

Resetting the pitch-analysis system. 2. Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones

Albert S. Bregman and Pierre A. Ahad

Department of Psychology, McGill University, 1205 Doctor Penfield Avenue, Montreal, Quebec H3A 1B1, Canada

Jean Kim

School of Human Communication Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec H3G 1A8, Canada

(Received 24 July 1992; revised 24 January 1994; accepted 11 July 1994)

Experiments on young adults studied the effects of suddenness of onset or offset on the discrimination of the order of pitches of individual tones in a 1-s, 4-tone cluster of overlapping pure tones. In experiment 1, the tones, all within a critical band, went on asynchronously. Each rose and decayed linearly in amplitude. Faster onsets, within the range 10 to 640 ms as measured on the first tone, increased the accuracy of the discrimination of the order of onsets, but 10-ms onsets were slightly worse than 40-ms onsets in early sessions. Experiment 2 found similar effects for the abruptness of offsets of tones in clusters whose components came on synchronously but went off asynchronously. Onset order was very much easier to detect than offset order. The auditory system may use neural onset and offset responses to reset itself and carry out new analyses at frequency-by-amplitude points of sudden amplitude change, thereby contributing to auditory scene analysis.

PACS numbers: 43.66.Hg, 43.66.Lj, 43.66.Mk [HSC]

INTRODUCTION

This is the second of two groups of experiments that examine how the rate of onset of a tone affects its segregation from other concurrent sounds. The first was by Bregman *et al.* (1994). Both sets of experiments were derived from a phenomenon observed in our laboratory in the course of experimentation with software that could reverse the order of samples in a digital sound file. When a tone with a rapid rise and a slow decay was played forward, it sounded like a clear plucked-instrument tone. When played backward, however, it sounded less definite. Even more interesting was the observation that when a cluster of sudden-onset tones, highly overlapped in time, were played in the forward order, they sounded distinct and their order was clearly heard. However when the file was played from back to front (slow onsets and fast offsets), the tones blended together and the temporal order was unclear.

In everyday listening, the importance of this response to sudden onsets may be in contributing to auditory scene analysis (Bregman, 1990), more specifically to the segregation of concurrent sounds in a mixture. It may help the system decide whether a rise in the intensity of a subset of frequencies in the spectrum of a sound is evidence for (a) a mere change in the spectrum of an ongoing sound, or (b) a new sound entering the mixture. If the change is slow, the auditory system may count this as evidence for the modification of the vibration pattern of an ongoing sound and hear it as a change in timbre. If it is sudden, it may treat it as symptomatic of the onset of a new sound in the mixture. A sudden offset of only a part of a spectrum, while rarer than a sudden onset, might be used as evidence for the existence of

a separate sound-producing vibration that has suddenly been damped.

Many single sounds have frequency components whose intensity changes are not exactly synchronized. For example in a trumpet tone, an analysis by Risset and Mathews (1969) showed that different harmonics reached their peak intensities at different times. Why were they not heard as separate sounds? Examination of their spectral analysis reveals that there were two properties of the rising set of harmonics that prevented any from being heard out: (a) there was only about 30 ms between the times at which harmonics reached their peak intensities; (b) these peaks did not occur in two distinct clusters but were spread out over the 30-ms interval.

The present experiments studied the discrimination of the order of onset of overlapping tones at different asynchronies and rise times. The effects of asynchrony on masking have also been studied by Rasch (1978), who asked subjects to discriminate between two possible targets. Each was a sequence of two complex tones whose fundamentals were at 500 and 750 Hz; so the target sequence either ascended or descended by a musical fifth. Each tone was generated with its first ten harmonics, with a spectral envelope that fell off at 6 dB/oct, and then were low-pass filtered at 4 kHz. The two-tone target sequence was presented more or less synchronously with two presentations of a single lower pitch, which acted as a masker. There was much better performance when the target and masker tones were asynchronous. The phase relations in our experiments (described under Methods) were arranged differently than those of Rasch. The phases of his components were set by sine addition for the low masker and cosine addition for the high-tone targets. The phase differ-

ence between coinciding harmonics in co-occurring high targets and low maskers was 90 deg. Our experiment overlapped four pure tones. Our phase relations were arranged so that whenever the tones overlapped, they were aligned exactly in phase.

Bregman and Pinker (1978) and Dannenbring and Bregman (1978) also studied the effects of asynchrony in the perceptual isolation of a target tone. When the target was more synchronous with an accompanying tone or tones, it fused with them and was less susceptible to being captured sequentially by a preceding tone. As in Rasch's study, rise times were not varied.

Kubovy (1981) reported a number of experiments in which a chord of pure tones was presented. When they were all on at the same time and steady in intensity, there was no perceptual domination by the pitch of any one. However, when one of the tones was lowered in intensity for 100 ms and then suddenly restored to its original value, its pitch dominated the experience of the listener. Kubovy called the phenomenon the "the onset-segregation" effect (Kubovy, 1976) and attributed it to the fact that there are specific neural responses to the onset of a sound. It is clearly this onset-segregation effect that we observed in our overlapping clusters of tones. We studied how this effect varied with the time taken by the tone to change in intensity; Kubovy did not vary this parameter.

The only experiment that did examine the effects of suddenness of onset on asynchronously starting tones was that of Pastore *et al.* (1982). They studied the discrimination of onset order using quite a different stimulus than the clusters used in the present experiments. Two sine-wave tones, 1650 and 2350 Hz, each with 0.5-ms rise time, started asynchronously and were mixed. Then the mixture as a whole was subjected to a large-scale amplitude envelope that rose, held a steady state for some duration, and ended with a 0.5-ms shutoff. Rise times for this envelope varied from 0.5 to 100 ms. As a consequence, the onset envelope of the second tone always consisted of two legs, an abrupt (0.5 ms) rise, and then a second (usually slower) rise followed the shape (unspecified) of the rising large-scale envelope. Listeners were asked to discriminate the order of pitches, low-high or high-low, by reporting which tone (high or low) came first. They found that the lowest level of asynchrony at which discrimination was possible was strongly affected by the large-scale envelope. The asynchrony threshold increased by a factor of about 2 between 10- and 25-ms rise times, but showed no change between 0.5- and 10-ms rise times, for the tone pairs that were most similar to ours in duration. There were two main differences between their experiment and the ones we will report, apart from the shape of the tone bursts. First, in theirs the subjects could make the discrimination by focusing on the pitch of the first onset, which, prior to the onset of tone 2, was present alone. At any given onset asynchrony, the faster the rise of the large-scale envelope, the greater would be the intensity of the first tone before being joined by the second. In our stimuli, however, the pair of tones whose order was to be discriminated were preceded and followed, in the cluster, by two bracketing tones. Thus, the critical tone was never presented first or last and was never more intense

than any of the tones that had preceded it. Therefore our studies determine whether the advantage of sudden onsets requires that a critical tone be the most intense one at some particular instant.

A second difference was that in the experiment by Pastore *et al.*, 1982 the tones were over two critical bands apart. One interpretation of their experiment could be as follows. In the low levels of the auditory system, information in different critical bands is separately represented; however, at higher levels, processes of auditory scene analysis (Bregman, 1990) tend to group the components that occur in different critical bands if they start at the same time. It may be that this process segregates the components when they start asynchronously and that the more sudden the onsets, the more strongly they are segregated. In our preliminary observations, however, we were surprised to observe that the sudden onsets enhanced the perception of the individual components of a cluster even when they were within the same critical band. Therefore we situated all our components within a single critical band. This was also done to prevent the stream segregation process from isolating tones from one another, given that a rapid sequence was to be employed. Had the tones been spaced further apart in frequency, their segregation into different streams might have isolated them so strongly that the effects of rise or fall time might not have been noticed.

Let us consider the implications of the fact that all our components were within a critical band. When we listened to versions of our stimuli that were not asynchronous (i.e., not like those used in the present experiments), we found it difficult to hear three different pitches in the mixture. Only when a component changed in intensity did its pitch become salient. This suggested to us that sudden amplitude changes might cause a "resetting" of the pitch-analysis mechanism. In a natural environment, sounds often are mixed and their onsets can follow one another in rapid succession. To cope with this, the pitch system is capable of registering more than one pitch at a time. This capacity to derive multiple pitches is what allows us to hear out one pitch in a chord formed of complex tones. It is possible that one of the cues that tells the pitch system to start to calculate a new pitch is a sudden rise of intensity in one or more spectral regions. If the output of frequency-specific "onset" cells in the auditory system were the main ones used in this pitch analysis, the components that had suddenly changed would be dominant in specifying the new pitch. Presumably the other pure tones that were also present would not interfere with the computation of the new pitch, because they would trigger no onset responses at that instant. It is also possible that onset responses both (a) directly supply the needed frequency information and (b) point to a narrow spectral region for enhanced analysis. The net effect might be that in cases where the relative intensities of parts of a spectrum changed slowly, only one event with a single pitch would be perceived, whereas when parts of the spectrum changed rapidly in intensity, a new event (with its own separate pitch) would be heard.

I. EXPERIMENT 1

This experiment studied the effect of both abruptness of onsets and synchrony of onsets in the perception of tone clusters. More rapid onsets were expected to result in clearer perception of the pitch pattern formed by the onsets.

A. Method

1. Stimuli

The basic stimulus was a cluster of four pure tones in which the tones came on at different times but went off at the same time. The listeners heard two such clusters in succession and were asked to decide whether the order of the component tones was identical in the two clusters. To prevent the task being solved by listening to the first or last tones of a cluster, which are always the most salient ones, all the clusters started and ended with the same tone (*X*). One of the middle tones was higher (*H*) in frequency than the bracketing *X* tones, and the other was lower (*L*). The frequency of the first and last tones (*X*) was 800 Hz, and those of the two middle tones, 750 Hz (*L*), and 850 Hz (*H*). There were two possible orders of onset of tones in a cluster, which we shall schematize as *XHLX* and *XLHX*, and the listeners were required to discriminate them, i.e., in one version of the cluster, called *HL*, the high tone preceded the lower one while in the other order, called "*LH*" it followed it. This arrangement was employed for the reasons described in the Introduction.

While the frequencies of the first and last tones were the same, they can be thought of as being separate tones because each was generated by turning on a separate (software) oscillator at a different moment and mixing the outputs. From the point of view of the resulting signal, since the first and last tones overlapped in time and were in phase, the amplitude envelope at the *X*-tone frequency was the sum of the envelopes of the first and last tones. Since the same four tones were involved in both tone orders, and the offset ramp of the four tones reached zero at the same time, this was an *onset-order* discrimination task.

Diagrams of the envelope shapes and the onset asynchronies for the four-tone clusters in two of the conditions are shown in Fig. 1. The top boxes show schematics of the amplitude envelopes of the four components of a single cluster. These are superimposed to show the timing relations. The bottom boxes show their spectrographic representations. The scale is linear amplitude, not decibels. Each tone (software oscillator) had a linear rise in its amplitude envelope, no steady state, and a linear decay. We refer to this as a triangular envelope. We varied two properties of the signals: (1) the "velocity" of the rise in intensity of the tones (defined below), and (2) the stimulus onset asynchrony (SOA) of the four tones in a cluster. All four tones in a given cluster had the same velocity of rise, in terms of amplitude units per second (defined below), and they all ended together. The four onsets were evenly spaced in time. Both of the clusters to be compared on a trial had the same values of onset asynchrony and velocity of rise of amplitude. They differed only in the order of their components. Shown at the left and right in Fig. 1 are different velocities of onsets for a fixed onset asynchrony.

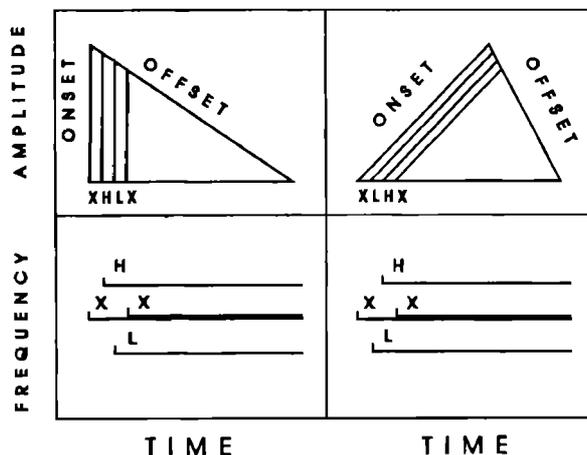


FIG. 1. Experiment 1: Diagram of two of the 1-s tone clusters. Top boxes show the amplitude and time relations by superimposing the four triangular amplitude envelopes used in generating the four components. Lower boxes show a spectrographic representation. Rise time of the first component is 10 ms (left), or 640 ms (right); onset asynchrony is 60 ms. Onset order is *XHLX* (left) and *XLHX* (right). $X=800$ Hz, $H=850$ Hz, $L=750$ Hz.

We expected that the orders would be easier to discriminate as the onset asynchrony became greater. Our only reason for including onset asynchrony differences was to create a range of difficulty for the subjects, whose discrimination abilities differed considerably, in order that no subject should either succeed or fail under all conditions. The values of onset asynchrony were 60, 80, and 100 ms.

The duration of the first tone of the sequence of four was one second. The other three tones started after the first, but ended at the same time. Therefore they were increasingly shorter. Only the first tone reached the maximum possible amplitude. This amplitude corresponded to ± 10 V out of the *D/A* converter of the computer, and was arbitrarily considered to have the value of 1 amplitude unit. All other amplitudes were scaled in terms of this value. Of course the actual amplitude at the ear of the listener depended on the setting of the amplifier; so the amplitude units are only meaningful as proportions of the maximum possible amplitude at that amplification. All tones after the first rose in intensity only until they reached the decaying amplitude envelope of the first, then decayed with it, following the same envelope (see Fig. 1). Therefore the maximum amplitudes of the three subsequent tones decreased, being limited by the decaying amplitude of the first one. This design guaranteed that none of the last three tones ever exceeded a previous one in amplitude. That is, they could not achieve their salience by momentarily "sticking out" above the amplitudes of their predecessors and dominating the waveform of the mixture.

Because of this aspect of the signal design, the rise of amplitude can be specified as "rise time" for only the first tone. For this tone, the rise times were 10, 40, 160, and 640 ms. However, another measure, "velocity of onset," expressed in terms of amplitude units per millisecond (AU/ms), can be applied to all the tones. [Note: Since the rise was linear in terms of amplitude, it could not be expressed in decibels per millisecond.] In terms of AU/ms, all the tones of a given cluster rose at the same rate (see the rising parallel

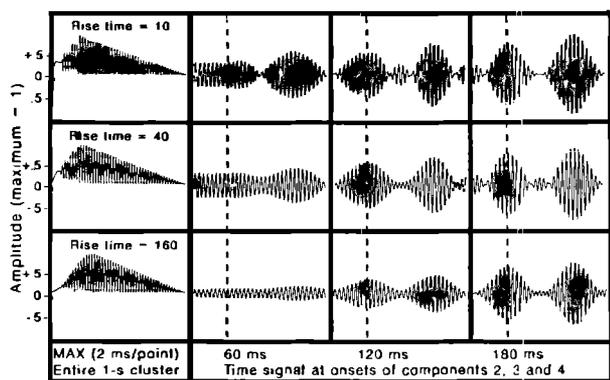


FIG. 2. Each row shows a cluster with a different rise velocity (labeled in terms of the rise time of the first component). All SOA's are 60 ms. The left-hand panels show plots of maximum values ("MAX") of successive 2-ms samples over the entire 1-s signal. The other three panels in the row are 40-ms segments extracted from the time signals, showing only the 40-ms segments in which the second, third, and fourth oscillators were "turned on" (at 10, 40, and 160 ms). Points of onset are shown by the dotted vertical lines. Segments run from 10 ms before each change to 30 ms after.

lines of Fig. 1). These rates were either 0.1, 0.025, 0.00625, or 0.00156 AU/ms. The triangular amplitude envelopes for the four frequency components (illustrated in Fig. 1) determined the relative amplitudes at which the tones were presented as a function of time.

Another fact should be mentioned: Due to the geometry of the envelopes, the maximum amplitudes reached by the second, third, and fourth oscillators were unaffected by the onset velocity. This can be seen by comparing the two clusters shown in Fig. 1.

In order to display the acoustic properties of the signal, we plotted, in Fig. 2, three of the conditions that had a 60-ms SOA. The rows differ in onset velocity (labeled in terms of the rise time of the first component). The three rows illustrate the clusters having different rise times (the three fastest). Abruptness of onset decreases from top to bottom. The left-most panel in each row allows the reader to visualize the amplitude envelope of the entire 1-s signal. This display was computed digitally by taking the maximum of the 32 sample points for each 2-ms segment. A line joining the peaks would be the amplitude envelope of the tone. The three panels to the right are on a more expanded time scale. They show 40-ms segments of the raw time signal at the points at which the second, third, and fourth components were added (i.e., at 60, 120, and 180 ms). The plotted segments extend from 10 ms before the onset to 30 ms after, and the points of entry are marked by vertical dotted lines. Both the 1-s amplitude display and the 40-ms segments of the time signal clearly show the 50-Hz amplitude modulation caused by the beating among the 50-Hz-spaced components whenever two or more are present. The top left panel shows the abrupt rises in energy when successive tones enter the mixture suddenly. It would be a mistake to consider the bump in the amplitude plot following the entrance of a tone to be at the frequency of that tone, since the signal can be analyzed as a carrier tone undergoing a 50-Hz AM, the frequency of the carrier gradually moving to a new value as the newly added tone gets stronger. Note that all three frequencies that were used (X ,

H , and L), are within a critical band and that the second and subsequent tones never exceed previous ones in amplitude. Despite this, when a new tone enters suddenly, it can be heard clearly as a separate tone by the experimenters. All the 40-ms time-signal displays show that there is no discontinuity in the waveform at points of onset. This was true because we controlled the phase relations in the signal, at the point of onset of each new component, by incorporating the following temporal parameters into the design of the stimuli: All components were sinusoidal and began in sine phase. Also, all frequencies were multiples of 50 Hz (20-ms period), and all asynchronies were multiples of the same 20-ms period. For these reasons, at the instant at which each later tone began, both it and the previous tones were at 0-degree phase. Thus as the intensity of each new tone grew, it did not introduce cancellation or sudden phase distortion. The relative contributions of the different frequency components, at any moment, are shown by the relative heights of the envelopes in Fig. 1.

As the second tone was added, the 50-Hz beating began. As the third tone was added, being 100 Hz away from the second one (e.g., 750 vs 850 Hz), a 100-Hz beating began. Note that these changes were all smooth, even for the most sudden rise time. They were much slower, of course, for the slower onset conditions. For example the one shown on the bottom left of Fig. 2, shows no visible bumps in the overall envelope when the successive tones enter the cluster, and the waveforms shown to its right change very slowly.

At the time of presentation, the varying asynchrony values and their effects on the attenuation of the later components caused the intensities of the clusters to vary between 57 and 60 dB, A-weighted. This intensity difference was correlated only with asynchrony of onset. The reader can verify, from the geometry of Fig. 1, that the maximum amplitudes reached by the later components within a cluster were not affected by rise time.

2. Task

The listeners carried out a same-different judgment task. There was one training session consisting of 12 trials, and four test sessions, each consisting of 48 trials. The task was the same in all sessions. Each trial presented two four-tone clusters, A and B , for comparison, and the subjects had to judge whether they were the same or different. An example of a "same" trial was $XLHX$ (800, 750, 850, 800 Hz) followed by itself a second time. An example of a "different" trial was $XLHX$ (800, 750, 850, 800 Hz) followed by $XHLX$ (800, 850, 750, 800 Hz). On half the trials, the clusters were the same and on half, the LH order was different. For half the "same" trials, the LH version of the stimulus was used, and for the other half, the HL version. On half the "different" trials the LH version came first, and on the other half, it came second. The trial began with a 100-ms warning tone of 800 Hz, then a 1-s silence, then cluster A , then a silence of from 1.3 to 1.5 s depending on the condition, then cluster B . This time variation was used so that the intercluster silence could be held approximately constant between the onsets of the fourth tone of cluster A and the first tone of cluster B . This tended to give a constant rhythm to the trials.

TABLE I. Experiment 1. Raw rated similarity scores and *C* scores. No. of cases=19.

Rise tone 1 (ms)	SOA (ms)	Comparison	Mean	s.d.	Std. err.	<i>C</i> score
10	60	Same	6.54	0.85	0.196	1.12
		Different	5.42	1.32	0.302	
10	80	Same	6.77	0.67	0.154	2.54
		Different	4.23	1.85	0.425	
10	100	Same	7.13	0.57	0.131	4.06
		Different	3.07	1.74	0.401	
40	60	Same	6.41	0.70	0.161	1.57
		Different	4.84	1.45	0.334	
40	80	Same	6.71	0.94	0.215	3.14
		Different	3.57	1.71	0.392	
40	100	Same	7.08	0.77	0.176	4.22
		Different	2.86	1.68	0.386	
160	60	Same	6.07	0.95	0.219	0.90
		Different	5.17	1.16	0.266	
160	80	Same	6.10	0.71	0.164	1.84
		Different	4.26	1.35	0.310	
160	100	Same	6.34	0.97	0.224	2.80
		Different	3.54	1.52	0.349	
640	60	Same	6.54	0.77	0.178	0.53
		Different	6.01	0.91	0.209	
640	80	Same	6.57	0.95	0.219	0.75
		Different	5.82	0.94	0.215	
640	100	Same	6.37	0.86	0.198	0.70
		Different	5.67	1.01	0.232	

The listener could take as long as necessary to judge the sameness or difference of the two clusters on an 8-point rating scale, in which 1 was labeled "very clearly different" and 8 was labeled "very clearly the same," with intermediate numbers representing intermediate judgments. The next trial started 1 s after the listener entered his or her judgment on a computer keyboard. An 8-point rating scale was used because we judged that it contained more information than a 2-point same/different judgment.

In the training period, an easier version of the task was used. The rise time of the first tone was always 10 ms, the most abrupt rise time to be used in the later test. The velocity of onset of the first tone and the subsequent tones was, therefore, 0.1 AU/ms. The training signals had one of three possible values of onset asynchrony: 150, 200, and 250 ms, which were greater than the asynchronies used in the test sessions. The training trials were not presented randomly, but in a fixed order of increasing difficulty (i.e., with decreasing asynchrony). Following this procedure, subjects were shown their scores.

In the test sessions, there was no feedback of results to the listeners. Forty-eight conditions were presented in random order in each test session: 4 onset velocities, 3 asynchronies, 2 same-vs-different cluster comparisons, and 2 possible orders in the first cluster, *LH* or *HL*.

3. Apparatus and subjects

All stimuli were digitally synthesized and output via a 16-bit digital/analog converter at 16 000 samples per second, and low-pass filtered with a Butterworth-design filter at 8 kHz with 48 dB per octave roll-off. Because of the low frequencies used, any aliasing components (all above 15.15 kHz) were attenuated by at least 44 dB. The stimuli were

presented binaurally over headphones (Sony model NRV7) in an Industrial Acoustics single-wall test chamber. The intensity of the signals coming from the headphones was measured by a SPL meter at fast A weighting, using a flat-plate coupler.

Twenty young adults, recruited from a university population, served as paid subjects. Their age range was 20 to 33 years. Before examining the data with respect to any of the experimental variables, we discarded the results of one subject who was unable to discriminate between same and different signals.

B. Results

The means taken over subjects for the raw sameness rating scores (1 to 8) are shown for the major experimental conditions in Table I together with standard deviations and standard errors. Higher values indicate a stronger judgment of "sameness."

Better performance is indicated, for the physically same stimuli, by a higher score, and for the physically different ones, by a lower score. More simply, the ability to discriminate same from different pairs of stimuli is seen in the difference between the sameness scores assigned to physically same and physically different pairs. If one subtracts the scores for different pairs from the one for same pairs within any condition, the resulting score represents discrimination. Accordingly, the raw scores were converted into an discrimination score by using the following formula for each asynchrony-by-velocity condition:

$$C = [(A + B) - (D + E)]/2,$$

A = rated sameness of *LH-LH* trial (i.e., *XLHX* paired with itself),

TABLE II. Experiment 2. Raw rated similarity scores and *C* scores. No. of cases=18.

Rise tone 1 (ms)	SOA (ms)	Comparison	Mean	s.d.	Std. err.	<i>C</i> score
10	60	Same	5.52	1.09	0.258	0.32
		Different	5.20	1.23	0.290	
10	80	Same	5.86	1.02	0.240	0.88
		Different	4.99	1.45	0.341	
10	100	Same	5.52	1.00	0.236	0.60
		Different	4.92	1.08	0.254	
40	60	Same	5.35	1.12	0.263	0.26
		Different	5.09	1.36	0.320	
40	80	Same	5.74	0.82	0.192	0.78
		Different	4.96	1.18	0.279	
40	100	Same	5.83	0.97	0.229	0.76
		Different	5.06	1.25	0.295	
160	60	Same	5.17	1.26	0.297	0.22
		Different	4.94	1.51	0.355	
160	80	Same	5.03	1.08	0.254	0.34
		Different	4.69	1.32	0.311	
160	100	Same	5.64	1.07	0.253	0.85
		Different	4.79	1.39	0.327	
640	60	Same	5.05	1.76	0.414	-0.19
		Different	5.25	1.75	0.412	
640	80	Same	4.95	1.60	0.377	-0.10
		Different	5.06	1.67	0.392	
640	100	Same	5.03	1.72	0.406	0.10
		Different	4.92	1.70	0.400	

B = rated sameness of *HL-HL* trial,
D = rated sameness of *LH-HL* trial,
E = rated sameness of *HL-LH* trial.

Note that on *A* and *B* trials the stimulus clusters are physically the same whereas on *D* and *E* trials they are physically different. The raw mean values shown in Tables I and II show that subjects tended to favor the higher parts of the scale and this might be interpreted as some sort of bias. However it could have been due to some irrelevant cause (such as the right-handedness of subjects). Our use of the difference between the means of the physically same and different stimuli (presumably biased in the same way) to obtain a *C* score was designed to eliminate this bias from the data. Therefore we will give an interpretation only to the *C* scores, and not to the raw means themselves.

C scores could range from +7 (perfect discrimination) through 0 (no discrimination) to -7 (calling all identical pairs "different" and vice-versa). They are shown as a function of onset velocity in Fig. 3. In this and all subsequent figures, for the sake of simplicity, the rise/fall velocities are expressed in terms of the rise time of the first *X* tone.

While it is not necessarily the case that the raw sameness scores are normally distributed, the distribution of any sum or difference formed out of 8 or more of such scores will approximate normality. Since the mean score for each subject in each asynchrony-by-velocity condition was based on 16 raw scores, it seemed appropriate to analyze such means by ANOVA. In analyzing Experiments 1 to 3 by ANOVA, the *F* ratios derived from both univariate and multivariate models were computed, but only the univariate statistics are reported below unless one model yielded statistical

significance and the other did not. In these cases, the conflict is noted.

Figure 3 shows a strong effect of onset velocity on discrimination, $F(3,54)=25.3$, $p<0.0001$. It appears that, apart from a slight decline for the most abrupt onsets, the more sudden the onsets the greater the discrimination of order. The small drop at the most abrupt onsets as compared with per-

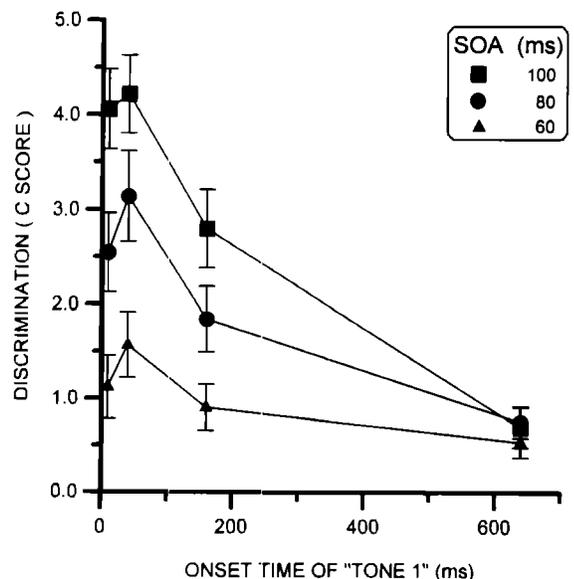


FIG. 3. Experiment 1: Discrimination scores averaged across 19 listeners as a function of the onset time of "tone 1" (the first tone in forward clusters) for asynchrony values of 60, 80, and 100 ms. 0 is chance performance, +7 is perfect. The pooled standard error (derived from ANOVA) was 0.69. Means above 1.14 are significantly different from zero ($p<0.05$, one-tailed).

formance at the next fastest rise was observed at every asynchrony. ANOVA showed that these differences were significant, $F(1,18)=6.13$, $p<0.05$. We thought originally that this may have occurred due to spectral spread caused by the very fast onsets, or by an increased forward masking of the later tones by the earlier ones. However, this drop in performance seemed to decline over sessions, presumably because of increasing familiarity with the task. By the last session, it had disappeared. In that session, the trend for better performance with more abrupt onsets was monotonic. This change over sessions was responsible for an interaction between the two variables, test session and the velocity of onset, $F(9,162)=3.58$, $p<0.001$. Therefore the detrimental effect of the 10-ms onsets may have been due to its effect on attention (perhaps distraction by perceptual clicks arising from the splatter of energy at the onsets, or by the alerting quality of the sudden onsets themselves), and the listeners seem to have been able to overcome it with experience. In general, discrimination improved a little as a function of test session, $F(3,54)=5.2$, $p<0.01$. Means for the four sessions were, respectively, 1.55, 2.12, 2.12, and 2.28. There was no significant interaction between sessions and onset asynchrony.

Finally, as expected, discrimination improved somewhat as the onset asynchrony became longer, $F(2,36)=65.9$, $p<0.0001$. The mean discrimination scores for 60, 80, and 100-ms asynchronies were, respectively, 1.03, 2.07, and 2.95. There was also a significant interaction between onset asynchrony and the abruptness of the onsets, $F(6,108)=15.94$, $p<0.0001$. This occurred because abrupt onsets were more beneficial at longer asynchronies (see Fig. 3). One way to interpret this is to say that the benefit of the more abrupt onsets was not visible when the overall performance was very low as a result of small onset asynchronies.

II. EXPERIMENT 2

The next experiment studied the effect of the suddenness of the offset, rather than onset, of tones in a four-tone cluster. We had noticed, in pretesting, that sudden offsets, occurring at a different time for each tone in the cluster, while not conveying as clear a sense of the pitch of the separate components as an onset did, still seemed to enhance them. Instead of creating a sequence of brief clear tones, as sudden onsets did, sudden offsets created "plops" with slightly different pitch colorations for the different tones.

A. Method

Experiment 2 employed the same stimuli as experiment 1, just reversing the signals digitally. In doing so, we guaranteed that any effects of "spectral spread" (or "splatter") would be the same for these offsets as for the onsets of experiment 1 (since the power spectrum is symmetrical with respect to time).

The reversed signals had synchronous onsets but asynchronous offsets, as one would see in a left-right reversal of Fig. 1. The waveforms were the reverse of those shown in Fig. 2. The experimental procedure was the same as in experiment 1, as was the structure of a trial. The variables were the same, except that onset asynchrony became offset asynchrony.

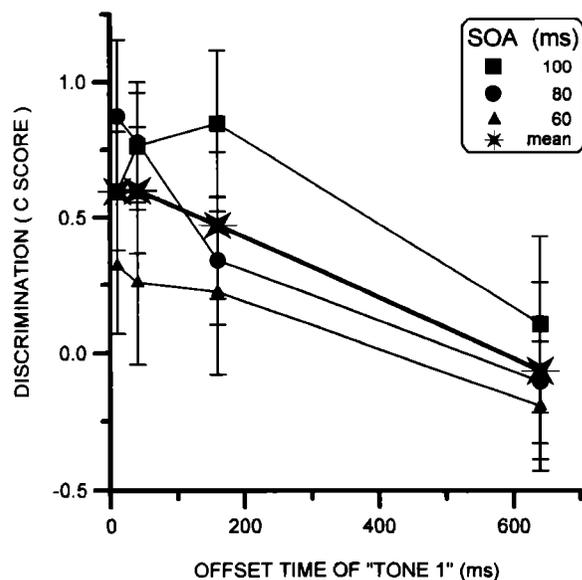


FIG. 4. Experiment 2: Discrimination scores averaged across 18 listeners as a function of the offset time of "tone 1" (actually the last tone in reversed clusters). 0 is chance performance, +7 is perfect. Pooled standard error (derived from ANOVA) was 0.57. Individual means must be higher than 0.94 to be significantly different from zero ($p<0.05$, one-tailed). Note that the vertical scale is greatly expanded relative to that of Fig. 3.

Eighteen young adults, recruited from a university population, and ranging from 20 to 29 years of age, served as paid subjects.

B. Results

The means over subjects for the raw sameness rating scores (1 to 8) are shown for the major experimental conditions in Table II together with standard deviations and standard errors. Higher values indicate a stronger perception of "sameness."

A discrimination score (C) was derived for each condition of offset velocity and offset asynchrony as in Experiment 1. Figure 4 shows these scores as a function of offset velocity, expressed in terms of the offset time of "tone 1", the final X tone. (In all the figures, "tone 1" refers to the first-rising tone of the forward stimuli or the last-falling tone of the reversed stimuli. This is the only tone that reaches an amplitude of 1 before decaying.) The standard error bars in Fig. 4 seem larger than in Fig. 3, but this is an illusion caused by the fact that the Y axis in Fig. 4 is stretched to allow more details to be seen in the data despite its compressed range.

These scores are very much lower than in experiment 1 where the best onset velocity, averaged across asynchronies, led to a mean score of about +3 (on a scale running from -7 to +7), where 0 indicates random performance and positive numbers reflect degrees of correct discrimination. In the present experiment the best offset velocity led to a score of only +0.6.

Due to the compressed range of scores, the offset effect is rather small as compared to the noise in the measurements, and if we separate out the three curves for different degrees of asynchrony, they look very irregular (see Fig. 4). Only if we combine them (heavy line with star symbols), thereby aggregating across more observations, do we see a mono-

tonic effect of onset velocity. Due to the noise, none of the points on the curve of "means" in Fig. 4 (i.e., averaged over onsets), when taken individually, was significantly different from zero (we can call this "result A"). Despite this, there was a significant effect of offset velocity on discrimination, $F(3,51)=5.5$, $p<0.01$ (call this result B). [We note that the Wilks' Lambda multivariate test gave a less significant probability, $p=0.023$.] The discrepancy between result A and result B is explainable, first because the power of the test that led to result B is higher, since it involves more observations and second, because the relation of these means to the zero point is not considered in computing it. However, despite their statistical weakness, we are encouraged to believe the observed pattern of means because they show the same result as do the curves of experiment 1: the more sudden the change in amplitude, the greater the discrimination of order. In contrast with experiment 1, there was no consistent drop in performance for the most abrupt offsets relative to the next most abrupt. It was observed in only two of the four test sessions and could have been due to chance. The means collapsed over sessions (as in Fig. 4) do not show it.

The results for the other variables were not significant: (a) There was no significant effect of sessions. The means for the four sessions were 0.12, 0.51, 0.62, and 0.34. (b) The effects of asynchrony of offset were not significant, although performance changed in the expected direction. The mean scores for asynchronies of 60, 80, and 100 ms were, respectively, 0.15, 0.48, and 0.58. (c) There were no significant interactions among any of the experimental variables.

III. SUMMARY AND GENERAL DISCUSSION

A. Summary

The findings concerning onset velocity can be summarized as follows: When clusters of pure tones of different frequencies were highly overlapped in time, the ability to judge the order of onset of the components was strongly affected by the suddenness of their onsets (experiment 1). When we ourselves listened to the clusters, the pitches of the more abrupt onset tones seemed more distinct than those of tones with slower onsets. As the onsets became more abrupt, the ability of our subjects to discriminate the order became more accurate up to and including the second fastest onset velocity used. At the fastest one, there was a small decline in performance. However, this decline could have been due to some disruptive effect caused by the most abrupt onsets, and tended to disappear in later sessions.

In experiment 1, the effects of onset *velocity* were strongest at the greater onset *asynchronies*. It is possible that the onset-velocity effect reduced the overlap of the rise portions of the four components; increased asynchrony would further reduce the overlap. We will come back to this issue in Sec. III C.

When the onsets were synchronous but the offsets asynchronous (experiment 2), abruptness of change had small, but detectable, effects. More abrupt offsets made it a little easier to discriminate the order in which the different pitches dropped out. However, the ability to detect the order of offset was much worse than for the order of onset, and the differ-

ence in performance between the best and worst offset times was very small. When the experimenters listened to the asynchronous offset signals, the sudden offsets sounded like "plops" colored by the pitch of the component that was dropping out.

Our experiment extends the findings of the experiment by Pastore and Harris (1982), described in the Introduction. Their subjects could have made the discrimination by focusing on the pitch of the first onset, which, before the sudden appearance of the second tone, was present alone. The more sudden the onset, the greater its maximum intensity in this time period. Our experiments showed that the critical tones did not have to include the first one, nor did they have to be more intense, at the highest point in their envelopes, than the other tones in the cluster at that instant. Second, in their experiment the tones were over two critical bands apart. The present experiments showed that this is not a necessary condition to obtain the benefit of sudden rise or fall envelopes.

B. Issues in interpreting the results

Rasch (1978) reported a result that is discrepant with those of the present experiment. Using concurrent target and masker tones, he found that the target tone had to start *before* the masker, rather than after, to obtain any benefit from the asynchrony. If Rasch's findings held true in our tone clusters, only the first tone should have been helped by the asynchrony. Yet the second and/or third tones must have been helped, since they were the ones essential for the discrimination. (It is possible that the subjects in our same-different discrimination task could have been using only the second or only the third tone; however, the contrast with Rasch's finding is unaffected by this). The apparent discrepancy with Rasch's findings may have resulted from our use of onset delays (60, 80, and 100 ms) that were much larger than his (10, 20, and 30 ms). It is well known that in simultaneous masking, a longer delay of the onset of the target relative to that of the masker will improve detection (the "overshoot" effect studied by Zwicker, 1965; Bacon and Moore, 1986; McFadden, 1989; Carlyon and White, 1992, and many others). Therefore it seems that asynchrony can help unmask the delayed tone as well as the advanced one, if the asynchronies are large enough.

The observation that sudden offsets facilitate order discrimination less well than sudden onsets do (and give less clear pitches to the experimenters) suggests two alternative explanations: (a) the offset responses in the auditory system carry weaker frequency-specific information, or (b) they select a spectral region for enhanced analysis, but by the time the auditory system is alerted to the sudden change, the energy has greatly diminished, the sensory memory has decayed, and there is less left to analyze. This second explanation is, however, incompatible with the observation that *faster* offsets produced the greatest clarity of tones. If the rapid disappearance of relevant energy had been the problem, the most rapid offsets should have been worst, not best.

At this point we should address the question of whether spectral "splatter" of the fastest rising or falling tones could account for the results. With our design, the effects of splatter in adding nontarget frequencies to the signal (perhaps

altering its timbre) could not be responsible for the effects of rise velocity. All tones in a cluster had the same rate of amplitude change, so that they would have similar degrees of splatter. Also the two clusters to be compared for the same-different judgment had the same rates of change, and the asynchronies were at the same place in the cluster (at the onset or offset). The only thing that distinguished the pair of clusters was the order of tonal components (*XHLX* vs *XLHX*); splatter-induced timbre change was useless for discrimination. To predict the present data using the concept of spectral splatter, one would have to argue that splatter in the spectral region of a tone somehow alerted the auditory system. That is, it is not the sudden rises *per se*, but their spectral consequences that trigger attention. However, this cannot be the entire explanation since splatter would be appreciable in size only for the 10-ms rise (or fall) time. Yet we see a continuing decline of performance between 40 and 640 ms. Finally, since splatter would have given the same advantage to the rapid *offsets* of experiment 2, performance should have been enhanced there too; it was not.

Another issue concerning interpretation is visible in Fig. 1 in the illustrations of the sudden rise times, shown on the left. In comparison with slower-onset clusters, at the instant at which any tone enters the mixture, the previous tone has reached a higher intensity level. Indeed, with the shortest rise times, the onset regions are completely nonoverlapping in time. If the auditory system is processing onset responses, their temporal isolation might be important in an order-judgment task. The stimuli of the present experiment are not suitable for testing this explanation against our own, which sees the rate of rise *per se* as the critical factor. For this reason, other research in our laboratory studied the effects of sudden amplitude changes using stimuli in which the tones *never* overlapped in time (Bregman *et al.*, 1994). In a first experiment, a sequence of two pure tones was presented against a background of masking noise. Their rise times did not affect the ease of isolating them from the masker. However, a second experiment using a chord in which two components were *incremented* in succession (cf. Kubovy, 1981) showed a sizable, monotonic effect of the rise time of the increments, more rapid changes leading to better discrimination. This occurred with increments of as little as 1 dB. The conditions of the present experiments with tone clusters were more similar to those of the second experiment of Bregman *et al.* than to those of their first, since the background of each component of the clusters in the present study consisted of tones rather than noise. The results of the Bregman *et al.* experiments, in which only one tone changed at a time, were clear and in close agreement with our present ones. This encourages us to believe that the increased separation of the rising (or falling) portions of stimuli in our experiments, resulting from abrupt changes, may have played a contributory role in clarifying the individual components, but was not solely responsible for producing the effects observed in the present experiments. If the increased nonoverlap with more abrupt onsets were the only factor responsible for the advantage of abruptness, we would expect this advantage to be found equally well at the offsets, yet the latter showed a much smaller effect. Clearly the rate of change of intensity

itself seems to be very important in the segregation of a sinusoidal component from a background of other sinusoidal tones.

In the introduction, we asked whether the auditory system might make use of the suddenness of a change in intensity to decide whether the change had resulted from a change in the manner of vibration of a single sound-producing event or from a second event starting or stopping. The present experiments and those reviewed in this paper suggest that it does.

C. General discussion

1. Adaptive significance of the sudden-change effect

The auditory system's special sensitivity to onsets and offsets may be mediated by cells that can be found at various levels of the system (see the review by Keidel *et al.*, 1983). Some respond exclusively to onsets and others to offsets. Olsen (1994) presented squirrel monkeys with a short noise burst whose amplitude envelope was supplied by a single cycle of a raised cosine. The duration of the noise was in the range 20 ms to 1 s (10 to 500 ms onset times). He found cells in the medial geniculate that responded better to the shorter bursts (which also had a faster rise time). More specifically, the maximum firing rate seen in any 20-ms temporal window in a poststimulus histogram declined with longer bursts (slower rise times). In another study of Olsen's, the stimuli were broadband noise bursts (0–20 kHz), 600 ms in duration. Rise/fall times were linearly ramped and ranged from 0 to 100 ms in duration. Olsen first presented a series of bursts with a fixed peak amplitude (at least 10 dB above the neuron's threshold for 0-ms rise time noise) and varied the rise time. The shortest rise times produced the lowest threshold responses. Then, the same rise time series was repeated at a higher peak amplitude (e.g., 20 or 30 dB above threshold). Most neurons responded in a transient way to the onset of the noise stimulus, but only for the shorter rise times. However, as peak amplitude was increased, the neurons responded to longer and longer rise times. Thus there appeared to be a trade-off between rise time and peak amplitude, suggesting that it was the *rate* of rise (what we have referred to, above, as onset velocity), rather than the absolute rise time, that was important in evoking a response.

Of what value would it be to an animal to respond to the *velocity* of the rise rather than its duration? Let us first point out why it might be important to detect the suddenness of an onset at all. In natural environments, many events happen concurrently and each can stimulate a number of frequency channels. The suddenness of increase of the energy in a particular frequency band may be important because it is signaling the onset of a new event that has begun to affect the total energy in that band, whereas a slow change might signal merely a changing manner of vibration of a single source. To make this distinction, the auditory system must register the suddenness of the rise in amplitude. Another use of onsets might involve the detection of a synchronous rise in a number of frequency bands, all at the same rate, which could indicate that the energy in them has probably all come from the same event. However, there would be a problem in using

the actual detected *time* of the rise for these purposes because its value would depend on the presence or absence of background noise; the early parts of the rise would be masked by the background sounds, and the auditory system would not be able to assess exactly when the rise had begun. On the other hand, the *rate* of increase (in terms of linear amplitude change) would be independent of the signal-to-noise level, and could be detected any time after the signal emerged from the noise floor.

2. Sudden onsets as a "resetting" effect

Responsiveness to change is found in all sense modalities and plays a role in their efficiency. The senses do not continue to analyze an stimulus after it ceases to contribute novel information. An example has been observed by Hafter and his colleagues, who have done a number of studies on "resetting the binaural system." They have found that this system habituates to a given interaural disparity so that the continuing stimulation contributes less and less to the accuracy of position judgments. However, any kind of "new information" or spectral discontinuity anywhere in the audible spectrum can retrigger the processing of interaural disparity (Hafter and Buell, 1990).

Similarly Clifton and her colleagues have observed an "resetting" effect that occurred as part of the auditory system's integration of an echo with the original sound. They presented a series of clicks or other sounds, with an delayed copy following each one and coming from a second loudspeaker at a different location (Clifton, 1987). With repeated click pairs, at first the pairs were heard as separate sounds; then there was a buildup of the tendency to fuse the original and its copy into a single experience (Clifton and Freyman, 1989; Freyman *et al.*, 1991). Presumably, the auditory system had "learned" from the constant delay in the induction pairs to treat the copy as an echo of the original acoustic energy, rather than the onset of a new sound-producing event. After integration had occurred, the experimenters could reset the mechanism, restoring the perception of two separate sounds, if they suddenly altered the spatial arrangement, (e.g., by interchanging the locations of the first and second sounds). After a few click pairs in the new condition had been heard, they again became integrated. Apparently, the switch in location re-established the click sequence as a novel event, canceling out the effects of the earlier learning and resetting the system.

Another "resetting" by a sudden change in properties has been observed by Rogers and Bregman (1991). When high and low tones were alternated in a sequence, there was a tendency to segregate the high from the low tones, and this tendency built up over time. However, if both tones in the sequence were suddenly made to come from a new location (apparent or real), the segregation process was reset and there was, for a short time, less stream segregation. In other words, the segregation process was reset and started again at its default setting of integration.

We can interpret the present experiment as revealing another resetting effect that belongs to this family. Our hypothesis is that our results occurred because pitch computations were reset at frequency-time points of sudden spectral change. While there may be other forms of interpretation of our data, for example physiological, ours links it to other resetting effects and gives it a functional significance, that of computing new properties for sound-producing events that are newly entering the ongoing mixture.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada, and by the FCAR program of the Quebec Ministry of Education.

- Bacon, S. P., and Moore, B. C. J. (1986). "Temporal effects in simultaneous pure-tone masking: effects of signal frequency, masker/signal frequency ratio, and masker level," *Hear. Res.* **23**, 257-266.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).
- Bregman, A. S., Ahad, Kim, J., and Melnerich, L. (1994). Resetting the pitch-analysis system. 1. Effects of rise times of tones in noise backgrounds or of harmonics in a complex tone," *Percept. Psychophys.* **56**, 155-162.
- Bregman A. S., and Pinker, S. (1978). "Auditory streaming and the building of timbre," *Can. J. Psychol.* **32**, 19-31.
- Carlyon, R. P., and White, L. J. (1992). "Some experiments relating to the overshoot effect," *Auditory Perception and Physiology: Proceedings of the 9th International Symposium on Hearing*, Carcans, France, 9-14 June 1991.
- Clifton, R. K. (1987). "Breakdown of echo suppression in the precedence effect," *J. Acoust. Soc. Am.* **82**, 1834-1835(L).
- Clifton, R. K., and Freyman, R. L. (1989). "Effect of click rate and delay on breakdown of the precedence effect," *Percept. Psychophys.* **46**(2), 139-145.
- Dannenbring, G. L., and Bregman, A. S. (1978). "Streaming vs fusion of sinusoidal components of complex waves," *Percept. Psychophys.* **24**, 369-376.
- Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (1991). "Dynamic processes in the precedence effect," *J. Acoust. Soc. Am.* **90**, 874-884.
- Hafter, E. R., and Buell, T. N. (1990). "Restarting the adapted binaural system," *J. Acoust. Soc. Am.* **88**, 806-812.
- Keidel, W. D., Kallert, S., and Korth, M. (1983). *The Physiological Basis of Hearing: A Review* (Thieme-Stratton, Stuttgart).
- Kubovy, M. (1976). "The sound of silence: a new pitch-segregation phenomenon," Paper presented at the 17th annual meeting of the Psychonomic Society, St. Louis, 11-13 November 1976, No. 66.
- Kubovy, M. (1981). "Concurrent-pitch segregation and the theory of indispensable attributes," in *Perceptual Organization*, edited by M. Kubovy and J. R. Pomerantz (Lawrence Erlbaum, Hillsdale, NJ).
- McFadden, D. (1989). "Spectral differences in the ability of temporal gaps to reset the mechanisms underlying overshoot," *J. Acoust. Soc. Am.* **85**, 254-261.
- Olsen, J. F. (1994). "Sensitivity of medial geniculate neurons in the squirrel monkey to rate of rise," *J. Neurosci. Abstr.* **20**, 321.
- Pastore, R. E., Harris, L. B., and Kaplan, J. K. (1982). "Temporal order identification: Some parameter dependencies," *J. Acoust. Soc. Am.* **71**, 430-436.
- Rasch, R. A. (1978). "The perception of simultaneous notes such as in polyphonic music," *Acustica* **40**, 21-33.
- Risset, J. C., and Mathews, M. V. (1969). "Analysis of musical instrument tones," *Phys. Today*, **22**(2), 23-30.
- Rogers, W. L., and Bregman, A. S. (1991). "Auditory streaming: a central change-sensitive process," *Can. Psychol.* **32**(2A), 290 (A).
- Zwicker, E. (1965). "Temporal effects in simultaneous masking by white-noise bursts," *J. Acoust. Soc. Am.* **37**, 653-663.