding is typically a function of the size of the context (the range of stimuli confronting the subject within a block of trials).

However, a stable qualitative change in perception should not be influenced by differences in the range of context coding, as determined in fixed versus roving discrimination tasks. In a fixed task only one standard and a limited set of comparison stimuli are presented within a block of trials. In a roving task, several different standards and a broader range of comparison stimuli are randomly mixed within a block of trials. Therefore, the stable threshold for short onset differences should be independent of ISI and task type (fixed vs. roving). However, when the onset difference is longer than TOT, we would expect discrimination to probably be based on judgments of relative duration, and thus to probably reflect changes in trace and context coding.

Present Research

In the two experiments reported in this paper, we asked subjects to identify which of two stimuli had the longer onset difference. If both stimuli have onsets that are less than TOT, the major source of potential information for performing the task should be some form of residual sensory trace from the first that is available during the presentation of the second. Such sensory trace information should decay as a function of time. Subjects may or may not be able to effectively use this trace information. If one stimulus, but not the other, has an onset difference that is less than TOT, there is a strong source of context information that should not be altered by increasing the ISI. Finally, if both stimuli have onset differences that are greater than TOT, listeners should be able to use both context and trace coding information, with only the latter type of information decaying as a function of ISI. We examine the latter conjecture in Experiment 1, evaluating temporal order discrimination relative to a standard stimulus with a brief (5-msec) onset difference as a function of ISI. Since the strength of a stimulus trace is a function of time, the importance of trace coding should be reflected in the change in performance as a function of ISI. Experiment 2 then evaluates whether discrimination results for the perception of order of onset under moderate uncertainty conditions (analogous to those found in typical studies of speech) follow the two-stage pattern predicted by the functioning of an absolute threshold for order identification and then, at longer onset differences, Weber's law. In order to vary context coding information, Experiment 2 was run under both fixed and roving discrimination conditions.

GENERAL METHOD

Subjects

A total of 11 subjects participated in the present experiments. All subjects were of student age (18–21 years), had normal hearing, were native speakers of English, and were paid for their participation. Subjects listened to the binaural, 78 dB:SPL stimuli over TDH-40 earphones in commercial sound chambers.

Stimuli

Two sets of 315-msec stimuli were synthesized (10-kHz sample rate, 12-bit converter). These stimuli are illustrated in Figure 2.

The first 15 msec are best understood in the context of the remainder of the stimulus, and thus are described below. The last 300 msec of the stimuli were sinewaves, with both frequency and amplitude changing as a function of time. This tonal portion of the stimuli had two major frequency components that roughly corresponded to the *F1* and *F2* resonances of the consonant-vowel (CV) (/bi/ and /gi/) syllables used by Volaitis and Miller (1992). The lower frequency component (*F1*) began at 180 Hz and rose linearly to a steady state of 330 Hz over the first 20 msec of the tonal portion of the stimulus. The *F1* component remained at 330 Hz for 180 msec, then declined to 300 Hz over the last 100 msec. The frequency of the higher (*F2*) component followed a similar temporal pattern of frequency change, except that the initial transition was 45 msec. The onset portion of *F2* (broken lines in Figure 2) either rose from 1800 Hz to a steady-state frequency of 2200 Hz or fell from an initial frequency of 2400 Hz to the same 2200-Hz steady-state frequency. The *F2* component was constant for 155 msec, then fell to 2000 Hz over the last 100 msec. With a common *F1* component, the two types of stimuli (rising and falling *F2*) therefore differed solely in terms of the onset portion of the higher frequency component.

The amplitude of the original *F1* component and the *F2* components rose as a linear function of duration over the first 45 msec. After maintaining a steady-state amplitude for 155 msec, the amplitude declined in a linear fashion to zero (ground) over the last 100 msec. An onset time continuum was created by replacing the first *n* msec ($n = 5, 10, \dots, 120$) of the *F1* component with silence, then having the amplitude grow in a linear fashion over the next 4 msec to the amplitude of the original stimulus at that point in time. The *F1* onset time continuum varied in steps of 5 msec, with 0 msec (onset synchrony) defined in terms of the original *F1* component. This procedure created 24 different *F1* onset durations.

With the goal of producing nonspeech stimuli that better resembled CV syllables, an initial 5-msec noise burst plus 10 msec of silence were added to the beginning of each of the two *F2* components. The noise burst was a segment of white noise band-pass filtered at two thirds of an octave geometrically centered approximately at either 1800 Hz (1.6–2.5 kHz) for the rising *F2* or 3000 Hz (2.5–4.0 kHz) for the falling *F2*, providing a crude analog to a stop consonant release burst.

Since prior research has indicated that TOT is between approximately 15 and 25 msec for stationary stimuli and for stimuli with different types of linear frequency onset transitions, one of our goals was to evaluate the possible role of an initial release burst. This goal can be accomplished only by eliminating specific aspects of the correlation between burst and *F2* parameters (i.e., holding the burst intensity and silence duration constant; note that the spectral width and centroid of the burst are still correlated with *F2*). Thus, the duration of the burst and the silent interval between the burst and the onset of the higher frequency tone are not necessarily typical of speech. Rather than having the stimuli differ in burst and silent interval duration, with a short interval of approximately 5 msec typical of a labial stop and a longer 10-msec burst and a 15-msec silent interval for a typical velar stop, we used a constant 5-msec burst and a 10-msec interval for both stimuli. Furthermore, the two bursts were equal in spectral intensity (E/N_0), rather than being much weaker for the stimulus patterned after a typical labial stop. If, under these conditions, TOT is both significantly longer than 15–25 msec and is equal for the two types of stimuli, we will have provided some initial evidence that the initial release and subsequent silent interval probably play a significant role in determining the location of the voicing boundary for initial stop consonants differing in place of articulation. Each *F1* component had a 15-msec segment of silence added to its beginning to temporally synchronize the *F1* and *F2* components.

Full stimuli were created by mixing one *F1* and one *F2* component. The selected *F1* and *F2* components were produced simultaneously over separate DAC channels (sharing a common time base), then mixed as analog signals to create stimuli that differed in the

Figure 2. Summary of stimuli used in present research. Stimuli are based on two components labeled *F1* and *F2*. Rising *F2* and falling *F2* stimuli differ in nature of initial noise burst and frequency transition at onset (broken line at beginning of *F2* component). Temporal continua are created by progressive elimination of onset (dashed) portion of *F1* component prior to mixing with specific version of *F2* component. (See text for additional details.)

(delayed) onset of energy in the *F1* component relative to the *F2* component. Thus, there was a rising and a falling *F2* stimulus continuum. A given subject listened to only one type of stimuli (rising or falling *F2*). None of the subjects reported hearing these stimuli as speech.

A nominal onset difference of n msec thus consisted of a 5-msec burst of noise, 10 msec of silence, n msec of the *F2* component only, and then the *F1* and *F2* combined for $300 - n$ msec. Alternatively, timing may be specified in terms more typical of VOT, and thus measured between the release of the syllable and the onset of voicing (marked by the end of the *F1* cutback); in this alternative metric, an n -msec offset difference for the current stimuli would correspond to a VOT of $n + 15$ msec.

Procedure

In both experiments in the present research, we used a two-interval, forced-choice (2IFC) task. On any given trial, two stimuli were presented, each with a different onset time (e.g., 5 vs. 45 msec). The stimulus with the shorter onset time was always the discrimination standard. The computer randomly selected the specific comparison (longer onset time) stimulus from a predetermined set, with the constraint of equal number of presentations of each longer onset stimulus (paired with each discrimination standard) within a block of trials. The computer also randomly determined the ordering of the stimuli in the two intervals, with the constraint that the discrimination standard was presented equally often in each of the two trial intervals (thus, each answer was always correct on 50% of all trials). Subjects were instructed to always indicate, by buttonpress, which of the two intervals contained the stimulus with the longer onset difference.

Subjects were trained initially with the extreme values of onset time differences (5 msec standard and 120 msec comparison) of the assigned stimulus type (rising or falling *F2*). Subjects ran in short blocks of trials, with feedback, until they could perform the task with a high degree of accuracy (at least 90%). Subjects then were run in a fixed condition using a 5-msec standard with the set of comparison stimuli differing in 25-msec steps across the full onset continuum. The goal of this second condition, which also included feedback, was to determine the approximate location of a difference threshold relative to a 5-msec onset asynchrony. These results were used to determine the limited set of comparison stimuli used in the experiments (i.e., a set of seven to nine stimuli centered on the estimated value of TOT).

The data for all conditions were analyzed in a consistent manner. A psychometric function, based on the z score transformed data, was generated for each subject under each condition. Linear regression was then used to estimate the onset difference matching a criterion of a z score of 1.0.

EXPERIMENT 1 Interstimulus Interval

The first experiment estimated the DL for TOT on the basis of a 5-msec discrimination standard, a stimulus minimally longer than the threshold for simultaneity. This estimate was made in the context of also evaluating the effects of varying ISI, and thus investigated whether the distinction between trace and context coding is impor-

tant for the discrimination of temporal order of onset. If subjects can use sensory trace information to improve performance, better performance should be observed for shorter values of ISI where both trace and context coding information exists. At longer values of ISI the sensory trace should have decayed, leaving only context coding information.

Method

Subjects

This experiment began with 8 subjects who were randomly assigned to one of the two treatment conditions. Two subjects (both trained with the falling *F2* stimuli) resigned from the experiment after training and were not replaced. Experiment 1 therefore is based on the data from 6 subjects (4 for rising *F2* and 2 for falling *F2*).

Procedure

A fixed discrimination procedure was used to generate a psychometric function for each of four time intervals between the pair of stimuli on each trial (ISI = 100, 300, 500, and 1,500 msec). There were 50 trials per subject for each comparison stimulus value at each ISI. On each trial, the minimum (standard) onset difference was 5 msec, with the comparison stimuli having longer onset differences whose range was based on the estimated threshold location for the individual subject. Order of ISI condition varied within subjects, except that all subjects initially ran with the 500-msec ISI. (In Experiment 2, with a 500-msec ISI, the fixed conditions resulted in an average DL located at 32 msec for the

5-msec standard, matching the 32.45 intercept of the ISI regression line, and thus indicating that order of running the ISI conditions was not a factor in the results.)

Results and Discussion

The results are summarized in Figure 3, which plots the threshold location for order discrimination as a function of ISI for the 5-msec discrimination standard (the DL is the ordinate value minus the 5-msec value of the standard). Symbols represent individual subjects, with the solid line plotting the mean across subjects.

Despite using the most lax possible criteria for finding significance, the main effects of stimulus type [$F(1,15) = 3.85, p = .14$], ISI [$F(3,15) = .45, p = .72$], and the stimulus type $\times$ ISI interaction [$F(3,15) = .38, p = .77$] were not significant. Although there was a difference in average thresholds for rising *F2* (mean of 4 subjects = 37.3) and falling *F2* (mean of 2 subjects = 26.4) conditions, equal threshold locations (34–35 msec) were found in Experiment 2 (below) for the two types of stimuli with the 500-msec ISI. Thus, the apparent (nonsignificant) differences in threshold locations for the falling *F2* stimuli would appear to be due to subject sampling.

Four of the 6 subjects exhibited threshold locations that were highest at the smallest ISI (100 msec); only 3 subjects exhibited large differences, and 1 (circles in Fig-

Figure 3. The threshold for magnitude of order of onset relative to a 5-msec standard onset difference is plotted as a function of the interstimulus interval (ISI) between the two stimuli presented on a given trial. The open symbols represent the rising *F2* stimuli and the filled symbols represent the falling *F2* stimuli. The solid line represents the mean for the two stimulus conditions.

ure 3) was in the opposite direction (lower than for longer ISI values). Thus, the small average difference over increasing ISI values was not significant, and the resulting psychophysical function was flat. The slope of the least squares regression line of the psychophysical function was -0.002 ($r^2 = .28$) for the full set of means, and a zero slope (to at least three decimal places, $r^2 = .002$) when the mean for the 100-msec ISI was excluded. These results tend to indicate that if trace information was available at 100 msec, it served to hinder, not help, the subjects. Thus, the distinction between trace and context coding is not important for temporal order discrimination. Probable reasons for the lack of importance for trace information will be discussed later.

EXPERIMENT 2

Testing Models

The second experiment provides a direct test of whether there is an "absolute" perceptual threshold by examining the DL as a function of the discrimination standard. If there is an "absolute" perceptual threshold, the location of the DL should be fixed or constant for discrimination standards that are less than the threshold. The threshold for longer discrimination standards should be consistent with standard findings for discrimination of differences in the duration of auditory stimuli. In addition, this experiment compared the size of the DL for fixed and roving procedures, thus further evaluating the underlying coding of information. If the DL for short onset delays is based on a fixed threshold (i.e., TOT), and not on a relative comparison of the stimuli with a variable or context-dependent standard, the estimated DL for such stimuli should be equivalent for fixed and roving procedures. However, the DL for stimuli with onsets longer than TOT were expected to be based on context-dependent judgments, and thus to be larger under roving than under fixed conditions.

Method

Subject

A total of 9 subjects completed this experiment. Five of these subjects had completed Experiment 1. The 4 new subjects were given the same training procedure as that in Experiment 1, but did not run in the ISI task (Experiment 1). Two new subjects and 3 old subjects performed the task with the falling $F2$ stimulus; 4 (2 old and 2 new) subjects performed the task with the rising $F2$ stimulus. The results from the old and new subjects were highly similar; thus participation in Experiment 1 was not a factor in Experiment 2.

Procedure

In this experiment ISI was fixed at 500 msec. Subjects ran in both the fixed and roving discrimination conditions described below.

Subjects participated in both a minimum (fixed) and maximum (roving) uncertainty condition. In the fixed condition, the discrimination standard (stimulus with the shorter onset time) was constant throughout the block of trials. There were eight separate fixed conditions as defined by the discrimination standards, which ranged from 5 to 85 msec in 10-msec steps, excluding 75 msec. The inclusion of an 85-, rather than a 75-, msec standard was meant

to provide a sampling of discrimination at a relatively long onset difference while still limiting the total number of conditions. In the fixed condition only one standard was used in a block of trials with 50 trials per comparison stimulus. In the roving condition, the discrimination standard randomly varied across trials within a block. There were two roving conditions defined by the set of discrimination standards. The short set had five discrimination standards (5, 15, 25, 35, and 45 msec). The long set had four discrimination standards (45, 55, 65, and 85 msec), but the same total 40-msec range as the short set. In the roving condition there were 30 trials per subject per comparison stimulus for each of the discrimination standards. Short practice blocks of conditions preceded the actual data collection for each of the fixed and roving trials.

Results and Discussion

Discrimination With Rising $F2$ Stimuli

The fixed and roving discrimination results are summarized in the two panels of Figure 4. In this figure, location of the discrimination threshold is plotted as a function of the discrimination standard. In the fixed condition, there was little variability across subjects for the different discrimination standards, and it is clear that there exist two separate functions for discrimination. For short onset differences the threshold location is close to constant at approximately 31 msec, consistent with the notion of an absolute threshold.

Once above this absolute threshold, the exact location of the threshold is a linear function of the standard. Since the slope of the linear regression line is 0.92 ($r^2 = .98$), the size of the DL for the longer discrimination standards is approximately constant. Both the constant size of the DL and the magnitude of the DL for these longer onset discrimination standards are consistent with the findings reported by Abel (1972) for single auditory stimuli with durations of 80 msec or less. The formula Abel reported using was employed to estimate the location of the just discriminable stimulus as a function of the discrimination standard. This function, based on the Abel results, is plotted as the broken line in both panels of Figure 4 (and Figure 5 below). (The major deviation from the Abel results was only for the shortest discrimination standards—below 31 msec—where the location of the threshold for temporal onset stimuli is constant.)

The results for the roving condition are summarized in the bottom panel of Figure 4. We again find that the threshold location is constant for small onset differences, indicating an absolute threshold. The value of this constant threshold, 35 msec, was slightly longer than for the fixed condition. We believe that this longer threshold location may reflect added (decision) noise in the roving task, rather than a judgment based on greater context variability. Specifically, discrimination for the 5-, 15-, and 25-msec standards can be based on a fixed perceptual limit, whereas discrimination of the 35- and 45-msec standards requires an actual judgment of change in duration. Therefore, the short roving condition mixes two different decision processes, and thus should be more demanding or noisier. For longer onset standards the location of the threshold was highly similar to the results from the fixed condition, with a regression line slope

Figure 4. The threshold for magnitude of order of onset is plotted as a function of the smaller difference in onset for the stimuli with the rising F_2 component. The two panels separately plot the results obtained under fixed and roving discrimination tasks. The separate symbols represent individual subjects. The two lines are the linear regression solutions for the sets of data for small and large initial differences in onset.

of 1.08 ($r^2 = 1.00$). The function from Abel (1972) provides a basis for comparison across the fixed and roving condition results. Thus, the results clearly indicate two separate functions, with discrimination at short onset differences based on an absolute perceptual threshold, and discrimination at longer onset differences based on a threshold for the duration of the earlier onset component (thus, equivalent to the results from Abel, 1972). Furthermore, the only differences between the fixed and roving condition results can be explained by the existence of the two types of discrimination processes.

Discrimination With Falling F_2 Stimuli

The fixed and roving discrimination results are summarized in the two panels of Figure 5. We again see that discrimination is constant at 32 msec for short discrimination standards. For longer discrimination standards, the location of the threshold is a linear function of the standard with a slope of 1.0 ($r^2 = .99$). With little variability across subjects for the different discrimination standards in the fixed condition, strong evidence again is found for the two separate functions of discrimination. The unity slope of the linear regression line indicates that

Figure 5. Threshold results for falling F_2 stimuli (see Figure 4 for details).

the DL is constant. The results for longer discrimination standards again closely approximate the finding of Abel (1972) (broken line) for the discrimination of duration of single stimuli with durations of 80 msec or less.

The results for the roving condition are summarized in the bottom panel of Figure 5. There is more variability in these results, and 1 subject (filled diamonds) could not perform the task for short discrimination standards, at least for the range of comparison stimuli employed. However, the overall results are still similar to those found in the fixed condition, and indicate two separate functions for discrimination. For brief onset standards the solid line is plotted at a constant 34 msec. For longer onset standards, the regression line has a slope of 0.95 ($r^2 = .96$), and the data again overlap the function from Abel (1972).

GENERAL DISCUSSION

Originally we expected that we would find a discrimination "threshold" for onset differences that exhibits a two-stage function. For short onset differences the location of the threshold would be constant, reflecting a perceptual threshold corresponding to the lower limit on temporal order identification originally reported by Hirsh (1959). The present findings confirm this first expectation.

At longer onset differences we expected the size of the DL, as with most stimulus continua, to follow some approximation of a Weber function. Instead, we found that the size of the DL for longer onsets was approximately constant, thus independent of onset difference, with the

function relating threshold location to the discrimination standard having a slope of approximately 1.0. Thus, detection of differences in onset for stimuli above the absolute threshold for identifying order of onset (or the nature of the earlier component) definitely did not exhibit a Weber function. However, as Figures 4 and 5 indicate, the size of the DL for these longer discrimination standards was quite similar to that reported by Abel (1972). Therefore, the present findings replicate results on temporal discrimination for very different types of acoustic signals. Furthermore, Zera and Green (1993), although investigating a different type of task, also reported results on the detection of differences in temporal onset for complex stimuli that are paralleled by those reported here.

In keeping with the analyses developed by Hirsh (1959) and Miller et al. (1976), we believe that the recognition of order of onset is taking place at a perceptual level of processing. The subjective reports of our subjects are quite consistent both with our own impressions about the current temporal order task and with other reports from other types of discrimination of temporal properties of stimuli (e.g., Divenyi & Hirsh, 1974). The subjects do not seem to be responding directly to temporal properties of the stimuli, but rather to qualitative or perceptual changes that are correlated with, or a function of, the temporal stimulus properties. Examining our stimuli from this perspective provides some insight into the nature of the temporal order task. In this discussion, we assume that long duration stimuli are sufficiently distinct (e.g., sufficiently separated in frequency) to be highly discriminable.

A brief stimulus that is less than 10 msec in duration typically is heard as a click, with such stimuli discriminable from each other only in highly controlled and limited laboratory situations. If one increases the duration of the stimulus beyond approximately 10 msec, the stimulus becomes a tone-pip with the beginning of an identifiable, pitch-like quality. A clear, identifiable pitch (without the click quality) is finally achieved with somewhat longer stimuli. This perceptual sequence is related to the spectrum of brief signals where the effective bandwidth of the signal is an inverse function of duration (in a Fourier analysis, the "sinc" function). Thus, one possibility is that the threshold for temporal order identification may reflect the need for a sufficient onset lead to accurately identify the quality (pitch) of the earlier portion of the stimulus. An alternative statement of this explanation is that order information is based on a slow, late process that requires the earlier, unambiguous identification of the nature of the component stimuli (e.g., high vs. low tone, or auditory vs. visual).

This perceptual sequence can be generalized from the task of identifying which of two stimuli had an earlier onset to the task of identifying the order of a sequence of three stimuli (see, e.g., Divenyi & Hirsh, 1974). With onset differences of less than a specific value, the earliest portion of the initial stimulus provides the subject with only broad-band information that is probably insufficient to identify the stimuli. Stated another way, sub-

jects require more than 10–20 msec of a signal in order to perceive an identifiable pitch-related quality. Thus the temporal onset difference must exceed this value for the subject to have sufficient information to identify (or at least to consistently apply a label to) which of the stimulus components has an earlier onset. If the stimuli have slow rise times or are dynamically changing in frequency composition, it is not surprising that even longer onset differences would be required for subjects to identify which of two stimuli has an earlier onset.

With extensive practice and a very limited set of stimuli, subjects may be able to apply unique labels to the full stimuli, thus allowing for better performance than is typically seen in more natural listening situations. Thus, when one uses very highly practiced subjects with a limited set of conditions, one would expect to be able to push the threshold toward much shorter onset differences. Such results have been reported for acoustic stimuli (Pastore et al., 1982) and for a pairing of acoustic and visual stimuli (Gengel & Hirsh, 1970). One would also expect to see high levels of performance, such as that reported by Kewley-Port et al. (1988). Conversely, when naive or minimally practiced listeners perceive highly complex stimuli that vary in composition (e.g., natural speech), it should not be surprising that limits on temporal order identification would require much longer onset differences.

Trace Versus Context Coding

The results from both the experiments reported here provide information about the importance of trace and context coding in the discrimination of temporal order of onset. In Experiment 1 we used a fixed-discrimination 2IFC task in which ISI was varied. Performance in this task should have been limited primarily by sensory noise. At the shortest ISI, information from a stimulus trace (or echoic store) could have been (theoretically) available as an additional source of information in performing the discrimination but also as an additional source of noise that may limit the use of this trace information. The ability to use trace coding thus would have been demonstrated by improved performance in terms of the DL at the shortest ISI. With the smaller onset standards we found a small, nonsignificant decrease in performance, with the DL located at a slightly longer onset difference. The lack of any strong evidence for use of trace coding may be partly related to the use of a 5-msec discrimination standard. The results from Experiment 2 indicate that temporal judgments for short onset discrimination standards reflect a perceptual limit on temporal stimuli. Such a perceptual limit should not be altered in a significant fashion by the existence of a sensory trace.

In hindsight, the concept of trace coding is probably not applicable to temporal judgments. In a pitch discrimination task, the residual or trace of one stationary stimulus can be important for the listener in judging the similarity or difference with a later stationary stimulus. Such trace information could also be useful in judging the relative intensity of two stationary stimuli. In contrast, the simple persistence of stimulus properties is of

little use in judging dynamic stimulus attributes. Thus, in the judgment of a difference in duration, only complex trace information could be useful. Specifically, if the listener could synchronously activate a stored trace with the onset of the second stimulus, and also have the duration of the trace be identical to that of the original stimulus, trace information would be expected to be important in temporal judgment. However, the notion of synchronous activation of a temporally accurate trace, or even of imagery, is radically different from most realistic conceptualizations of these concepts. Therefore, trace coding, or even imagery, could not assist in temporal judgments unless they possessed the timing properties already necessary to perform the temporal task.

Context Coding and Range Effects

The lack of difference between fixed and roving conditions in Experiment 2 provides a different indication of the nature of the discrimination of differences in temporal onset. In a single interval task, fixed discrimination is limited in theory by sensory noise, whereas roving discrimination is limited by sensory, trace, and context noise (see, e.g., Macmillan et al., 1987). In the present experiments, we used a 2IFC task that allowed trace information to be utilized in the fixed discrimination condition. In Experiment 2 we found that the location of the DL for short onset discrimination standards was slightly longer under roving, than under fixed conditions (e.g., 34–35 vs. 31–32 msec). This small difference in threshold location might be attributed to context noise related to the range of discrimination standards and comparison stimuli within a block of trials. However, as noted earlier, the difference in absolute threshold location may reflect the probability that in the roving condition, judgment of order of onset probably reflects the use of two different criteria. Specifically, when the discrimination standard is quite brief, judgment of the longer stimulus should be based on whether or not that stimulus exceeds threshold, whereas for the longer discrimination standards within the short range (the 35- and 45-msec standards), subjects must process the stimuli in a manner consistent with a difference threshold rather than an absolute threshold. Without prior information about which information to utilize, the listener has to process both types of information, thus leading to greater processing noise and a higher threshold.

Although any differences due to range effects in context noise should be reflected in the difference for all suprathreshold fixed and roving conditions, the results for longest discrimination standards were equivalent. This later finding is surprising, since it would seem to imply that context variance, in terms of variability in the discrimination standard, did not play a major role in determining the size of the DL.

Our task seems to have required that, for longer discrimination standards, subjects estimate the duration of the critical component within the signal (i.e., the portion of the *F2* stimulus that occurred subsequent to the release burst and silent interval at onset and prior to the

onset of the *F1* component). This interval was probably extracted for each of the two stimuli on a given trial and compared directly, independent of the total range of stimuli within a block of trials. This judgment of difference in duration does not seem to have been a function of the complexity of the stimuli, or the fact that the judgment was based on an explicitly defined portion of a longer stimulus, since the results for our longer discrimination standards are quite compatible with the 2IFC discrimination results reported by Abel (1972) for tones and noise varying in duration (see Figures 4–5). Therefore, for longer onset differences, context coding, at least for discrimination of duration differences measured in a 2IFC task, is not an important factor.

Contrast With Earlier Temporal Order Findings

Early work on temporal order identification (e.g., Hirsh, 1959; Miller et al., 1976; Pisoni, 1977) consistently reported temporal order thresholds in the range of 15 to 20 msec—considerably shorter than the category boundary locations usually reported for speech stimuli contrasted in voicing, where the category boundary is defined in terms of VOT (see, e.g., Pastore, 1987b). This difference in threshold or boundary locations was the basis of an important criticism of conjectures about a possible relationship between temporal order limits and the voicing contrast (Summerfield, 1982). The present results are in direct contrast to the very short threshold locations reported in the early temporal order threshold literature. What accounts for the difference between the 35-msec temporal order threshold identified under roving conditions in the current research, and the 15- to 20-msec thresholds identified in the early temporal order research?

Clearly, part of the difference can be accounted for by the use of dynamic, rather than stationary, stimuli. Changing from the stationary stimulus components of the early temporal order studies to stimuli with dynamic frequency changes at onset (Hillenbrand, 1984), and then adding gradual amplitude envelopes (Pastore et al., 1988), has been demonstrated to result in a lengthening of the temporal order identification threshold. In addition, a release burst at onset (followed by a 10-msec silent interval between burst offset and *F2* onset) seems to lengthen TOT. The burst may have a small forward masking effect at a sensory level, and probably makes it more difficult to perceptually identify the nature of the onset portion of the *F2* stimulus (with gradual rise time) prior to the onset of the *F1* component. It is reasonable to expect that differences in burst parameters associated with changes in place of articulation would result in different threshold locations, consistent with those found in the literature.

One implication of this research is that it suggests that the temporal order identification threshold is not independent of all stimulus parameters, as first reported by Hirsh (1959), but, rather, is a function of both the nature of the release burst and the dynamic stimulus properties that are inherent in the articulation of speech stimuli.

Thus, depending on the specific nature of the components within a complex stimulus (and assuming other than extremely highly practiced listeners under minimal uncertainty conditions), the perceptual limitation on the ability to identify the order of onset of stimuli appears to range from the 15–20 msec reported in the early research to the 34-msec estimate for the present, dynamic stimuli, and possibly to slightly longer estimates for our previous research (Pastore et al., 1988). However, to provide a direct comparison with the VOT continuum, one should add the duration of the release burst and subsequent silent interval, which are typically included in defining VOT for voicing lag conditions. However, in contrast to most research on the voicing contrast, Miller (1977) evaluated the voicing boundary for synthetic stimuli that lacked any initial burst (and therefore subsequent silent interval). Although the voicing boundary still differed as a function of place of articulation, the boundaries for alveolar and velar stimuli were relatively short (under a 30-msec VOT). Therefore, the burst and silent intervals do seem to contribute to, but do not completely determine, the difference in voicing boundary as a function of place of articulation.

For the current stimuli, making TOT equivalent to VOT would require adding the duration of the release burst (5 msec) and the duration of the silent interval (10 msec) to the actual lead time necessary to identify the initial component (34 msec). This would result in a threshold at approximately 50 msec. If anything, this temporal threshold boundary may be too long, rather than too short, relative to the category boundary for voicing contrasts. Clearly, eliminating some differences in burst parameters does eliminate threshold differences that have been found as a function of differences in F_2 onset parameters for both speech and nonspeech stimuli. However, use of more realistic burst parameters should further improve the correspondence between TOT and the VOT voicing boundary. What is clear is that the temporal order threshold location represents a parameter-dependent range of values that encompasses, and is only slightly larger than, that found for voicing contrasts.

Symmetry of Order Thresholds

The speech literature on voicing contrasts indicates that English has a category boundary for only voicing lag, whereas Thai has a category boundary for both voicing lead and voicing lag (Lisker & Abramson, 1970). Furthermore, infants presented with a continuum of VOT stimuli exhibit category boundaries for both voicing lead and voicing lag, but not in the identical locations (Aslin, Pisoni, Hennessey, & Percy, 1981). Hirsh (1959) reported that the temporal order threshold is independent of which stimulus has the earlier onset, and thus is symmetric around onset synchrony. If a temporal order identification threshold is one factor that contributes to the perception of these voicing contrasts, why are the category boundary locations not symmetric around onset synchrony?

Although voicing lead and voicing lag may be conceptually symmetric in terms of production considerations, the acoustic properties of the resulting stimuli are not symmetric around onset synchrony. Unlike the original symmetric stimulus continua utilized in the Hirsh (1959) research, where either the lower or the upper frequency component could have an earlier onset, voicing lead and voicing lag are not along a single physical continuum. Voicing lead is characterized by the voiced source activating a relatively stationary, low-frequency resonance prior to the release of the stop consonant. With voicing lag, the silent stop period is followed by an initial release burst, activation of the relatively higher frequency, dynamically changing vocal tract resonances by aspiration noise alone, and then activation by all formants by the addition of voicing. Thus, from acoustic and perceptual perspectives, as well as the parameters influencing the temporal order threshold location, voicing lead and voicing lag are quite different. Since the threshold for stationary and dynamic stimuli have now been demonstrated to be quite different, it should not be surprising that the location of the category boundary for voicing lead and voicing lag should differ, or that some languages may not incorporate distinctions based on one or the other stimulus property.

We thus return to a weak version of the original notion proposed by Hirsh (1959) that a limitation on the ability to identify the order of onset of stimuli may contribute to the perception of the voicing contrast. Furthermore, other changes in the acoustic properties of speech stimuli may function by altering the location of the temporal order identification threshold, and thus may exhibit a trading relation. Clearly, the conjecture of a perceptual threshold on temporal order identification is not the only factor in defining the location of the voicing boundary, but it may be an important factor that still needs to be considered from a theoretical perspective and evaluated empirically.

Categorical Perception

Can the results plotted in Figures 4 and 5 account for categorical perception, as has been demonstrated for VOT continua? First, there do appear to be two perceptual categories. Under random presentation of stimuli (e.g., roving condition) with a category boundary at an onset difference of approximately 34 msec, stimuli with onsets greater than 34 msec can be labeled in terms of the component with the earlier onset. For stimuli with onset differences less than 34 msec, the nature of the onset is relatively indistinct, and thus can be categorized as being equivalent to, yet different from longer onset stimuli. Thus, labeling should be of the form reported by Miller et al. (1976) and Pisoni (1977) for temporal onset continuum, but at a location appropriate to the relevant stimulus parameters.

Should there be a discrimination peak located at approximately the category boundary and troughs of chance discrimination within categories? Figure 6 provides the basis for an affirmative answer to this question. The data

in Figure 6 are based on the roving discrimination results plotted in Figures 4–5. In the upper panel, the size of the DL is estimated by subtracting the standard (abscissa value in Figures 4–5) from the location of the difference threshold (ordinate value of points in Figures 4–5) for the best fitting regression lines.

Let us assume that we run a standard categorical perception discrimination task in which we present a given standard and a comparison stimulus that differs by a constant amount (e.g., the line at 12 msec in the upper panel of Figure 6). The difference between the 12-msec stimulus difference and the DL provides an indication of the ability of subjects to perform the task.² These differences are plotted in the lower panel of Figure 6.

For stimulus standards of less than approximately 20 msec, the 12-msec stimulus difference would be considerably less than the DL, and performance should approach, or be at, chance. Discrimination between a 25-msec stimulus and a 37-msec stimulus (plotted at 25 msec in Figure 6) would represent a peak of high performance in the discrimination function. Performance returns to quite low levels of discrimination for stimuli greater than approximately 35 msec (although for one of the two stimulus sets, the extreme standards of 65 and 85 msec return to above-chance levels).

The pattern of discrimination in Figure 6 meets the criteria of troughs of chance discrimination within each of the two categories and a discrimination peak near the category boundary. However, the discrimination peak does not precisely correspond to the 34-msec category boundary. This discrepancy is due, in part, to the transformed data being plotted as a function of the lower onset difference, rather than the typical categorical perception procedure plotting the midpoint; had we plotted the results at the midpoints, the discrimination peak would be at 31 msec, which is in close proximity to the discrimination threshold. Furthermore, a precise correspondence between the discrimination peak and the category boundary would be quite atypical of categorical perception results, where the actual discrimination peak is usually offset somewhat from the actual category boundary. Therefore, the present results do provide a basis for categorical perception findings for temporal onset continua, including speech, tones, and noise-buzz sequences.

Final Comment

In the temporal order identification task a subject must acquire sufficient information about the stimulus with the earlier onset to be able to consistently apply the appropriate label to the stimulus. Although the research may specify the threshold in terms of the independent variable duration, the task of the subject probably is the acquisition of sufficient stimulus information to perform a type of identification task based on stimulus quality. This threshold defines a lower limit on the judgment of order of onset. If it is based on there being sufficient stimulus information (duration) to achieve a distinct identifiable quality, the threshold for temporal order

Figure 6. Discrimination performance, as a function of onset differences, for fixed stimulus differences derived from the results of Experiment 2. In the upper panel, the size of the difference limen (DL) is estimated by subtracting the standard (abscissa value in Figures 4–5) from the location of the difference threshold (ordinate value of points in Figures 4–5) for the best fitting regression lines. The line at 12 msec represents an example of a fixed stimulus difference that might be used in a standard categorical perception discrimination task. The difference between the 12-msec stimulus change (solid line from upper panel) and the DL provides an indication of subjects' ability to perform the task. These differences are plotted in the lower panel of the figure indicating troughs of chance discrimination (shaded area) within each of the two categories and a discrimination peak near the category boundary.

identification could be considered a perceptual, or even cognitive, limit on performance. The notion of a perceptual or qualitative limit is not unique to this temporal identification task. In a careful analysis of vowels from the perspective of the Durlach-Braida model, Macmillan et al. (1987) found that categorical perception for vowels seems to be based on an anchor at the category boundary, with the anchor being perceptual or cognitive rather than sensory in nature.

With the longer values of temporal onset differences, the limits on perception of pitch is no longer relevant in defining the DL. Instead, subjects probably are required

to make a judgment based on the duration of the initial stimulus component. For these longer stimuli, the temporal order task becomes equivalent to the types of temporal detection or discrimination tasks studied by Abel (1972) and by Zera and Green (1993). Therefore, the temporal order discrimination task seems to involve three separate processes. For very brief onsets, the task is to detect any cues associated with onset asynchrony. For slightly longer, but still very brief, onset differences, the task requires acquisition of sufficient stimulus information to identify the earlier component. At longer onset differences, the earlier onset component seems to function as a separate acoustic stimulus that can be judged in terms of differences in duration.

REFERENCES

- ABEL, S. M. (1972). Duration discrimination of noise and tone bursts. *Journal of the Acoustical Society of America*, **51**, 1219-1223.
- ACKER, B. E. (1994). *Discrimination of musical chord components*. Unpublished master's thesis, State University of New York, Binghamton.
- ASLIN, R. N., PISONI, D. B., HENNESSEY, B. L., & PERCY, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development*, **52**, 1135-1145.
- 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.
- DURLACH, N. I., & BRAIDA, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, **46**, 372-383.
- GENGEL, R. W., & HIRSH, I. J. (1970). Temporal order: The effect of single versus repeated presentations, practice, and verbal feedback. *Perception & Psychophysics*, **7**, 209-211.
- HILLENBRAND, J. (1984). Perception of sine-wave analogs of voice onset time stimuli. *Journal of the Acoustical Society of America*, **75**, 231-240.
- HIRSH, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America*, **31**, 759-767.
- HIRSH, I. J. (1967). Information processing in input channels for speech and language: The significance of serial order of stimuli. In C. H. Millikan & F. L. Darley (Eds.), *Brain mechanisms underlying speech and language: Proceedings* (pp. 21-39). New York: Grune & Stratton.
- HIRSH, I. J. (1974). Temporal order and auditory perception. In H. R. Moskowitz, B. Scharf, & J. C. Stevens (Eds.), *Sensation and measurement* (pp. 251-258). Dordrecht, The Netherlands: Reidel.
- HIRSH, I. J., & FRAISSE, P. (1964). Simultanéité et succession de stimuli hétérogènes [Simultaneity and succession of heterogeneous stimuli]. *L'année Psychologique*, **64**, 1-19.
- KEWLEY-PORT, D., WATSON, C. S., & FOYLE, D. C. (1988). Auditory temporal acuity in relation to category boundaries: Speech and non-speech stimuli. *Journal of the Acoustical Society of America*, **83**, 1133-1145.
- KLUENDER, K. R. (1991). Effects of first formant onset properties on voicing judgements result from processes not specific to humans. *Journal of the Acoustical Society of America*, **90**, 83-96.
- KLUENDER, K. R., LOTTO, A. J., & JENISON, R. L. (1995). Perception of voicing for syllable-initial stops at different intensities: Does synchrony capture signal voiceless stop consonants? *Journal of the Acoustical Society of America*, **97**, 2552-2566.
- LISKER, L. (1975). Is it VOT or a first-format transition detector? *Journal of the Acoustical Society of America*, **57**, 1547-1551.
- LISKER, L., & ABRAMSON, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences* (pp. 563-567).
- LISKER, L., LIBERMAN, A. M., ERICKSON, D. M., DECHOVITZ, D., & MANDLER, R. (1977). On pushing the voice-onset-time (VOT) boundary about. *Language & Speech*, **20**, 209-216.
- MACMILLAN, N. A., BRAIDA, L. D., & GOLDBERG, R. F. (1987). Central and peripheral processing in the perception of speech and nonspeech sounds. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 28-45). Boston: Nijhoff.
- MILLER, J. D., WIER, C. C., PASTORE, R. E., KELLY, W. J., & DOOLING, R. J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, **60**, 410-417.
- MILLER, J. L. (1977). Nonindependence of feature processing in initial consonants. *Journal of Speech & Hearing Research*, **20**, 519-528.
- MOORE, B. C. J. (1993). Temporal analysis in normal and impaired hearing. In P. Tallal et al. (Eds.), *Temporal information processing in the nervous system: Special reference to dyslexia and dysphasia* (Annals of the New York Academy of Sciences, Vol. 682, pp. 119-136). New York: New York Academy of Sciences.
- OSGOOD, C. E. (1953). *Method and theory in experimental psychology*. New York: Oxford University Press.
- PARKER, E. M. (1988). Auditory constraints on the perception of voice-onset time: The influence of lower tone frequency on judgments of tone-onset simultaneity. *Journal of the Acoustical Society of America*, **83**, 1597-1607.
- PASTORE, R. E. (1987a). Categorical perception: Some psychophysical models. In S. Harnad (Ed.), *Categorical perception* (pp. 29-52). New York: Cambridge University Press.
- PASTORE, R. E. (1987b). Possible acoustic bases for the perception of voicing contrasts. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 188-198). Boston: Nijhoff.
- PASTORE, R. E. (1988). Burying straw men without graves: A reply to Kewley-Port, Watson, and Foyle (1988). *Journal of the Acoustical Society of America*, **84**, 2262-2266.
- PASTORE, R. E., HARRIS, L. B., & KAPLAN, J. K. (1982). Temporal order identification: Some parameter dependencies. *Journal of the Acoustical Society of America*, **71**, 430-436.
- PASTORE, R. E., LAYER, J. K., MORRIS, C. B., & LOGAN, R. J. (1988). Temporal order identification for tone/noise stimuli with onset transitions. *Perception & Psychophysics*, **44**, 257-271.
- 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.
- RASCH, E. (1978). Principles of categorization. In E. Rasch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.
- ROSEN, S., & HOWELL, P. (1987). Is there a natural sensitivity at 20 msec in relative tone-onset-time continua? A reanalysis of Hirsh's (1959) data. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 199-209). Boston: Nijhoff.
- SCHOUTEN, M. E. H., & VAN HESSEN, A. J. (1992). Different discrimination strategies for vowels and consonants. In M. E. H. Schouten (Ed.), *The auditory processing of speech: From sounds to words* (pp. 309-314). New York: Mouton de Gruyter.
- SINEX, D. G., & McDONALD, L. P. (1988). Average discharge rate representation of voice-onset time in the chinchilla auditory nerve. *Journal of the Acoustical Society of America*, **83**, 1817-1827.
- SINEX, D. G., McDONALD, L. P., & MOTT, J. B. (1991). Neural correlates of nonmonotonic temporal acuity for voice onset time. *Journal of the Acoustical Society of America*, **90**, 2441-2449.
- SUMMERFIELD, Q. (1982). Differences between spectral dependencies in auditory and phonetic temporal processing: Relevance to the perception of voicing in initial stops. *Journal of the Acoustical Society of America*, **72**, 51-61.
- SUMMERFIELD, W., & HAGGARD, M. P. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, **62**, 435-438.
- UCHANSKI, R. M., MILLIER, K. M., REED, C. M., & BRAIDA, L. D. (1992). Effects of token variability on resolution for vowel sounds. In M. E. H. Schouten (Ed.), *The auditory processing of speech: From sounds to words* (pp. 291-302). New York: Mouton de Gruyter.
- VOLAITIS, L. E., & MILLER, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, **92**, 723-735.

- WARREN, R. M. (1982). *Auditory perception: A new synthesis*. New York: Pergamon.
- WATSON, C. S., & KEWLEY-PORT, D. (1988). Some remarks on Pastore (1988). *Journal of the Acoustical Society of America*, **84**, 2266-2270.
- WOODWORTH, R. S., & SCHLOSBERG, H. (1954). *Experimental psychology*. New York: Holt.
- ZERA, J., & GREEN, D. M. (1993). Detecting temporal onset and offset asynchrony in multicomponent complexes. *Journal of the Acoustical Society of America*, **93**, 1038-1052.

NOTES

1. The 15–30 msec temporal order threshold (TOT) can be contrasted with a much longer threshold (several hundred milliseconds)

for correctly assigning the order labels to a sequence of four or more stimuli repeated in sequence (Warren, 1982), where the limit is related to identifying and storing a label for each stimulus element in the order presented without interference from the repeated sequence.

2. The DL here is based on the stimulus difference that yields $d' = 1.0$ in a 2IFC task, or approximately 69% correct. The equivalent conditions should yield lower, but above chance, performance in an ABX task. In order to have a reasonable model for discrimination, we have selected a somewhat larger physical stimulus difference than is typical of categorical perception research for voicing contrasts; if a 12-msec difference yields performance a bit above chance, then a 10-msec difference should better approximate chance.

(Manuscript received February 17, 1995;
revision accepted for publication August 29, 1995.)