THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2006, 59 (3), 543–556

Psychology Press
Taylor & Francis Group

Effects of unit formation on the perception of a changing sound

Poppy A. C. Crum and Albert S. Bregman
McGill University, Montréal, Canada

The effects of subunit formation on adult listeners' ability to notice changes in a continuous spectral gradient of sound were studied. Results of this experiment support the idea that the auditory system processes information differently within a unit, and that this processing does not occur unless the perceptual system detects unit boundaries. In this experiment, silences were inserted into a continuously changing sound to cause the formation of short units. Listeners noticed the change earlier in conditions with silences inserted than in to conditions where the transition was either unbroken or broken by loud noise bursts. Results are discussed in terms of two processes, one that accentuates stimulus properties present at moments of onset and offset, and a second that uses onsets and offsets to signal the beginnings and ends of units and reduces the change perceived within units.

The study of the perceptual organization of sound has so far focused on two aspects of dealing with (organizing) the incoming sensory information. The first is the allocation of the simultaneously received components of the spectrum into different concurrently perceived sounds. An example is the division of the spectrum emanating from a musical ensemble into separate integrated sets of components, in which each set groups the spectral components of an individual instrument—this latter process being referred to as the "integration of simultaneous components" (Bregman, 1994, Chap. 3). The function of a second process is to connect, over time, sound that is part of a sequence of events arising from the same source in the environment. This latter process is called "sequential integration", and it forms separate auditory streams (Bregman, 1994, Chap. 2). An example is the connecting, over time, of the notes that form a single melodic line or "voice". These aspects of the subdivision of sensory information are strongly interactive, although they can be distinguished conceptually.

However, there is a third aspect of subdividing the sensory input that has been largely neglected—namely, breaking it into temporal units. When we think of perceptual grouping, it is tempting to think of perceptual organization as forming clusters out of already existing temporal units. The typical experiment on grouping starts with some discrete units called "tones". But this way of proceeding hides a deep question: Where do

Correspondence should be addressed to Poppy A. C. Crum, School of Medicine, Department of Biomedical Engineering, Johns Hopkins University, 720 Rutland Avenue, Traylor 412, Baltimore, MD 21205-2109, USA. Email: poppy@jhmi.edu

This work was supported in part by a research grant (R01 MH52254-03) from the National Institute of Health to ASB. We wish to thank Pierre Ahad for his help in the preparation of this manuscript and Dr. Philip B. Stark for his help with statistical methods discussed in this manuscript, as well as Dr. David Wessel, Dr. Willliam Prinzmetal, Dr. Anne-Marie Bonnel, Dr. Sven H. Khatri, and anonymous reviewers for providing helpful comments and suggestions.

http://www.psypress.com/qjep
DOI:10.1080/02724980443000737

units come from? Are they a preexisting property of the sound—say, any arbitrary 10-ms segment? Are all long tones derived by integrating some underlying shorter tones? Analogously, in vision, is a long line just an unbroken succession of short lines laid end to end? If we think of perceptual organization as "grouping", we have little choice—we must begin with short units and group them to form larger units.

In contrast with this approach, we base our argument on the Gestalt idea that the default state of the visual field is undifferentiated, and that visual contours arise as a consequence of discontinuities in the field. Transposing this concept to audition, auditory units are perceived as a result of discontinuities in the auditory input. The information-processing role of perceptual units is that each unit should represent a simple environmental event. For example, a piano tone represents a distinct impact of a hammer on one or more strings. These events/units can be incorporated into larger perceptual patterns. When perceptual units are grouped into streams, this indicates that the auditory system is treating all the events, in a succession of events in the environment, as coming from the same sound source. To summarize, units are not already existing entities that *enter into* the perceptual analysis of sensory information; they come *out* of the analysis.

There are two general questions that can be asked in the study of unit formation: (a) What stimulus features cause units to form? (b) What consequences does the formation of units have for our perception of the flux of sound?

Regarding the first question, we made the simple assumption in the following study, that a rapid rise in intensity causes the listener to hear the beginning of an event, and that a rapid drop in intensity causes the unit to end, as found in previous research (Bregman, Ahad, & Kim, 1994a, 1994b; Nakajima, 1996; Royer & Robin, 1986). The present study used this assumption to examine the second question: the effects upon perception of the formation of units. For the perceptual domain, we chose nonspeech timbre because of its relation to both natural sound and music. In particular, we studied whether the formation of units within a continuously changing timbre (a) exaggerated the differences in timbre between units (Gestalt notion of "sharpening") and (b) reduced perceived differences within the same unit (Gestalt notion of "levelling").

Research investigating the effects of the formation of units has been carried out primarily in the visual modality, and it is helpful to consider the treatment of a unit within the visual system as a possible parallel to the treatment of a unit within the auditory system. In the complex visual scene it is often true that a single visual object may have many different levels of luminance within a single surface area. Variables such as interacting light sources or reflections from other objects can cause a single surface to have regions that are darker or lighter in luminance. Previous demonstrations of Gestalt psychology in vision suggest that the visual system tends to form units within regions bracketed by abrupt changes of luminance (Koffka, 1915, p. 46). We tend to perceive a unit as more homogeneous in surface colour than it really is. Unit boundaries, when introduced into a visual gradient, appear to function as cues that the visual system uses to identify regions of space that should belong together and appear similar. If this description is essentially correct, the formation of a visual unit should affect the perceived properties of the surface that comprises it.

We can use Benussi's rings as an example, in the visual system, of the tendency of regions enclosed by unit-defining contours to remain perceptually homogeneous (Koffka, 1915, p. 46). When a neutral grey ring is overlaid—half on top of a black background and half on top of a white background—the ring appears more or less a homogeneous grey (Figure 1A) despite contrast relationships that might be expected to cause portions of the ring to be perceptually darker (against the white background) or lighter (against the black background). However, if a thin black line divides the ring precisely at the border of the two contrasting backgrounds, the expected contrast relationships are now perceived, and the portion of the ring on the white background appears darker than the portion of the ring on the black

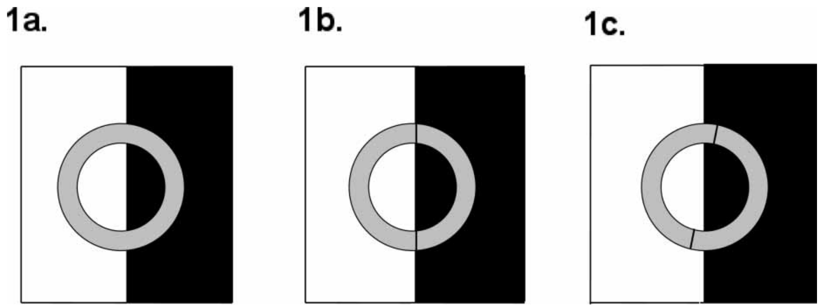

Figure 1. *Black-and-white version of Benussi's rings (Koffka, 1935, p. 134). In Figure 1a the ring appears a uniform grey; in Figure 1b contrast relationships appear; in Figure 1c contrast relationships appear but extend into the opposing background.*

background (Figure 1B). In a further demonstration, it is possible to divide the ring into two halves, but not precisely at the border of the abrupt change in background luminance. In this instance, each half-ring protrudes a little into the adjacent background. The brightness does not appear to change at the border of the two backgrounds as might be expected; instead, each half-ring remains homogeneous despite its protrusion into the opposite background (Figure 1C). This suggests that the visual system tends to simplify the visual region within a formed unit (each half-ring) by processing a single average luminance for the area within unit boundaries.

It was suggested by Koffka (1915, p. 47) that the visual system tends to use abrupt changes in luminance as contours that define individual units. Perceived brightness is more homogeneous within a perceptual unit than the changing physical intensities in light across the surface would warrant. This suggests that a perceptual simplification has occurred. For example, suppose we have a horizontally oriented band that builds gradually in luminance (gets lighter) from left to right, but is broken into smaller rectangles by vertical stripes the same colour as the background that hide portions of the changing gradient, as seen in Figure 2A. Each smaller rectangle appears to have a different, fairly homogeneous, level of brightness, indicating that perceptual simplification has taken place within the smaller rectangles; the brightness appears to change in discrete steps across the band. As a result, two regions of the gradient appear to be more different when they appear within different "steps" than they do when they are within an unbroken gradient (Figure 2B). Consequently the change in brightness from left to right is more marked in Figure 2A. While the change in brightness in Figure 2B is visible, the perception is unstable, and one cannot see differences between closely adjacent regions.

We wanted to explore whether the auditory system also tends to perceptually simplify regions that are treated as units. To do this, we looked at whether the auditory system would respond to a change in the spectral balance of a sound over time in a way analogous to the visual system's response to a change in luminance across a region of space.

Unit-derived perceptual effects within the auditory system have had relatively little study (Nabelek, Nabelek, & Hirsh, 1973a, 1973b; Nakajima, 1996; Royer & Robin, 1986). It has been suggested that the auditory system may use neural onset responses occurring at points of sudden amplitude change to recognize the occurrence of new events and subsequently points at which to begin new analyses and processing of

Figure 2. *Figure 2a shows a continuous luminance gradient broken by vertical lines that cause the formation of smaller subunits. Figure 2b shows an unbroken continuous luminance gradient.*

new informational groups (Bregman et al., 1994b; Hafter & Buell, 1990). In this way, the perception of the physical stimulus would be influenced by the auditory system's detection of a unit boundary.

One might ask, then, whether the auditory system provides a similar change in the perception of a continuous spectrum of sound when it is broken by auditory contours. It has been suggested by Bregman (1994, p. 72) that the auditory system forms a perceptual representation of a given unit by processing only the information that is confined within its contours.

In the present experiment, it was expected that using silences to break a larger region of smoothly changing timbre into smaller regions of change would cause the smaller regions to be perceived as individual units, and that the changes within these units would be simplified. This should allow listeners to detect a change in the timbre earlier than when the smaller units are absent.

The purpose of the present experiment was to introduce onsets into a changing gradient of timbre. Timbre has been described by Wessel (1979) as the quality of sound that is typically divorced conceptually from pitch and loudness. Studies of timbre (Grey, 1977; Grey & Gordon, 1978; Plomp, 1970) have demonstrated it to be a multidimensional property of sound. The present study will manipulate one salient dimension of timbre, the "spectral centre of balance", or spectral centroid. To do this, the fundamental remains constant, and the relative intensities of higher and lower harmonics are changed gradually across time. It was expected that the introduction of a series of silences into the gradually changing timbre would, by producing a succession of onsets, cause smaller regions to be formed that would be treated as individual units. Let us symbolize a gradient of timbre by the range of real numbers from 1–7 (shown in Figure 3a), where an increase in numbers represents a change in spectral balance favouring higher harmonics as we progress from 1 to 7. Suppose the section 3–5 is removed and is replaced by a silence leaving 1–3,_,5–7 (Figure 3b). We assume that 1–3 and 5–7 are now perceived as distinct units. Since, according to the simplification hypothesis,

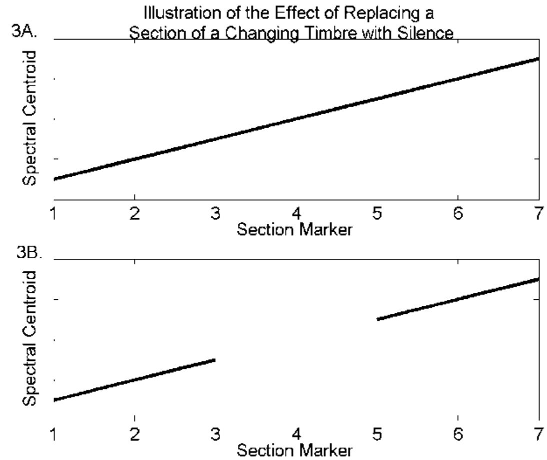

Figure 3. *Figure 3A illustrates a sound with a continuously changing spectral centre of balance, or spectral centroid, from section marker 1 to section marker 7. Figure 3B illustrates the same changing sound with the region between section marker 3 and section marker 5 replaced with silence.*

the properties of any region in a unit are influenced by the properties of other regions within the same unit, the region just before 3 is now influenced by the region from 1–2 causing it to be perceived as perceptually "duller" in timbre than before, and the region just after 5 is now influenced by 6–7, making it perceptually brighter than before. The net effect is that the difference in timbre between the region just before 3 and the region just after 5 is greater after the silence has replaced the region from 3–4. In our experiment, the introduction of a series of silences was expected to assist a listener in perceiving a change in the timbre of the sound at an earlier point in time than if the gradient appeared continuous.

This decrease in the time necessary to detect a change was expected as a consequence of the smaller units tending to appear more different from one another than do the corresponding regions in an unbroken changing gradient. The spectral variable manipulated in this study can be referred to as "brightness" (Wessel, 1979). Using this terminology, we expected that if listeners were asked to detect a change in the brightness of a gradually changing sound, A, they should respond earlier when a series of silences was introduced as interruptions within A.

Three different interruption types were tested. In *Condition 1*, no interruptions were introduced, and A appeared continuous. In *Condition 2*, silences were introduced, and A appeared "broken". In *Condition 3*, noise bursts were inserted in place of the silences of Condition 2. Condition 3 was included in the experimental design in order to provide a condition that included interruptions that were not expected to create smaller units. Rather, it was expected that the perception of Condition 3 would appear continuous and more closely resemble the perception of Condition 1 (Ciocca & Bregman, 1987; Dannenbring & Bregman, 1976; Warren, Obusek, & Ackroff, 1972). Condition 3 was required in order to observe whether any change in the listeners' responses could be attributed solely to the presence of interruptions, possibly allowing listeners to compare regions (the 3 and 5 in the sequence 1−3,_,5−7) that were further apart in time, and hence in timbre, instead of changes in the perceptual properties that arose due to the formation of smaller auditory units.

It was expected that listeners' reaction times should be fastest in Condition 2 in which silences, causing onsets and producing units, were introduced. In both Conditions 1 and 3 no unit-forming cues were given, and the effects of the formation of smaller units were not expected to occur. Therefore, we predicted that Conditions 1 and 3 would produce slower reaction times than Condition 2.

Method

Listeners

The listeners were 26 adults who reported normal hearing. They were recruited from a university student and faculty population and were compensated for participating. The data of two subjects were discarded due to failure to follow the instructions supplied by the experimenter. Therefore, the data of 24 subjects were included in the analysis.

Stimuli

Four fundamental frequencies (160, 165, 170, and 175 Hz) were used (on different trials) in the three testing conditions. For each fundamental, two harmonic spectra were mixed in varying proportions to create a sound that gradually changed from a sound with more intense low harmonics, producing a "dull" sound, to a sound with more intense "high" harmonics, producing a "bright" sound. Both Harmonic Spectrum 1 and Harmonic Spectrum 2 consisted of Harmonics 5 through 14 of the fundamental used for that trial. Harmonic Spectrum 1 had its maximum amplitude at Harmonic Number 5 with the amplitudes of the harmonics declining linearly down to Harmonic Number 14, such that Harmonic Number 14 was 10% of the amplitude of Harmonic Number 5. This spectrum produced a "dull" sound. In contrast, Harmonic Spectrum 2 had its maximum amplitude at Harmonic Number 14 with the amplitudes of the harmonics declining linearly down to Harmonic Number 5, such that it was 10% of the amplitude of Harmonic Number 14. This spectrum produced a "bright" sound. The two spectra were then mixed together in different proportions over time to create a sound that continuously changed in its spectral energy distribution. The summed relative amplitudes of an individual harmonic at any time were always equal to one. Figure 4 shows a representation of the amplitude of the individual harmonics across time as the two spectra are mixed together in a linear combination. All stimuli consisted of a 5-s sound that always began with Spectrum 1 and gradually changed from a pure Spectrum 1 to a mixture of Spectrum 1 and Spectrum 2 during the first 4 seconds of each stimulus, remaining at the 4-s value for the last 1 s of the stimulus. There was no change during the final second of the stimulus in order to allow for any latency in a listener's response. In testing, five different levels of spectral change were used (40, 50, 60, 70, and 80%). Each percentage is represented in Figure 4 by a dotted line indicating the relative harmonic amplitudes for the specified percentage after a 4-s period of change. Further, the first quarter of

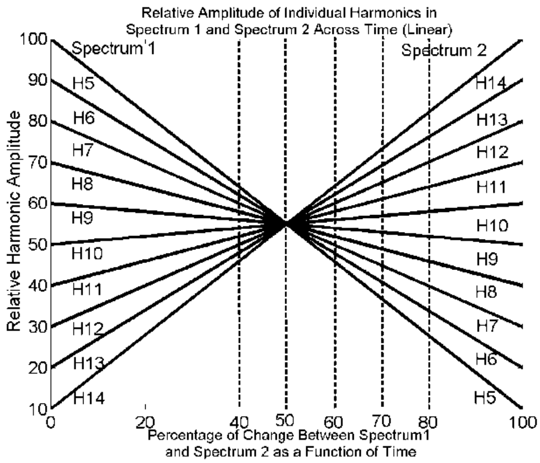

Figure 4. *Figure 4 shows the relative amplitude of the individual harmonics of Spectrum 1 and Spectrum 2 as they are mixed across time in a linear combination. The level of each harmonic is represented as a percentage of the amplitude of H14 (Harmonic Number 14) and H5 (Harmonic Number 5) for Spectrum 1 and Spectrum 2, respectively.*

a sine-wave function (0 to 90 deg) was used to control the rate of change of the amplitude multipliers of Spectrum 2. The effect of this function on the overall rate of change as a function of time is shown for the five different levels of spectral change in Figure 5. This "warping" was done in

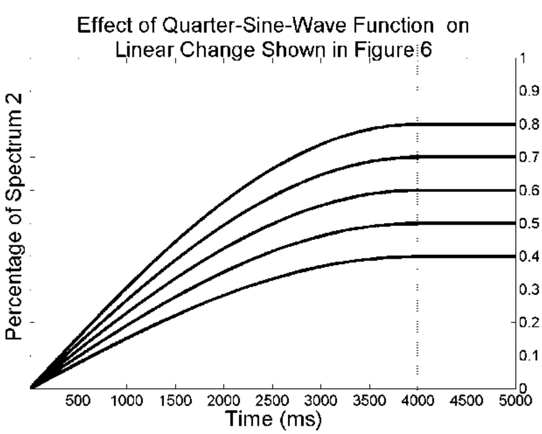

Figure 5. *Rate of change levels used for stimuli are shown following the addition of the quarter-sine-wave function used to warp the perceived rate of change. The dotted line indicates the point after which no additional timbre change occurs.*

order to better approximate a linear perception of the timbral change than would occur with a physical linear change between the two spectra (the quarter-sine-wave function caused the change in the spectrum to become increasingly slower). This type of transition was introduced because data from pilot studies showed that the distribution of reaction times was highly clustered towards the end of the 4-s stimulus when a linear function was used as a means of controlling the amplitude of Spectrum 2. In these linear-change conditions, listeners appeared to perceive the greatest part of the timbral change as occurring during the last portion of the stimulus. The quarter-sine-wave function was introduced in order to distribute the perception of change more evenly throughout the stimulus duration. While this function did not achieve perfect perceptual linearity, it did provide a considerably more even distribution of reaction times across the 5-s duration. Additionally, this "warping" was present in all conditions and therefore would not be expected to have contributed to the differential effect of the conditions.

The overall change is described by the following equation:

$$S(t) = x_1\beta(t) + (1 - \beta)(t)x_2,$$

where

$$\beta(t) = M\sin(t), \quad \text{for } t \in \left[0, \frac{\pi}{2}\right]$$

$S(t)$ represents the weighted sum of the Spectrum 1 and Spectrum 2 signals across time (t). This value is a product of the amplitude of the harmonics of Spectrum 1 and a weighting function β summed with the amplitude of the harmonics of Spectrum 2 and the complement to β, or $(1 - \beta)$. The weighting function β changes as the sine of (t) for all values between 0 and $\pi/2$ producing the first quarter of a sine-wave. In the testing conditions it was never the case that the change from Spectrum 1 to Spectrum 2 was 100%, but, rather, listeners heard five percentages of change between Spectrum 1 and Spectrum 2 that were always less than 100%. A scaling percentage, M, with a value of .40, .50, .60, .70, or .80, was used to alter the amount of change that occurred during the duration of the trial. It is helpful to think of the value of M as controlling the maximum allowable amount of Spectrum 2. M is the maximum permitted transition between Spectrum 1 and Spectrum 2 on a given trial. Whenever M is less than 1.0, there is still, at 4 s, a substantial amount of Spectrum 1 present (see dotted lines in Figure 4). After 4 s there is no further change in the spectrum of the signal; the final value is held for an additional second in order to allow for any latency in the listener's response. (The effects of the quarter-sine-wave function on an underlying linear change for each level of M are shown in Figure 5; however, for simplicity, further discussion of the stimulus refers only to the underlying linear change before "warping".)

Additionally, for every combination of fundamental frequency and maximum change, three interruption types were constructed. All were derived initially from one of the possible combinations of fundamental frequency and maximum change that had produced a continuous gradually changing timbre spectrum.

All sounds were presented in each of three different interruption types. *Condition 1—continuous*: This was a continuous sound gradually changing in timbre, which contained no interruptions (as represented by Figure 6, panel 1). This condition represents a stimulus without interruptions. *Condition 2—interrupted by silence*: This was a continuous sound gradually changing in timbre, which contained 20-ms silences interrupting every 290 ms (as represented by Figure 6, panel 2). Additionally, a 10-ms onset and offset were added to the overall stimulus. This produced an alternating sequence of 16 tones and 17 silences across the entire stimulus duration of 5 s. The spectrum continued to change smoothly despite these interruptions. Silences were used to introduce auditory contours. *Condition 3—interrupted by noise*: This was similar to Condition 2, but with noise bursts inserted during the silences (as represented by

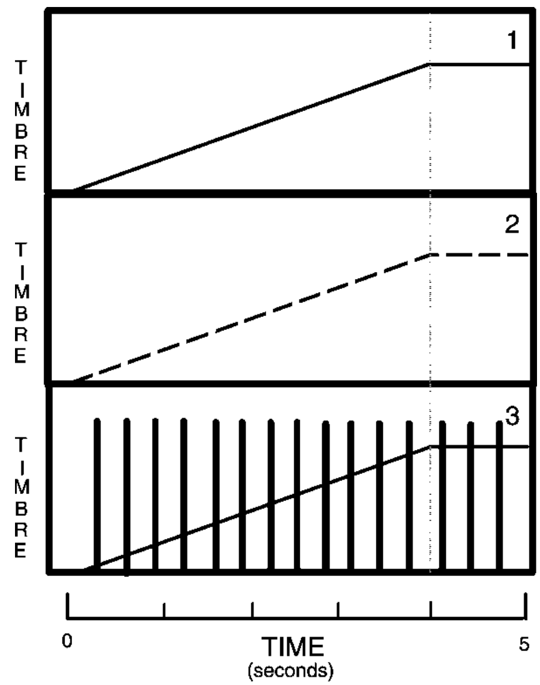

Figure 6. *Stimulus design for the three different interruption types. Panel 1: Continuous. Panel 2: Interrupted by silence. Panel 3: Interrupted by noise.*

Figure 6, panel 3). This produced an alternating sequence of 16 tones and 17 noise bursts across the entire stimulus duration of 5 s. This condition was used as a control condition that presented interruptions in the same manner as the silences in Condition 2, but, instead, was expected to produce a perception in which the tone appeared continuous (as in Condition 1) beneath the interrupting noises. In previous studies it has been found that listeners will perceive tones as maintaining continuity over breaks of up to 300 ms filled with loud noise (Dannenbring & Bregman, 1976; Nakajima & Sasaki, 1996; Warren et al., 1972). The breaks used in the stimuli of the present study were 30 ms in duration, and it was expected that the tone would appear to maintain a perception of continuity. The perceived continuity of the sounds was verified during the experiment. Upon completing the test sessions, all subjects reported that they had perceived the tone as continuous in the conditions with noise bursts present.

All stimuli had a duration of 5 s with 10-ms onsets and offsets. The 10-ms attack and decay envelopes of all the 5-s stimuli were controlled by a linear function. Stimuli were presented at 70 dBA. The noise was presented at 80 dBA. White noise was bandpass filtered with a bandwidth of 2,100 Hz around a centre frequency of 1,500 Hz (with −3-dB cutoff points at 450 and 2,550 Hz). This filter passed at full intensity all spectral components of the noise burst that fell within the frequency range of the spectral components of the complex tone. In Condition 2, a 5-ms onset and offset were added to the tones immediately before and after an interrupting silence as displayed in Figure 7A. In Condition 3, a 5-ms onset and offset were added to each noise burst so that it overlapped the onset and offset of the immediately preceding and following tones as displayed in Figure 7B.

On a single trial the listener heard only one 5-s condition. The conditions were divided into four blocks by fundamental frequency: 160, 165, 170, or 175 Hz. (Small variations in fundamental frequency were made in order to avoid any results due to the use of a single frequency. We were not testing the effects of fundamental frequency; however, we did want to avoid any potential habituation created by the use of a single frequency.) A block consisted of five replications of each of 15 conditions (each level of *M* within each condition type), which were randomly presented to produce 75 trials. Each stimulus lasted 5 s. Between the completion of a stimulus and the onset of the next stimulus a 2-s pause was inserted. The total duration of each block was slightly longer than 8 min. Listeners were given 5-min breaks between the four blocks, making the total duration of the experiment approximately 50 min.

Procedure

Listeners were asked to detect a change in the "brightness" quality of each presented sound. Before beginning the testing sessions, the experimental task was explained to all listeners through both verbal and written instructions. Diagrams of the condition types were provided, and "brightness" was explained qualitatively to the listeners as "tinniness". During training, listeners heard repeated presentations of each of the three interruption types: Condition 1—continuous, Condition 2—interrupted by silence, and Condition 3—interrupted by noise, at the minimum and maximum levels of spectral change that would be used in the main experiment. During the presentation of the 40% stimulus ($M = .4$), they were told that it represented a small change in the brightness of the sound, and during the presentation of the 80% stimulus ($M = .8$), they were told that it represented a larger change in the brightness of the sound. In this initial training session, most listeners were

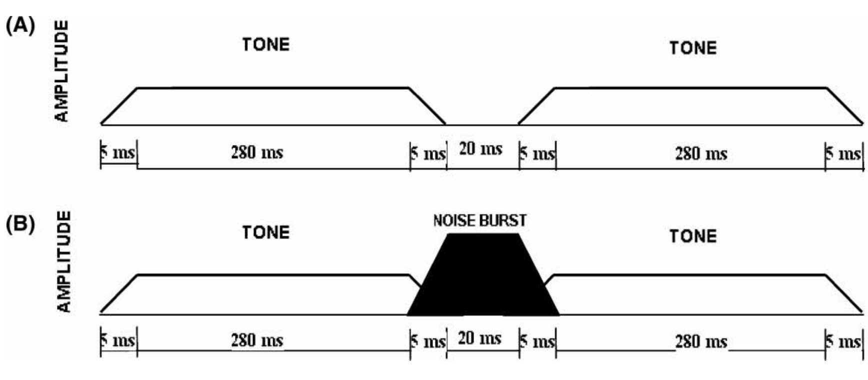

Figure 7. *7A: Stimulus construction for Condition 2—interrupted by silence. 7B: Stimulus construction for Condition 3—interrupted by noise.*

unable to perceive a change in the sound quality of the 40% stimulus condition. When the stimulus of 80% change was presented, all but one listener were able to recognize the change in the quality of the sound. Those that were able to perceive the change in quality were instructed to consider the quality change that they perceived to be a change in brightness, as defined by the experimenter.

Each listener was then told that the sound might or might not change in brightness during each trial of the testing session. If a change in the brightness quality of the sound was detected, the listener was instructed to respond immediately by pressing the keyboard spacebar. If a change in brightness was not detected, the listener was instructed not to respond and to wait for the next trial. After receiving the instructions, each listener was given a practice block of one randomly ordered presentation of 15 conditions. This block was presented at 170 Hz, the highest of the four fundamental frequencies used in the testing blocks. Once the training session was complete the listeners were tested in four separate 8-min blocks with 5-min pauses between blocks. The presentation order of the four testing blocks (each with a different fundamental) was counterbalanced. It required six listeners for the completion of a full testing of all possible orderings of blocks. A single block consisted of 75 trials. Each trial was one of the three interruption types that had been presented to the listeners during the prior training session. During the testing session the computer recorded the listeners' responses as they pressed the keyboard spacebar when they detected a change in brightness. (At this point it is important to note that all trial conditions contained a spectral change. While inclusion of conditions of "no change" would have enabled a measure of the observer's accuracy and potential bias, we chose to exclude them due to reports by the observers and observations by the experimenters that unchanging conditions appeared to change in an illusory manner and get "duller" perceptually within the context of the other experimental stimuli. This phenomenon was not the variable of interest, and its inclusion would interfere with the desired measurement of brightness change. Therefore, only conditions of true spectral change were tested.)

Apparatus

Listeners were tested in an audiometric test chamber and were presented with stimuli binaurally over Sennheiser HD414 headphones. The experiment was conducted on a PC-compatible 486 computer. All stimuli were synthesized prior to the experiment at a sampling rate of 20 kHz using the MITSYN programming software Version 8.1 (Henke, 1981). The stimuli were then presented to listeners through a program written in the M.A.P.L.E. language, which also recorded and timed the responses of the listeners. The intensities of the stimuli coming from the headphones were measured using a General Radio Company Type 1565-B sound-level meter set at "A" weighting and "fast" reading using a flat-plate coupler. The route mean square amplitudes of all stimuli (different combinations of Spectrum 1 and Spectrum 2) were equalized at the time of synthesis.

Results

Interruption Type (Conditions 1, 2, and 3). The purpose of the present experiment was to determine whether in Condition 2 (interrupted by silence), which should allow the formation of smaller auditory units, the listener could detect a change in the brightness level of the stimulus earlier than in Conditions 1 or 3 (continuous and interrupted by noise), which should not produce smaller units. For each trial, the time it took for the listener to detect and respond to a change in the brightness of the stimulus was recorded. The recorded response time was restricted to be at least 0.001 s. Response times of 5 s or longer (including nonresponses) were recorded as 5 s. As shown in Figure 8, listeners detected a change in brightness significantly earlier in Condition 2 than in either Conditions 1 or 3. Furthermore, Figure 6 shows that this was true for all levels of spectral change.

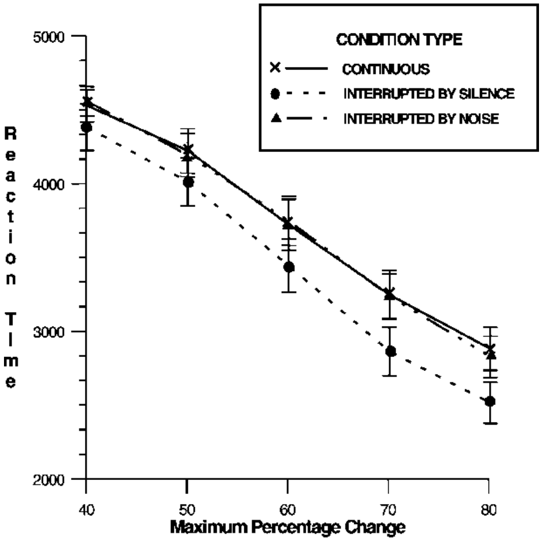

Figure 8. *Reaction time vs. maximum percentage change averaged across fundamental frequency. Standard error of estimation is indicated for each condition type and level of percentage change.*

Mean recorded response times were longest for the conditions expected to be the most difficult for detecting brightness changes. Response times of 5 s and longer (i.e., those coded as 5.000) were strongly associated with the type of stimulus: A chi-square test for association with Condition 1, Condition 2, and Condition 3 yields $\chi^2(2) = 15.0$, $p < .001$ (Table 1).

The mean reaction times for each condition shown in Figure 8 are combined across fundamental frequency. Each mean represents the average reaction time taken by listeners to detect a change in the "brightness" level of the sound at a fixed value of *M*.

Table 1. *Total number of 5.000 responses at each level of maximum percentage change and condition type*

	Condition type		
Maximum percentage change	*1*	*2*	*3*
40	347	340	347
50	274	244	267
60	181	131	165
70	86	40	71
80	41	17	30

Note. Each cell represents the number of nonresponses out of 480 trials. In total there were 7,800 trials across all levels of maximum percentage change and condition type.

To test whether the listener could detect a change in the brightness level of the stimulus earlier in Condition 2 (interrupted by silence) than in Conditions 1 or 3 (continuous and interrupted by noise), which should not produce smaller units, we used a nonparametric permutation test (Manly, 1997). This test stratified on all other variables: subject, fundamental frequency, and value of *M*. The test statistic was the amount by which the mean response time for the broken condition was exceeded by the smaller of the means of the response times for the other two conditions. If detection is indeed easiest for the broken condition, we would expect this difference to be positive, which it is: The observed value is 270 ms. If detection is just as easy in one of the other conditions, we would expect this difference to be zero. If detection is easier in one of the other conditions, we would expect this difference to be negative. Under the null hypothesis that there is no difference in the difficulty of detecting brightness change in the three conditions, the labelling of the three responses—holding subject, fundamental frequency, and value of *M* fixed—is arbitrary. That is the basis of the permutation test: For each subject, fundamental frequency, and value of *M*, we randomly assigned the three observed responses to the three conditions. Once we had randomly assigned the three responses in all 480 strata, we calculated the value of the test statistic:

min(mean response time for Condition 1, mean response time for Condition 3) − mean response time for Condition 2

We approximated the null distribution of this statistic by repeating this randomization of the 7,200 responses and calculating the corresponding value of the test statistic 10,000 times. This permutation test is nonparametric: It is valid whether or not the distribution of recorded response times is approximated well by a normal distribution. Moreover, the stratification within the test controls for the effects of subject, fundamental frequency, and value of *M*. The largest value of the test statistic in the 10,000 random replications was 109 ms, far smaller than the observed value of the statistic, so the estimated p value is less than $1/10{,}000 = .0001$.

The results shown in Figure 8 suggest that the difference between Condition 2 and Conditions 1 and 3 became increasingly larger as the degree of spectral change increased. It is necessary to note that the mean reaction times for all fundamentals at all levels of spectral change were faster in Condition 2, as shown in Figure 8. The difference between Condition 2 and Conditions 1 and 3 is progressively greater at increasing levels of spectral change (Figure 8). The trend in Figure 8 suggests that the observed increase in detectability of a brightness change during Condition 2, relative to Conditions 1 or 3, gets larger with increasing values of *M* and, subsequently, perceived change in brightness.

Spectral change. It is important to assess the presence of an inverse relationship between the degree of spectral change (and hence the rate of spectral change) and the reaction time necessary for listeners to detect a change in the brightness of the sound. As the level of spectral change increased, listeners' reaction times decreased significantly (as seen in Figure 8). As discussed earlier, a coding of 5,000 ms was used to indicate a condition of "nonresponse" in the change-detection task. A chi-square test for association with the value of *M* as .4, .5, .6, .7, and .8 (maximum percentage change between Spectrum 1 and Spectrum 2) yields $\chi^2(4) = 1{,}214.9$, $p < .0001$ (Table 1). This result, combined with observation of the trend in Figure 8, suggests that listeners were more able to detect a change in the brightness level of the sound as the maximum percentage change between Spectrum 1 and Spectrum 2 increased and, more importantly, that listeners were sensitive to the variable being measured in the task. Additionally, this test supports the use of nonresponse trials as indicative of the listener's difficulty in detecting a change and justifies their inclusion in the permutation test.

Discussion

Results of the present experiment demonstrate that the formation of smaller units in a longer period of a changing sound enables listeners to recognize a change in the timbre of the sound faster than when the sound appears to change continuously. This suggests that the formation of units causes the timbral difference between parts of the overall transition to be more pronounced when those parts are packaged in separate units. Further, we speculate that it is this change in the perception of the sound that enables comparisons between adjacent units, which, in turn, lead to the observed decrease in reaction time. A condition where the gradient appears continuous is apparently lacking the necessary reference points to allow for a comparison between adjacent regions. For example, the addition of noise bursts to the changing spectrum did not create smaller perceptual units, and, in turn, the results of the noise conditions were more similar to the results of the continuous conditions. This suggests that the detection of unit boundaries in the auditory system's analysis of a sound may significantly affect the perception of information within a unit.

It is important to recognize two distinct processes that may be operating when unit boundaries, such as onsets and offsets, are introduced into a sound. In Process 1, their introduction creates a change in the perceptual representation by directly accentuating the perceptual features at the moments of onset and offset. In contrast, Process 2 uses the introduction of an onset and offset into a sound to create a unit. The formation of the unit causes any change that occurs within it to be perceptually minimized.

In both Process 1 and Process 2, the region within the introduced onset and offset can suffer a perceptual loss of detail. Process 1 causes this perceptual change by the accentuation of the information in the beginning and ending points relative to the information that appears in the middle. In Process 2 the perceptual shift is more dependent on the properties of the unit arising from the onsets and offsets, rather than a change in perception that is primarily a product of the onsets and offsets themselves.

Process 1. In Process 1 it is important to consider the formation of onsets and offsets from bracketing silences and their role in the perception of a sound. We are unaware of research that looks at the perceptual effects caused by the temporal position of the parts of a single unit; however, it may be relevant to consider the role of temporal positioning in the perception of a series of discrete sounds (Bregman & Rudnicky, 1975; Divenyi & Hirsh, 1974, 1978; Warren & Bashford, 1993; Watson, Kelly, & Wroton, 1975; Watson, Wroton, Kelly, & Benbassat, 1974). Consistently, elements in the first and last serial positions appear to have richer perceptual representations, which allow for increased identification and memory. Conversely, those imbedded in the middle of a pattern appear to experience a loss in individual perceptual identity (Bregman, 1994).

It is possible to think of the onsets and offsets of a unit as the first and last of a series of positions within the sound and, therefore, as information that may have an accentuated perceptual representation (Bregman, 1994, pp. 110–112). For instance, the results of the present experiment suggest that the onsets introduced by the silences provide listeners with information that enables them to recognize a change in the timbral gradient faster than if the onsets were not present. Process 1 could have produced these results by perceptually accentuating the onsets of the smaller regions of sound in comparison to the middle regions. This would allow listeners to make point-to-point comparisons of the onsets of adjacent units and, in turn, to detect a change in the timbral gradient earlier than without the points of reference.

Physiological evidence indicates that certain cells in the auditory system are selectively responsive to sudden rates of change (i.e., rapid onsets and offsets; Chimoto, Kitama, Qin, Sakayori, & Sato, 2002; Phillips, 1988) and in some instances to onsets with specific rates of rise (Olsen, 1994). It is possible that the accentuation of onsets, implied by Process 1, is the result of the activation of "onset" cells found throughout the auditory pathways. This information may receive an enhanced analysis by the auditory system, and it may be a mechanism that could achieve the onset accentuation required for Process 1. Response of such cells may serve as a cue used by the auditory system to engage a process that places greater emphasis on the acoustic information present at the time of the cell's activation.

Process 2. Process 2, in comparison, increases the uniformity of the region defined as a unit. Therefore it suggests a perceptual description of the region within a unit that is simpler than the descriptions of the acoustic properties of the constituent parts of the unit region. In turn, adjacent units would appear more different. Process 2 would also account for the present results, however, in contrast to Process 1, it would imply that listeners were making a unit-to-unit comparison of adjacent regions rather than an onset-to-onset comparison.

Additionally, it is not necessary that Process 1 and Process 2 are mutually exclusive, and they could occur simultaneously. In other words, one process could give a stronger perceptual weighting to the starting and ending points of a sound, while the other process created properties that were unique to the formed unit, at the same time reducing the perceptual representation of changes within the unit. The formation of the unit appears to affect the qualitative perception of the region between the onsets and offsets by perceptually simplifying the middle region.

If both processes are active in the perception of a unit of sound, a further investigation of their interaction with changes in stimulus variables, such as duration, amount of change, rate of onset, and the distance of the onset from the nearest temporally adjacent sounds, should be carried out. For instance, the strength of the properties arising due to either Process 1 or Process 2, the latter causing the formation of an auditory unit, may vary, as observed by Royer and Robin (1986) and Hafter and Buell (1990), continuously and be dependent on the isolation of the onset from the surrounding sounds.

Possibly one of the most interesting implications of the present experiment is a fairly direct and general practical application. When one is confronted with the problem of how best to convey the most information from a time-varying sound, it appears that breaking the sound into smaller units may actually increase the amount of information that the listener is able to extract in a given period of time. This is slightly counterintuitive in that it suggests that a sound broken into units, which in actuality lacks portions of information, may in fact be able to convey different, more useful, information than a continuous sound that presents the information in its entirety. The creation of smaller units within a larger region of change appears to allow subtler details of the changing region to emerge. This may reflect an important role of onset detection in information processing.

Original manuscript received 23 July 2004
Accepted revision received 28 September 2004
PrEview proof published online 28 June 2005

REFERENCES

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: Bradford Books, MIT.

Bregman, A. S., Ahad, P. A., & Kim, J. (1994a). Resetting the pitch-analysis system. 2. Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones. *Journal of the Acoustical Society of America, 96*, 2694–2703.

Bregman, A. S., Ahad, P. A., Kim, J., & Melnerich, L. (1994b). Resetting the pitch-analysis system. 1. Effects of rise times of tones in noise background or of harmonics in a complex tone. *Perception and Psychophysics, 56*, 155–162.

Bregman, A. S., & Rudnicky, A. I. (1975). Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance, 1*, 263–267.

Chimoto, S., Kitama, T., Qin, L., Sakayori, S., & Sato, Y. (2002). Tonal response patterns of primary auditory cortex neurons in alert cats. *Brain Research, 934*, 34–42.

Ciocca, V., & Bregman, A. S. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception and Psychophysics, 42*(5), 476–484.

Dannenbring, G. L., & Bregman, A. S. (1976). Perceived auditory continuity with alternating rising and falling frequency transitions. *Canadian Journal of Psychology. 30*, 99–114.

Divenyi, P. L., & Hirsh, I. J. (1974). Identification of temporal order in three-tone sequences. *Journal of the Acoustical Society of America, 56*, 144–151.

Divenyi, P. L., & Hirsh, I. J. (1978). Some figural properties of auditory patterns. *Journal of the Acoustical Society of America, 64*, 1369–1385.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America, 61*, 1270–1277.

Grey, J. M., & Gordon, J. W. (1978). Effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America, 63*, 1493–1500.

Hafter, E. R., & Buell, T. N. (1990). Restarting the adapted binaural system. *Journal of the Acoustical Society of America, 88*(2), 806–812.

Henke, W. L. (1981). MITSYN: A coherent family of command-level utilities for time signal processing [Computer program]. Belmont, MA: Author.

Koffka, K. (1915). Zur Grundlegung der Wahrnehmungspsychologie. Eine Auseinandersetzung mit V. Benussi. *Zeitschrift fur Psychologie, 73*, 11–90.

Koffka, K. (1935). *Principles of gestalt psychology.* New York: Harcourt, Brace & World.

Manly, B. F. J. (1997). *Randomization, bootstrap and Monte Carlo methods in biology.* New York: Chapman & Hall/CRC.

Nábêlek, I. V., Nábêlek, A. K., & Hirsh, I. J. (1973a). Pitch of sound bursts with continuous or discontinuous change of frequency. *Journal of the Acoustical Society of America, 63*, 1305–1319.

Nábêlek, I. V., Nábêlek, A. K., & Hirsh, I. J. (1973b). Pitch of tone bursts of changing frequency. *Journal of the Acoustical Society of America*, *48*, 536–553.

Nakajima, Y. (1996). *A simple grammar for auditory organization: Streams, events and subevents.* Paper presented at the Proceedings of the XXVI International Congress of Psychology, Montreal, Canada.

Nakajima, Y., & Sasaki, T. (1996). An illusory reconstruction of auditory elements. *Journal of the Acoustical Society of America*, *100*, (4), 2751.

Olsen, J. F. (1994). Sensitivity of medial geniculate neurons in the squirrel monkey to rate of rise. *Journal of Neuroscience Abstract, 20*, 321.

Phillips, D. P. (1988). Effect of tone-pulse rise time on rate-level functions of cat auditory cortex neurons: Excitatory and inhibitory processes shaping responses to tone onset. *Journal of Neurophysiology, 59*, 1524–1539.

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and detection of hearing*. Leiden, The Netherlands: A. W. Sijthoff.

Royer, F. L., & Robin, D. (1986). On the perceived unitization of repetitive auditory patterns. *Perception and Psychophysics, 39*, 9–18.

Warren, R. M., & Bashford, J. A. (1993). When acoustic sequences are not perceptual sequences: The global perception of auditory patterns. *Perception and Psychophysics, 54*(1), 121–126.

Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science, 176*, 1149–1151.

Watson, C. S., Kelly, W. J., & Wroton, H. W. (1975). Factors in the discrimination of tonal patterns. II. Selective attention and the learning under various levels of stimulus uncertainty. *Journal of the Acoustical Society of America, 60*, 1176–1188.

Watson, C. S., Wroton, H. W., Kelly W. J., & Benbassat, C. A. (1974). Factors in the discrimination of tonal patterns. I. Component frequency. Temporal position, and silent intervals. *Journal of the Acoustical Society of America, 57*, 1175–1185.

Wessel, D. M. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3*, 45–52.