Temporal order judgment of auditory stimulus offset

RICHARD E. PASTORE
State University of New York, Binghamton, New York

The temporal order judgment threshold for the offset asynchronies of pairs of tones at least 10 msec in duration was found to follow a pattern of parameter dependencies which is similar to that found for the onset of acoustic stimuli. Both thresholds increase with increased stimulus duration, but both appear not to depend upon the intensity or frequency of the component stimuli. Temporal order judgment thresholds for offset asynchronies appear to be briefer than onset asynchronies, probably due to the availability of some form of echoic information in the offset condition. Implications for models of temporal order perception and for the perceptions of temporally contrasted CV and VC syllables are discussed.

Investigations of the limits on temporal order judgments (TOJ) of auditory stimuli have focused on asynchronies of stimulus onset, ignoring TOJ of stimulus offset asynchronies. It is of theoretical importance to be able to compare TOJ for onset and for offset of similar stimuli. While similarities in onset and offset conditions might be indicative of a single process, or a common set of processes, underlying the perceptions of temporal order, differences could help us understand the possible nature of the processes involved in order perception. The current research focuses on TOJ threshold for offset asynchronies.

Temporal Order Onset Perception

Hirsh (1959) studied the threshold times for the identification of order of stimulus onset for pairs of stimuli with both stimuli being tones or with one being a tone and the other a brief noise burst. Hirsh reported that the TOJ threshold was approximately 17 msec and that this threshold was independent of the nature and intensity of the stimuli. In later papers, Hirsh (1974; Julesz & Hirsh, 1972) noted that there were a number of different types of TOJ phenomena. These phenomena include the type of stimuli that were studied by Hirsh (1959) (which are the focus of the present research) plus a number of other phenomena, including lateralization associated with brief interaural temporal (phase) differences, and the minimum duration of the components of complex, repeated stimuli whose order must be identified (Warren, Obusek, Farmer, & Warren, 1969). Another related phenomenon is the perception of the order differ-

ences of brief (less than 10 msec) stimuli, with which the threshold has been found to be dependent upon relative intensity, duration, and spectral composition (Patterson & Green, 1970). Each of these phenomena is characterized by different patterns of parameter interactions and by different magnitudes of thresholds.

The Hirsh-type TOJ limitation has been conjectured to be one possible basis for voice-onset-time (VOT) contrasts among speech stimuli (J. D. Miller, Wier, Pastore, Kelly, & Dooling, 1976). Like a VOT continuum, a temporal order continuum exhibits categorical perception with approximately a 20-msec boundary (Pisoni, 1977) and exhibits selective adaptation (Pisoni, 1980). One problem with this analysis is the absence of evidence for cross-adaptation between temporal order continua and VOT continua (Pisoni, 1980). A second problem is the finding that the VOT boundary is not fixed, but depends very much upon the specific stop consonant (Lisker & Abramson, 1967), the specific vowel context (Lisker, Liberman, Erickson, & Dechovitz, 1978; Summerfield & Haggard, 1977), and the rate of articulating other contextual stimuli (Summerfield, 1981). A somewhat similar pattern of temporal parameter interactions has been reported for the voicing boundary of final position consonants (Raphael, 1972; Summerfield, cited in J. L. Miller, 1981). Recently, we (Pastore, Harris, & Kaplan, 1982) demonstrated that the threshold for the temporal order identification of the onset asynchrony of tone pairs also is not fixed, but rather is a direct function of the duration of the tones and, at least for rise times greater than 10 msec, also of rise time. Replicating Hirsh (1959), we found an approximately 17-msec TOJ threshold for 300-msec stimuli and no effects of tone frequency. These findings improve the correlation between TOJ phenomena and the perception of initial stop consonants with voicing contrasts.

Models of Temporal Order Perception

A number of different models have been proposed to account for the limits on temporal order judgments. Most of these models can be conceptualized as variants of specific models summarized by Sternberg and Knoll (1973) within their broadly defined general theoretical framework. Sternberg and Knoll proposed a multistage framework which begins with transmission processes feeding a central processing mechanism whose output is subjected to a decision-type analysis. In the transmission stage, the information is transduced from its external representation to a neural representation and then is conveyed or transmitted to the central timing mechanism. Sternberg and Knoll limited their concern to those situations in which the information about the two critical stimuli is transmitted over independent sensory channels. At this point, the various models begin to differ. Some models assume differences in channel transmission rates; for instance, Rutschmann (1973) conjectured that the central timing mechanism was accurate, with TOJ being limited by the variability of sensory arrival times. Other models also begin with the transmission difference, but assume further limitations with the central timing mechanism; for instance, some models assume a sampling of information by the timing mechanism in terms of discrete perceptual "moments" (e.g., Stroud, 1955). Kristofferson (1970) proposed an attention-switching model in which the major constraints on temporal order perception are the temporal requirements necessary to attend to the signal in one channel and then switch attention or processing to the other channel. There are a number of variants of these and other models.

Our earlier research on the temporal order identification of onset asynchrony can be examined in the context of this general outline. The first question concerns whether our stimuli were processed through separate input channels. Channel independence is a basic assumption of the temporal judgment models summarized above. The question of independent or separable channels is very difficult to answer, but must be a major concern for any model or theory attempting to employ the concept of attention. In an earlier study, we demonstrated that, for the detection of tones, frequency separation by several critical bandwidths provides a better approximation to independent auditory input channels than does presentation to separate ears (Puleo & Pastore, 1978). If one can generalize from the detection of tones to the identification of the order of perfectly detectable signals, then the separation of tones by several critical bandwidths of frequency should meet the assumption of channel independence. However, it has yet to be demonstrated that one can generalize the definition of channel independence across such tasks, and, in fact, there is some evidence that an empirically based definition of channel independence for an intensity dimension may not be adequate for a temporal dimension (Eijkman & Vendrik, 1965). The finding (Hirsh, 1959) that TOJ is independent of the nature of the component sounds may seem to imply channel independence, but this probably is not the case. In our previous study (Pastore et al., 1982), we found a direct dependency of TOJ on duration which was orthogonal to our replication of Hirsh's finding of spectral independence for pairs of tones separated by a broad sampling of frequency differences ranging from several critical bands to considerably less than one critical band. We must conclude that either none of the frequency differences we employed, or all of these differences, resulted in channel independence for TOJ (we suspect the former to be the case, but will return to this topic later). In conclusion, we do not know if we have met the assumption of these models of transmission channel independence, and thus we must continue to entertain the possibility that our stimuli are interacting prior to any central timing mechanism or decision process.

TOJ for Onset

For the moment, let us assume channel independence. Our previous study with onset asynchronies found a direct relationship between TOJ and both stimulus duration (10 $\leqslant$ t $\leqslant$ 300 msec) and rise time (for rise time > 10 msec, but not for rise time < 10 msec). One might expect differences in transmission rate as a function of rise time. For instance, changes in rise time may result in differences in the perception of stimulus onset time (e.g., see Vos & Rasch, 1981).[1] Thus, the rise-time differences are consistent with most of the models of temporal order perception. The duration effects are more difficult to reconcile with the models.

There is no reason to conjecture that the differences in duration after onset should directly affect the transmission rate of onset information. It is possible that transmission rate is a function of total energy integrated across stimulus duration. However, Hirsh (1959) found no effect of the relative intensity of the tones used in his TOJ task (for a 20-phon range). He also found that it did not matter whether he created the onset asynchrony with two tones having simultaneous offsets (thus two tones with equal duration except for the onset asynchrony) or with one of the stimuli being a very brief burst of noise (thus one stimulus being considerably shorter than the other stimulus). Based upon Hirsh's findings, we conclude that neither the relative instantaneous intensity of the component stimuli nor the relative intensity integrated across given stimulus durations significantly alters the transmission rate within, or the attention switching times between, input channels. If, however, we are not dealing with independent input channels, then the various stage analyses may be inappropriate for modeling this specific type of task behavior.

We also can draw some positive conclusions from our earlier finding that overall stimulus duration influences the TOJ threshold for durations of 10, 30, 100, and 300 msec. We do not know whether this relationship between TOJ and overall duration asymptotes between 100 and 300 msec or extends beyond 300 msec. We can conclude only that the decision concerning order of stimulus onset must be delayed by at least 100 msec (and possibly 300 msec or longer) following the onset of the second stimulus component. That decision also must involve more than simply the monitoring of differences at stimulus onset. It would appear that the subjects are responding to either the whole stimulus or the first t msec of the stimulus (where t > at least 100 msec). It is well known that the auditory system integrates stimulus information over a period of approximately 200 msec (Zwislocki, 1960), and there is evidence that an auditory stimulus trace lasts at least 200 msec (Massaro, 1972, 1973; Plomp, 1964). One implication of these findings is that subjects are using some kind of information storage in performing the task. Subjects may be responding to some spectral characteristics of the stimuli integrated over the first t msec of the stimuli. Alternatively, subjects may be responding to a stored representation of the onset characteristics of the stimuli, with the decision stage being influenced by the later characteristics of the stimuli. In the last case, and in most specific implementations of the former, we would not expect to find a similar dependency upon duration for the identification of stimulus offset. In TOJ for stimulus offset, the critical stimulus information is followed by silence, which, unlike the tones defining the stimuli in the judgment of onset, should have no masking or interaction effects on the perception of the stimuli.

Summary

The present experiments concern the discrimination of stimulus offset. The first three experiments are analogous to our previously reported onset-discrimination experiments. In that research, TOJ threshold increased with increasing stimulus duration, an apparent violation of the independence assumption of the Steinberg and Kroll theoretical framework. Offset discrimination offers a more direct test of this assumption, since the critical portion of the stimulus is not subject to backward masking. To look ahead, the first experiment reveals the same dependency on duration for offsets as for onsets, but with smaller thresholds. This violation of the channel independence assumption might be due to the fact that the two stimuli were similar in frequency and equal in intensity; Experiments 2 and 3 investigate these possibilities by varying relative intensity and frequency. The last experiment investigates the possibility that the smaller thresholds for offset discrimination relative to onset discrimination may be due

to the absence of backward masking effects for offset, but not for onset, conditions.

EXPERIMENT 1: BASE DURATION

The threshold for the judgment of onset order is a direct function of the base duration (the duration of the stimuli after the onset of the second stimulus). The first experiment investigates the analogous condition for stimulus offset, mapping TOJ as a function of base duration (the duration of the stimuli before the offset of the first stimulus). We selected tones that are typical of the center frequencies of the second and third formants in speech, although no phonetic contrast depends explicitly on the order of offset of these formants.[2]

Method

Subjects. The three subjects (two female and one male) were undergraduates with no known hearing deficits. The subjects were paid for participating in the study. The two female subjects were experienced psychophysical listeners. The subjects were run in a commercial sound chamber.

Stimuli. All stimuli were tones generated by Wavetek 132 function generators, gated with .5-msec rise-fall times and bandpass-filtered (1,000-3,150 Hz) through Ithaco 4302 filters. The stimuli were segments of 1,800- and 2,150-Hz tones which were presented monaurally (left ear for one subject and right ear for the other two subjects) at 70 dB SPL through Telephonic TDH-39(300) earphones. The base duration of the stimuli was always fixed during a block of trials and was 10, 30, 100, and 300 msec. Each base duration was equal to a 3,500-msec interstimulus interval plus both the base duration of the stimuli and the offset asynchrony.

Procedure. We employed a Levitt up-down adaptive procedure with a 2:1 rule (see Levitt, 1971) and a 2-msec step size for adjusting temporal differences in offset. This procedure estimates the 70.7% threshold. The order of tone offset was randomly determined for each trial, with tone onsets always being simultaneous. The subjects were run in blocks of 120 trials. To complete a condition, a subject's data for two consecutive blocks of trials had to meet the following criteria, as suggested by Levitt (1971): the total range of means (for the two sequences per block of 120 trials) and the standard deviations (within each sequence) all had to be less than 2.5 times the step size.

Results and Discussion

The results are summarized in Figure 1. Plotted also are the results from the analogous onset condition (with four different, but equally experienced, subjects and 1,650-2,350-Hz tones). The difference in stimulus frequency is probably not an important parameter (see both Experiment 2 below for offset conditions and Pastore et al., 1982, for onset conditions). In both the initial and final position of the stimuli, the threshold for temporal judgment is a direct function of the duration of the stimuli. Thus, the TOJ for both onset and offset asynchronies seems to involve a similar type of interaction with base duration.

In the final position, the judgment of order appears to be better than in the initial position by a factor of almost two. This superiority of final position TOJ is too large and consistent to dismiss on the basis of uncontrolled factors such as the use of dif-
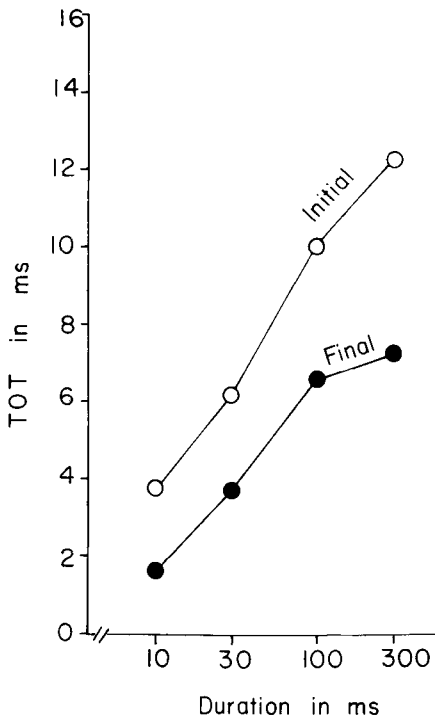
Figure 1. Temporal order identification threshold as a function of stimulus duration for initial (temporal onset) and final (temporal offset) stimulus position.

ferent subjects. Rather, we believe that there is a significant difference between TOJ for onset and TOJ for offset. That difference could be the presence of a masking stimulus (the tones which define the task) following the onset asynchrony but not following the offset asynchrony. Experiment 4 tests this possibility.

EXPERIMENT 2: FREQUENCY SEPARATION

Temporal order judgment research with stimuli at least 10 msec in duration has reported thresholds which are independent of the frequency of the stimulus components (Hirsh, 1959; Patterson & Green, 1970; Sternberg & Knoll, 1973). This previous research

is based upon judgments of the order of onset asynchrony. Experiment 2 investigates the generalization of such frequency independence to the offset asynchrony condition with the four base durations of the previous experiment. We will use frequency separations ranging from a minimum of 350 Hz, which is less than the critical ratio estimate of the critical band at 1,900 Hz (Hawkins & Stevens, 1950), but is slightly greater than the direct estimate of the critical band (Zwicker, Flottorp, & Stevens, 1957), to a maximum of 1,750 Hz. If frequency separation is important in defining input channel independence, then, based upon the findings of Puleo and Pastore (1978) for detection, we would expect a separation of less than 2-3 critical bands, that is, less than 800 to 1,000 Hz, to reflect channel interaction and wider separation to reflect channel independence, as assumed by the models discussed above.

Method
The subjects and procedures were identical to the first experiment. In this experiment, we used a number of different tones to define the offset asynchrony. The tones, and resulting frequency separations of the tones, are summarized in the left column of Table 1. All data for a given tone pair were collected before changing to the next tone pairs.

Results and Discussion
The mean TOJs (and standard deviation), averaged across subjects, are summarized in Table 1. These results replicate the results of the first experiment, with a significant effect of base duration and with TOJ magnitudes being comparable to those found in Experiment 1. The results failed to demonstrate any significant effects of tone frequency or frequency difference. Therefore, TOJ for offset, like TOJ for onset, would seem to be independent of tone frequency (at least for the frequencies examined). Either all conditions represent input channel independence, as required by the Sternberg and Knoll conceptual framework, or all conditions involve a similar interaction between the tones defining the offset asynchrony (and thus an absence of channel independence).

The magnitude of the standard deviations increased with increasing base durations. All subjects exhibited

Table 1
TOT as a Function of the Frequency and Duration of the Stimuli

| Frequency (in Hertz) | | | Base Duration of Two Stimuli | | | | | | | |
| | | | 10 msec | | 30 msec | | 100 msec | | 300 msec | |
Two Stimuli		Difference	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1800	2150	350	1.71	.07	3.75	.36	6.54	2.40	7.33	2.30
1600	2350	750	1.60	.03	3.18	.78	6.67	2.00	9.56	5.47
1600	2850	1250	1.98	.31	4.55	1.91	7.37	1.76	7.84	3.98
1100	2350	1250	1.91	.13	4.02	.80	7.14	2.51	9.87	2.31
1100	2850	1750	1.78	.11	4.55	1.23	8.65	4.13	8.42	3.86

nearly identical means for the 10-msec base duration condition, and all exhibited monotonically increasing mean thresholds as a function of increased base duration. However, the slopes of the monotonic function differed across subjects, thus producing the increased variability with increased base duration. It would appear that the cues used by the subjects in performing the task are much more salient, and much less subject to individual differences, with the shorter base durations.

EXPERIMENT 3: RELATIVE INTENSITY

Hirsh (1959) reported that TOJ for onset was independent of the intensity of the component stimuli, at least for a 20-phon range of relative loudness. This experiment examines a similar condition for the temporal order judgment of offset. If the information transmission models discussed above are to be considered reasonable, then one would expect that relative stimulus intensity would be an important parameter, since intensity is expected to affect neural transmission speed. Furthermore, with TOJ of offset, the transmission of information would be in the absence of a backward masking stimulus. Thus, one might expect to find a dependency on intensity with offset asynchronies even when such a dependency is absent with onset asynchrony.

Method

The subjects and procedure again were identical to those of Experiment 1. In this experiment, we used the 1,600-2,350-Hz tone pair to define the task. We created five intensity conditions by keeping one of the tones at 70 dB and varying the other tone in 10-dB steps from 50 to 70 dB (see Table 2). For each intensity condition, we ran the four base duration conditions before changing to the next intensity condition.

Results and Discussion

The trend in the data for all subjects was very similar and is reflected in the overall means summarized in Table 2. There again is a major effect for base duration, but there is no effect of relative intensity. We would expect any changes in stimulus intensity to alter both the mean TOJ threshold and the variability within

adaptive up-down sequences. Neither statistic was altered in a systematic fashion. As with TOJ for onset, we must again conclude either that intensity does not alter information transmission rate as proposed in some models of order perception or that some other assumption of the models (such as channel independence) is not valid.

EXPERIMENT 4: ECHOIC MEMORY

One major difference between TOJ for onset and for offset is the possible presence of a backward masking effect for onset, but not offset, asynchronies. Both our current and previous research with TOJ indicates that TOJ involves a delayed decision, and thus probably involves some form of short-term storage of the information concerning the temporal asynchronies. Such short-term storage of information could be defined as a type of echoic memory and probably is a type of preanalyzed or "precategorical" acoustic storage (Massaro, 1972, 1973). In TOJ for stimulus offset, such echoic information exists during a period of silence and thus is available without interference. In TOJ for stimulus onset, this echoic store must compete with the continuing stimuli. With a reduced quality of an echoic representation of asynchrony information with TOJ for onset, the subject would require a greater temporal asynchrony to perform the task. If this analysis is correct, then the addition of a masking stimulus during the otherwise silent interval following the offset asynchrony should be expected to interfere with the memory trace or echoic storage. With less echoic information available, the subjects should require a greater actual stimulus difference to perform the task. This should result in increased thresholds for TOJ for offset. We further would expect that the backward masking of TOJ for offset should decrease with increasing processing time as defined by increasing delays between final signal offset and masker onset. Based upon previous research on echoic memory (Massaro, 1972, 1973), we would expect the echoic information to be available for no more than 200 to 250 msec. Therefore, a backward masker delayed by more than

Table 2
Average TOJ Thresholds for Offset With Tones of Varying Intensity

| Intensity (in Decibels) | | | Base Duration (in Milliseconds) | | | | | | | |
| | | | 10 | | 30 | | 100 | | 300 | |
1600	2350	Difference	Mean	SD	Mean	SD	Mean	SD	Mean	SD
50	70	20	2.2	.7	4.8	2.1	5.8	1.2	8.6	3.3
70	50	20	2.0	.2	4.6	1.0	7.1	3.5	11.4	5.9
60	70	10	1.7	.1	3.1	.6	7.1	2.2	10.4	4.2
70	60	10	1.8	.3	3.7	.5	7.1	2.5	8.3	4.1
70	70	0	1.7	.1	3.7	.4	6.5	2.4	7.3	2.3

Note—Average between-subjects standard deviation within each block of trials run by each subject under the given base duration = .5, 1.2, 2.2, and 2.6 for the 10-, 30-, 100-, and 300-msec durations, respectively.

250 msec should not interfere with echoic memory. Only at brief masker delays, at which the masker interferes with echoic store actually used in performing the task, should a backward masking stimulus alter TOJ for offset. Experiment 4 tests this prediction.

Method

Three subjects were run in this study.[3] To varying degress, all were practiced psychophysical observers. The procedures were the same as in Experiment 1, except that the stimuli were segments of 1,500- and 2,000-Hz, 70-dB tones. Each subject ran a block of 100 trials with no masker, followed by a similar block of trials with a 50-msec segment of an 80-dB filtered (1,000-3,150-Hz) Gaussian noise beginning 1 msec after the offset of the longer tone. The noise and the tones had .5-msec rise-fall times. Once a subject completed five blocks of the nonmasked/masked conditions for the 10-msec base duration, the subject ran the 30-msec, then the 100-msec, and finally the 300-msec base duration condition. Based upon the finding of significant differences between the nonmasked and masked conditions, the procedure was repeated with a pair of masker conditions; the masker was delayed by 50 and by 300 msec. The 300-msec condition was selected to exceed the assumed duration of echoic storage for simple acoustic stimuli.

Results and Discussion

The results are summarized in Table 3. The standard deviation for each TOJ threshold estimate ranged from less than .1 to 2.2 and were roughly in proportion to the magnitude of the threshold. We again have a main effect of base duration. In addition, the masker at a 1-msec delay interferes with TOJ for all base durations. In fact, the masker at 1 msec nearly doubles the threshold values (an average multiplicative change of 1.8). We had conjectured that TOJ both for offset with this minimally delayed masker and for onset represent two conditions without the availability of an echoic store of the stimulus asynchrony. Both of these masked conditions resulted in data which are nearly double those for the unmasked conditions for offset which we assume can be performed using echoically stored information.

With a masker delay of 50 msec, TOJ for offset still is slightly elevated relative to the no-mask conditions, while at 300 msec, TOJ equals the no-mask conditions. With the 50-msec delay, the subject would have considerable echoic information available, and therefore does not require a significant increase in the physical stimulus difference to perform the task. At 300 msec, either the echoic information has faded

or the processes required for the judgment of order have been completed; in either case, the masker no longer has any effect.

GENERAL DISCUSSION

The results of the present set of offset asynchrony experiments, interpreted in the context of the results from onset asynchrony experiments, allow us to eliminate some accounts of the nature of auditory temporal order judgments. First, it is clear that subjects are not making temporal order judgments based strictly and directly upon the nature of the temporal asynchrony. TOJ threshold for both onset asynchronies and offset asynchronies is a direct function of total stimulus duration. The temporal asynchrony should have little relationship to stimulus duration, yet we clearly are dealing with some kind of integration or smearing of stimulus information in time over the stimulus duration. This dependency of TOJ upon duration and the independence of TOJ on frequency lead us to doubt that we are dealing with the transmission of stimulus information through independent channels. This was a basic premise in the Sternberg-Knoll theoretical framework.

What, then, is the nature of such stimulus integration? One possibility is that the integration of a longer stimulus results in greater total stimulus energy which produces greater nonsimultaneous masking of the asynchrony. Integration models can explain many nonsimultaneous masking findings (Penner, 1979, 1980). In the present study, the results from Experiment 4 would lend support to the notion that TOJ for onset does involve some form of backward masking. Thus, we might conjecture that longer, and therefore more intense, stimuli produce greater backward masking which results in a longer TOJ threshold. By analogy, we could argue that TOJ for offset involves some degree of forward masking between the stimuli and the asynchronous offset. However, the stimulus parameters are not fully consistent with such an interpretation as a primary explanation for the duration effect on TOJ. Assuming perfect integration of stimulus energy in time, our stimuli represent a total change of 14.8 dB between the 10- and 300-msec conditions. One would expect that an equivalent (or pos-

Table 3
TOJ Thresholds for No Masking (NM) and Various Masker Delay Conditions

Masker Delay	10-msec Base				30-msec Base				100-msec Base				300-msec Base			
	NM	1	50	300	NM	1	50	300	NM	1	50	300	NM	1	50	300
Observation 1	1.7	3.0	1.6	1.2	2.7	5.2	2.8	1.8	2.7	6.2	4.4	3.1	2.6	4.6	3.2	2.0
Observation 2	1.9	4.1	2.2	1.3	3.4	4.5	4.2	3.1	5.5	7.3	6.6	5.1	7.3	12.4	7.0	5.9
Observation 3	2.2	5.9	7.9	2.0	6.4	11.2	12.1	8.5	10.4	17.1	13.5	7.6	12.5	14.3	12.6	9.5
Mean	1.9	4.3	3.9	1.5	4.2	7.0	6.4	4.1	6.2	10.2	8.2	5.3	7.5	10.4	7.6	5.8

sibly an even greater) change in threshold data would be produced by a direct change in stimulus (and therefore masker) intensity, yet a decrease in the intensity of either stimulus by 20 dB does not change TOJ thresholds for onset (Hirsh, 1959; Pastore, Harris, & Kaplan, 1982) or for offset (Experiment 3). These same findings provide a basis for an argument against other models based upon integrated intensity differences. For instance, for stimulus offset asynchronies one might conjecture that longer (therefore more intense) stimuli require greater decay times than shorter (less intense) stimuli, with subjects responding on the basis of the residual sensation following the decay of one of the stimuli to a subliminal magnitude. In an alternative model described in the introduction, difference in total stimulus magnitude (including the asynchronous portions of the stimuli) might produce differences in the neural transmission times which are the basis for the temporal order judgments. Such models would have to predict analogous results due to changes in stimulus intensity, yet we find no such changes in thresholds.

In the introduction we conjectured that subjects might be responding to a stored representation of the asynchronous characteristics of the stimuli, with the evaluation of the stored information being judged relative to later-occurring stimulus information. Experiment 4 supports the conjecture that the subjects are using some form of stored stimulus information or echoic memory. Our conjecture, however, was an attempt to account for the dependency of TOJ for onset on stimulus duration. Since we find a similar dependency with TOJ for stimulus offset, the critical variable probably cannot be the nature of the stimulus present at the time the response decision is made.

Despite abandoning this delayed comparison hypothesis, our best conjecture still is that subjects are making a relative judgment in performing the TOJ task. Our problem then is to specify the nature of the relative judgment. We already have examined the relationship between stimulus duration and effective stimulus intensity. Another possibility is the relationship between duration and bandwidth. The effective bandwidth of a time-limited segment of a tone is described in terms of a sine [or (sine wjt)/wjt] function and thus is an inverse function of the stimulus duration. Longer segments of tones have narrower spectra. Unfortunately, hypotheses based upon this type of analysis quickly break down for several reasons. First, given constant center frequencies, the longer duration (narrower bandwidth) stimuli should be more discriminable from each other than shorter duration (broader bandwidth) tones, yet longer duration stimuli yield larger TOJ thresholds. Also, TOJ for both onset (Hirsh, 1959; Pastore et al., 1982) and offset (Experiment 2) was found to be independent of the frequency of the stimuli defining the task, at least

for the broad range of stimulus parameters investigated. We still suspect that subjects are performing the task based upon judgments of the perceptual quality of the temporal asynchrony. This, in effect, would be a spectral judgment rather than a direct temporal judgment. Our subjects all claim that they were listening to differences in the quality, rather than the duration, of the stimulus. However, we are at a loss to specify the nature of the spectral differences which might serve as the basis for such judgments.

Speech Perception

The current findings have some direct implications for speech perception. Raphael (1972) demonstrated that changes in vowel duration altered the voicing boundary of final-position consonants. Raphael produced voiceless analogues to final-position voiced consonants in CVC context by cutting back the offset of the final F1 transition; increasing the magnitude of the F1 cutback increased the tendency to hear the final consonant as voiceless. For a given F1 cutback, shortening the vowel duration also increased the tendency to perceive the final consonant as voiceless. In effect, Raphael found a trading relationship between F1 cutback and vowel duration.

Our results provide the basis of an acoustic explanation for the Raphael findings. It has been suggested that voicing contrast may be mediated by a TOJ threshold (Miller, Wier, Pastore, Kelly, & Dooling, 1976; Pastore, 1976; Pisoni, 1977). For instance, if an initial consonant is perceived as having a delayed onset for F1, the consonant is recognized as being voiceless. If the order of onset of the formant transitions cannot be identified, the stimulus is perceived as being voiced. This analysis, if valid, should apply equally well to final consonants. With short-duration vowels, the TOJ threshold is short; a consonant with a small F1 cutback that exceeds this threshold would be perceived as being voiceless. Lengthening the vowel duration would increase the TOJ threshold; a consonant with an F1 cutback that had just exceeded the TOJ threshold in the context of the short-duration vowel now would be subliminal (because of the shifted threshold) and thus would be perceived as being voiced. Furthermore, since continuous changes in vowel duration produce continuous changes in TOJ threshold, the voicing boundary should be a continuous function of vowel duration; Raphael reports just such a relationship. When vowel duration is held constant, TOJ threshold, and therefore the voicing boundary, is stationary and should be categorical for changes in VOT (Pastore, Ahroon, Baffuto, Friedman, Puleo, & Fink, 1976, provide a discussion of the relationship between thresholds and the finding of categorical perception.)

One possible criticism of this analysis is that the TOJ thresholds reported in this paper for stimulus

offset and reported earlier for stimulus onset (Pastore et al., 1982) are considerably shorter than the voicing boundaries reported for speech stimuli. We believe that these discrepancies can be attributed to differences in procedures. In our studies, we employed highly practiced subjects who were very familiar with the task and the use of sinusoidal stimuli. Whenever we have started new subjects on a TOJ task, we always have found TOJ thresholds to be a direct function of duration, but the initial thresholds always are on the order of 100 msec or more. After some experience with this, or with other psychophysical tasks involving sinusoidal stimuli, we find that the TOJ thresholds are reduced to approximately 20-40 msec. It is only after reasonably extensive practice that one finds thresholds of the magnitude reported in this paper. Most speech perception studies involve the use of subjects who are not familiar with the task but who have had some experience with the given stimulus continuum. We would expect their effective TOJ threshold to be greater than the thresholds reported in this paper, but less than those for our most naive subjects. Furthermore, we would expect practice to result in a change in the location of the category boundary for speech stimuli as well as the discriminability between stimuli within a perceptual category. Such findings have been reported by Carney, Widen, and Viemeister (1977) and Samuel (1977). Therefore, our TOJ threshold findings are not inconsistent with the nature and magnitude of the boundaries for voicing constructs in speech.

In Experiment 4, we found that the addition of a noise stimulus immediately following the stimuli produced a masking effect as measured by an increase in TOJ threshold. We find a direct analogy to this condition in the Raphael (1972) study. While some of his stimulus continua were based upon CVC syllables, other stimulus continua had noise added at the end of the syllable to create either a final fricative or a final consonant cluster. Generalizing from our findings, we would predict that the addition of such noise should increase the TOJ threshold, thus increasing the voiced perception of stimuli. Comparing conditions with the same final stop, but differing in terms of the presence of an added fricative, we note that the category boundary is at a shorter vowel duration when the fricative is present. A shorter vowel duration would correspond to a shorter TOJ threshold which would be required to compensate for the increased TOJ threshold due to the masking effects of the added noise. Therefore, our findings provide the basis for one possible explanation of these results reported by Raphael.

While we do not believe that TOJ thresholds provide anything like a complete explanation of voicing contrasts for speech, we do believe that they may provide a possible acoustically based explanation for many of the influences on voicing contrasts. We feel that there are sufficient similarities between VOT and TOJ phenomena to warrant further investigation. For instance, Summerfield (1981) reports that initial stop consonants differ in VOT boundary as a function of place of articulation. His stimuli had a complex onset envelope, rising approximately 60 dB during the first 2 msec, and then rising another 15 dB over the remainder of the transition duration. The overall rise-time therefore is approximately equal to the transition duration. He reports that velar consonants, with a 50-msec rise time, have the longest VOT boundaries, while labial consonants, with 30-msec rise times, have the shortest VOT boundaries. In Pastore et al. (1982), we found that the TOJ threshold for onset asynchronies was a direct function of stimulus rise time. Might the differences in categorical VOT boundary as a function of place of articulation be mediated by the changes in TOJ threshold caused by the stimulus rise-time differences correlated with changes in place of articulation? This type of analysis may allow us to better understand the similarities and the differences in the perception of such acoustic and speech stimuli.

REFERENCES

CARNEY, A. E., WIDIN, B., & VIEMEISTER, N. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, **62**, 961-970.

EIJKMAN, E., & VENDRICK, A. J. H. Can a sensory system be specified by its internal noise. *Journal of the Acoustical Society of America*, 1965, **37**, 1102-1109.

HAWKINS, J. E., JR., & STEVENS, S. S. The masking of pure tones and of speech by white noise. *Journal of the Acoustical Society of America*, 1950, **22**, 6-13.

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

HIRSH, I. J. Temporal order and auditory perception. In H. R. Moskowitz (Ed.), *Sensation and measurement*. Boston: Reidel, 1974.

JULESZ, B., & HIRSH I. J. Visual and auditory perception—An essay of comparison. In E. E. David & P. N. Denes (Eds.), *Human communication: A unified view*. New York: McGraw-Hill, 1972.

KRISTOFFERSON, A. B. Attention. In R. M. Patton, T. A. Tanner, J. J. Markowitz, & J. A. Swets (Eds.), *Applications of research on human decision making*. Washington, D.C: NASA, 1970.

LEVITT, H. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 1971, **49**, 467-477.

LISKER, L., & ABRAMSON, A. S. The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Science* (Prague), 1967, 563-567.

LISKER, L., LIBERMAN, A. M., ERICKSON, D., & DECHOVITZ, D. On pushing the voice-onset-time boundary about. *Language & Speech*, 1978, **20**, 209-216.

MASSARO, D. W. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 1972, **79**, 124-145.

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

MILLER, J. D., WIER, C. C., PASTORE, R. E., KELLY, W. J.,

& DOOLING, R. J. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, 60, 410-417.

MILLER, J. L. Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, N.J: Erlbaum, 1981.

PASTORE, R. E. Possible psychoacoustic factors in speech perception. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives in the study of speech*. Hillsdale, N.J: Erlbaum, 1976.

PASTORE, R. E., AHROON, W. A., BUFFUTO, K. J., FRIEDMAN, C., PULEO, J. S., & FINK, E. A. Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception & Performance*, 1976, 3, 686-696.

PASTORE, R. E., HARRIS, L. B., & KAPLAN, J. K. Temporal order identification: Some parameter dependencies. *Journal of the Acoustical Society of America*, 1982, 71, 430-436.

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

PENNER, M. J. The ear's two temporal integrators, operating in series. *Journal of the Acoustical Society of America*, 1979, Suppl. 1, 66, S7-8.

PENNER, M. J. The coding of intensity and the interaction of forward and backward masking. *Journal of the Acoustical Society of America*, 1980, 67, 608-616.

PLOMP, R. Rate of decay of auditory sensation. *Journal of the Acoustical Society of America*, 1964, 36, 277-282.

PISONI, D. B. 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*, 1977, 61, 1352-1361.

PISONI, D. B. Adaptation of the relative onset time of two-component tones. *Perception & Psychophysics*, 1980, 28, 337-346.

PULEO, J. S., & PASTORE, R. E. Critical-band effects in two-channel auditory signal detection. *Journal of Experimental Psychology: Human Perception & Performance*, 1978, 4, 153-163.

RAPHAEL, L. J. Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. *Journal of the Acoustical Society of America*, 1972, 51, 1296-1303.

RUTSCHMANN, R. Visual perception of temporal order. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973.

SAMUEL, A. G. The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics*, 1977, 22, 321-330.

STERNBERG, S., & KNOLL, R. L. The perception of temporal order. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973.

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

SUMMERFIELD, Q. Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception & Performance*, 1981, 5, 1074-1095.

SUMMERFIELD, Q., & HAGGARD, M. P. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 1977, 62, 435-448.

VOS, J., & RASCH, R. The perceptual onset of musical tones. *Perception & Psychophysics*, 1981, 29, 323-335.

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

ZWICKER, E., FLOTTORP, G., & STEVENS, S. S. Critical band width in loudness summation. *Journal of the Acoustical Society of America*, 1957, 29, 406-415.

ZWISLOCKI, J. Theory of auditory summation. *Journal of the Acoustical Society of America*, 1960, 32, 1046-1060.

NOTES

1. We varied the rise time within the stimulus envelope with the component stimuli having a rise time of .5 msec within the constraints of the envelope. Therefore, the later stimulus had a faster rise time, which could have produced a more rapid transmission.

2. When first designing this study, we conceptualized voicing contrast as being analogous to a temporal order judgment between the onset of the aspiration noise and the onset of voicing. In that context, the selection of frequencies roughly in the F2 and F3 frequency range made sense. It is just as likely that the appropriate analogy to TOJ is between F1 onset (F1 cutback) and F2 onset. In the case of voicing contrast for offset, the latter would seem to be the case (Raphael, 1972). Based upon pilot work with different frequencies, plus published data by Hirsh (1959) and Patterson and Green (1970), we believe that the use of a low-frequency tone (analogous to F1) would not have altered the results of any of our experiments.

3. The data from a fourth subject were omitted from Experiment 4. This subject often failed to show up for sessions, failed to complete the study, and reported having run for a period of time with ear discomfort due to a kernal of unpopped popcorn in her external auditory canal. Her data had been highly variable.