

## The effect of tone duration on auditory stream formation

MICHAEL W. BEAUVOIS

Laboratoire de Psychologie Expérimentale, CNRS, Université René Descartes, Paris, France  
and IRCAM, Paris, France

In a study in which the effect of tone duration on the formation of auditory streams was investigated, subjects were presented with 15-sec alternating pure-tone sequences (ABAB...) and were asked to orient their attention over the duration of the sequence toward hearing either a temporally coherent or a segregated percept. At stimulus offset, the subjects indicated whether their percept at the end of the stimulus had been that of a temporally coherent ABAB trill or that of segregated A and B streams. The experimental results indicated that the occurrence of stream segregation increases as (1) the duration of the A and B tones increases in unison and (2) the difference in duration between the A and B tones increases, with the duration differences between the tones producing the strongest segregation effects. A comparison of these experimental results with those of other studies strongly suggests that the time interval between the offset and onset of consecutive tones in the same frequency range is the most important temporal factor affecting auditory stream formation. Furthermore, a simulation of the experimental results by the Beauvois and Meddis (1996) stream segregation model suggests that both the tone duration effects reported here and Gestalt auditory grouping on the basis of temporal *proximity* can be understood in terms of low-level neurophysiological processes and peripheral-channeling factors.

Auditory stream segregation, in its simplest form, occurs when listeners are presented with a continuous alternating-tone sequence composed of two pure tones of different frequencies (A and B). If the tone-repetition time (TRT, or the time interval between the onset of consecutive tones in the sequence) is long and the frequency separation between the tones ( $\Delta f$ ) is small, listeners hear a connected series of tones, or a musical trill (ABAB...), a condition known as *temporal coherence* (van Noorden, 1975). However, if TRT is decreased, or if  $\Delta f$  is increased, the tones composing the trill segregate into two parallel sequences, or *streams*, one low and one high in pitch (an A-stream and a B-stream). Here, the listener's attention can be focused on only one stream at a time, and both streams appear to have a longer periodicity, equal to twice the TRT. The most important factors influencing stream formation are  $\Delta f$ , TRT, and attentional set (van Noorden, 1975). However, a number of studies have found that tone duration (TD) also has a significant effect on sequential grouping in pure-tone sequences. Van Noorden, using himself as a subject, kept the TRT of an ABAB sequence constant and increased the TD of the A and B tones in unison. His results indicated that stream segregation increased as TD increased. In a study by Hartmann and Johnson (1991), an interleaved-melody paradigm was used

to investigate whether peripheral-channeling or source-grouping factors were responsible for promoting stream segregation. Peripheral channeling was defined as being established in the auditory periphery and was considered to be tonotopic (frequency) or lateral (left-ear/right-ear) in nature. Source grouping was defined as a higher level process that grouped tones into streams when it seemed likely that the tones came from the same sound source. Hartmann and Johnson's data not only showed that peripheral-channeling factors were of "paramount importance" in the formation of auditory streams, but also that there was a secondary effect created by a source-grouping factor—TD. They found that if the TD of one set of tones was increased relative to the other, stream segregation increased and that this occurred even when the lengthened tones overlapped the other set of tones in time.

In this paper, the results are presented from a study that investigated the effect of TD on the formation of auditory streams in ABAB pure-tone sequences by systematically varying the individual durations of the A and B tones ( $TD_A$  and  $TD_B$ ). The study itself had four main goals. The first goal was to confirm van Noorden's (1975) findings by using a larger subject pool to examine the segregation effects produced when  $TD_A$  and  $TD_B$  were increased in unison, so that  $TD_A = TD_B$  (global-TD). The second goal was to investigate more fully the secondary effect of TD that was found by Hartmann and Johnson (1991), by using a greater number of differences between  $TD_A$  and  $TD_B$  ( $\Delta TD$ ). The third goal was to determine which of the two factors, global-TD or  $\Delta TD$ , has the greater effect on auditory stream formation. The fourth goal was to gather experimental data in a form that

---

This research was supported by a grant from the Fyssen Foundation, Paris, France. Thanks go to Chris Darwin, Brian Moore, and an anonymous reviewer for their comments on an earlier version of this paper. Correspondence may be addressed to M. W. Beauvois, Department of Psychology, Leicester University, University Road, Leicester LE1 7RH, England (e-mail: mwb5@le.ac.uk).

not only fulfilled the first three goals, but also could be easily simulated by the Beauvois and Meddis (1996) model of auditory stream segregation. The final goal was to use the operational characteristics of the model to account in neurological terms for the empirical data obtained from listeners.

## METHOD

In the experiment described below, the subjects listened to 15-sec alternating pure-tone sequences (ABAB . . .) that were presented at a fixed TRT of 130 msec, where  $\Delta f$ ,  $TD_A$ , and  $TD_B$  were varied independently. The subjects were asked to orient their attention over the duration of the sequence toward hearing either a temporally coherent percept (coherence) or a segregated percept (segregation). At stimulus offset, the subjects indicated whether their percept at the end of the stimulus had been that of a temporally coherent trill or of segregated A and B streams. From van Noorden's (1975) findings, we would expect an attentional set of coherence to result in a greater mean number of coherence responses ( $M$ ) than would an attentional set of segregation,  $M$  to decrease with increasing  $\Delta f$ , and  $M$  to decrease with increasing global-TD. From Hartmann and Johnson's (1991) findings, we would also expect  $M$  to decrease as  $\Delta TD$  increased.

### Subjects

There were 17 subjects, between 20 and 32 years of age, all of whom were already familiar with the concepts of temporal coherence and stream segregation, who had participated in previous auditory stream segregation experiments. All the subjects reported having normal hearing.

### Stimuli

The stimulus format was the same in every trial and consisted of a 5-sec silence, followed by a 15-sec ABAB . . . pure-tone sequence. In each trial,  $f_A = 1000$  Hz, TRT = 130 msec,  $\Delta f = 100, 280, \text{ or } 460$  Hz (i.e.,  $f_B = 1100, 1280, \text{ or } 1460$  Hz),  $TD_A = 40, 80, \text{ or } 120$  msec, and  $TD_B = 40, 80, \text{ or } 120$  msec. Included in the values of  $TD_A$  and  $TD_B$  were 5-msec rise and fall times for each tone, implemented with raised cosine ramps, in order to prevent onset and offset clicks. All stimuli were presented diotically at 65 dB SPL.

### Apparatus

The stimuli were generated on a NeXT computer at a sampling rate of 44.1 kHz, using the IRCAM Musical Workstation (Lindemann et al., 1991). The NeXT's digital output was fed into a Digidesign PRO10 D/A converter and then into a Canfort audio amplifier. The audio signal was then relayed to the subjects, via Beyer DT-49 headphones, while they were seated in a Soluna SN1 double-walled sound-attenuation booth.

### Procedure

The subjects were told that they would hear tone sequences for which there would be two possible percepts. One percept was that of a *warble* or *trill*, in which both tones contributed to the percept (temporal coherence). The other percept was that of *two streams*, in which the test sequence had split into separate A and B streams (stream segregation). The subjects were asked to use one particular attentional set (coherence or segregation) over the duration of the stimuli in each experimental block and to respond on the basis of the percept they heard as the stimulus ended. At stimulus offset, a computer terminal that was present in the booth prompted the subjects to press a button indicating which percept they had heard and warned of the next stimulus onset. Between the time at which the

subjects made their decision and the beginning of the next stimulus, there was a silent interval of 5 sec. Each subject was given two practice trials for each attentional set before the results were recorded, so that they could get used to the procedure and the stimuli.

All 17 subjects completed 10 blocks for each attentional-set condition, each block containing all possible combinations of the three  $\Delta f$  values, the three  $TD_A$  values, and the three  $TD_B$  values, in random order. The coherence and segregation blocks were presented to each subject in a different random order, and each possible combination of  $\Delta f$ ,  $TD_A$ , and  $TD_B$  was presented to the subjects once in each experimental block.

## RESULTS

An analysis of the data indicated that, as expected from van Noorden's (1975) findings, an attentional set of coherence produced significantly more coherence than did an attentional set of segregation and that perceived coherence decreased significantly as  $\Delta f$  increased. The analysis also indicated that perceived coherence decreased significantly as either  $TD_A$ ,  $TD_B$ , global-TD, or  $\Delta TD$  increased. There were also a number of significant interactions between factors. These findings are given in greater detail below.

A repeated-measures analysis of variance (ANOVA) was conducted on the number of coherence responses given by subjects in response to the stimuli, with attentional set (coherence, segregation),  $\Delta f$  (100, 280, 460 Hz),  $TD_A$  (40, 80, 120 msec), and  $TD_B$  (40, 80, 120 msec) as repeated factors. The mean number of coherence responses was found to be significantly greater for the attentional set of coherence than for that of segregation [ $F(1,16) = 32.7, p < .0005$ ], and mean coherence responses decreased significantly as  $\Delta f$  increased [ $F(2,32) = 83.7, p < .0005$ ]. Mean coherence responses also decreased significantly as  $TD_A$  increased [ $F(2,32) = 10.0, p < .0005$ ] and as  $TD_B$  increased [ $F(2,32) = 4.4, p < .05$ ].

The analysis also showed a number of significant interactions between factors. The first significant interaction was between attentional set and  $\Delta f$  [ $F(2,32) = 8.4, p < .005$ ] and is shown in Figure 1. Figure 1 shows  $M$  as a function of  $\Delta f$ , with attentional set as parameter. This interaction expresses itself as a sharp decrease in  $M$  between  $\Delta f = 100$  and 280 Hz for an attentional set of segregation and as a more gradual decline in  $M$  across  $\Delta f$  for an attentional set of coherence. These differences in attentional set across  $\Delta f$  could be due to the difference between the temporal coherence and the fission boundaries described by van Noorden (1975). Above the fission boundary, listeners are able to hear a segregated percept at will; above the temporal coherence boundary, listeners are unable to hear a coherent percept at will (van Noorden, 1975). Although these perceptual boundaries vary widely between listeners and with procedure, the fission boundary typically occurs at a far smaller  $\Delta f$  (75–155 Hz) than does the temporal coherence boundary (300–1000 Hz) when  $f_A = 1000$  Hz and TRT = 130 msec. In effect, this means that a listener's ability to hear a segregated percept

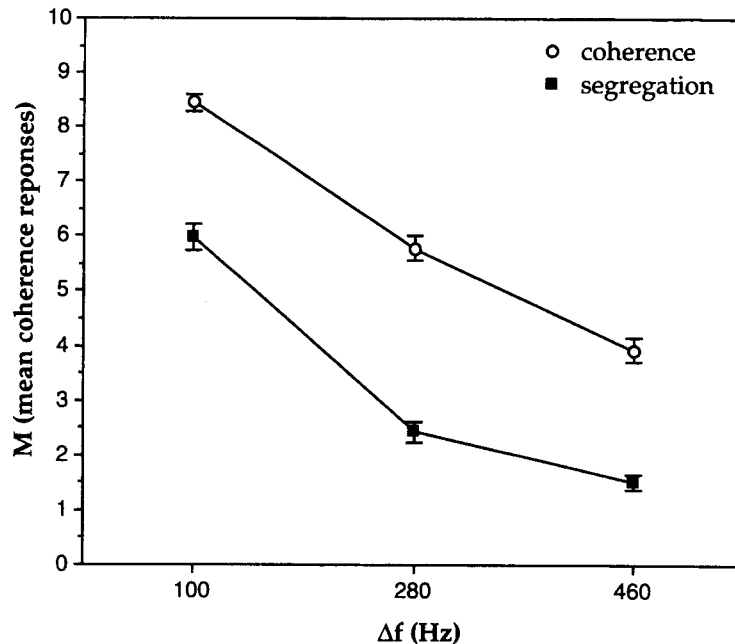


Figure 1. Mean number of coherence responses ( $M$ ) as a function of  $\Delta f$ , with attentional set as parameter. Standard-error bars are shown for each data point.

will become markedly easier at a small  $\Delta f$  once the fission boundary is crossed, leading to the steep decline in  $M$  values at small  $\Delta f$  values that is shown in Figure 1. On the other hand, a listener's ability to hear a coherent percept will gradually decrease with  $\Delta f$  until the temporal coherence boundary is crossed, leading to the gradual decline in  $M$  values across  $\Delta f$  that is shown in Figure 1. The pattern of results shown in Figure 1 therefore suggests that the interaction between attentional set and  $\Delta f$  is caused by the effect of the perceptual boundaries described by van Noorden.

The second significant interaction was between  $TD_A$  and  $TD_B$  [ $F(4,64) = 54.0, p < .0005$ ] and is shown in Figure 2. Figure 2 shows  $M$  as a function of  $TD_B$ , with  $TD_A$  as parameter, and indicates that mean coherence decreased significantly as both global-TD and  $\Delta TD$  increased. For global-TD, this can clearly be seen by considering the  $M$  values in Figure 2 for the  $TD_A/TD_B$  combinations 40/40, 80/80, and 120/120. For  $\Delta TD$ , this can clearly be seen by considering the  $M$  values for the  $TD_A/TD_B$  combinations 40/40, 40/80, and 40/120 (see note 1). Figure 2 also indicates that  $\Delta TD$  had a stronger effect on perceived coherence than did global-TD. If the 40/40 combination is taken as a reference point,  $\Delta TD$  values of 40 and 80 msec (e.g.,  $TD_A/TD_B = 40/80$  and 40/120) produced a greater decrease in mean coherence than did global-TD increases of 40 and 80 msec (e.g.,  $TD_A/TD_B = 80/80$  and 120/120). The stronger effect of  $\Delta TD$  is also shown by the lowest mean coherence for the stimuli being produced by the maximum  $\Delta TD$  value of 80 msec ( $TD_A/TD_B = 40/120$  and 120/40). Finally, Figure 2 indicates that mean coherence decreased linearly as  $\Delta TD$  increased, with the slope of the line being flatter for

a longer  $TD_A$  (120 msec) than a shorter  $TD_A$  (40 msec). In order to make the effects of global-TD and  $\Delta TD$  more apparent, Figure 2 is replotted in Figure 3 so as to show  $M$  as a function of  $\Delta TD$  ( $TD_B - TD_A$ ), with  $TD_A$  as parameter. When  $\Delta TD = 0$  msec, the decrease in mean coherence with increasing global-TD can clearly be seen as  $TD_A$  increases from 40 to 120 msec. Figure 3 also clearly shows the stronger segregation effects of  $\Delta TD$  as  $\Delta TD$  increases from 0 to 80 msec.

The third significant interaction was between  $\Delta f$ ,  $TD_A$ , and  $TD_B$  [ $F(8,128) = 2.5, p < .05$ ] and is shown in Figure 4. Figure 4 shows  $M$  as a function of  $TD_A$  and  $TD_B$  with  $\Delta f$  as parameter. A comparison of Figures 2 and 4 indicates that, even though there is a weak interaction with  $\Delta f$  that changes specific details of the  $TD_A/TD_B$  pattern, the general form of the interaction between  $TD_A$  and  $TD_B$  shown in Figure 2 is independent of  $\Delta f$ . These results therefore indicate that the effects of global-TD and  $\Delta TD$  are independent of  $\Delta f$ .

The last significant interaction (not shown) was between attentional set,  $\Delta f$ ,  $TD_A$ , and  $TD_B$  [ $F(8,128) = 3.5, p < .005$ ]. Here, the general form of the  $\Delta f/TD_A/TD_B$  interaction, shown in Figure 4, remained largely independent of attentional set, with specific details of the  $\Delta f/TD_A/TD_B$  pattern changing as a result of trying to hear coherence or segregation. The main effect of attentional set was to produce lower overall coherence scores for an attentional set of segregation. None of the other interactions was significant.

In conclusion, the results indicate a significant effect of global-TD and  $\Delta TD$  on auditory stream formation, with coherence decreasing as both global-TD and  $\Delta TD$  increase, and with  $\Delta TD$  producing stronger segregation ef-

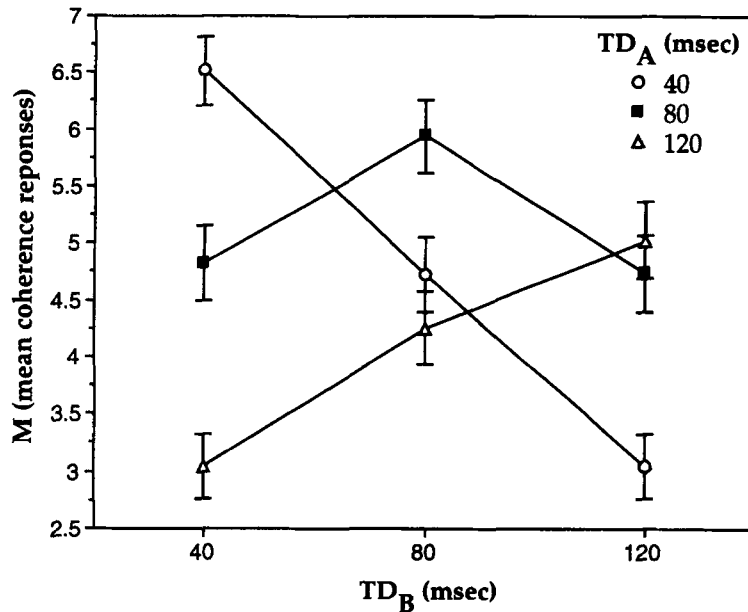


Figure 2. Mean number of coherence responses ( $M$ ) as a function of  $TD_B$ , with  $TD_A$  as parameter. Standard-error bars are shown for each data point.

facts than does global-TD. These findings are in general agreement with those of previous TD studies (Hartmann & Johnson, 1991; van Noorden, 1975).

### MODEL SIMULATION

The Beauvois and Meddis (1996) computer model of auditory stream segregation is designed to work specif-

ically on ABAB pure-tone sequences and uses low-level auditory analysis to account for some stream segregation phenomena associated with these stimuli. The model itself is based on simple circuits that are analogous to physiological systems present in the lower auditory system and the auditory periphery and exhibits behavior that is similar to that of human listeners for certain simple stimuli. Using the same set of parameter values, the

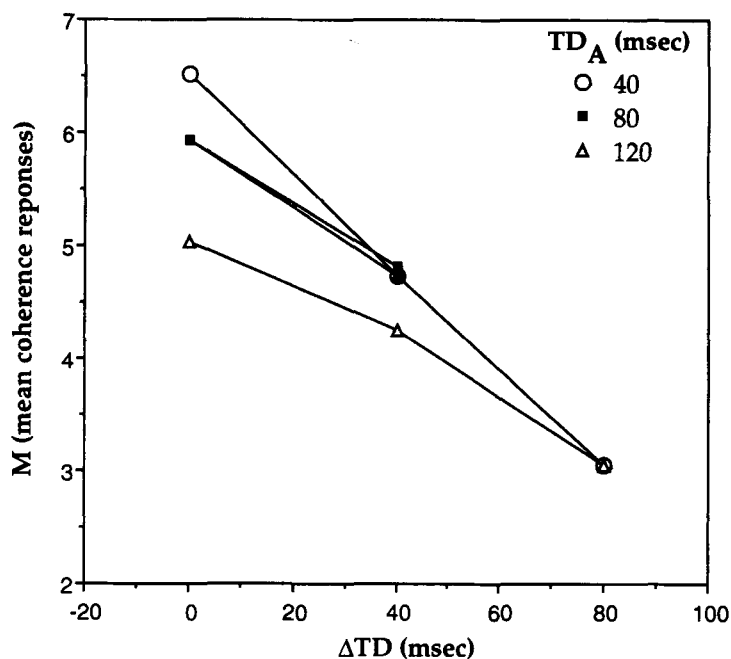


Figure 3. Mean number of coherence responses ( $M$ ) as a function of  $\Delta TD$  ( $TD_B - TD_A$ ), with  $TD_A$  as parameter.

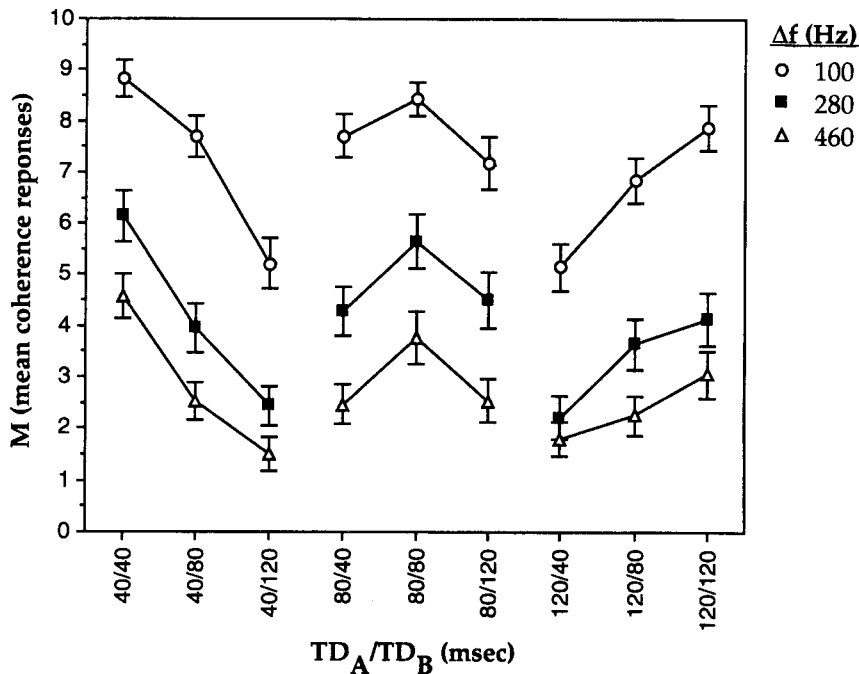


Figure 4. Mean number of coherence responses ( $M$ ) as a function of  $TD_A/TD_B$  combination, with  $\Delta f$  as parameter. Standard-error bars are shown for each data point.

model is able to reproduce a number of phenomena associated with auditory stream segregation, including the buildup of stream segregation over time, the temporal coherence and fission boundaries obtained from human listeners, and the trill threshold (Beauvois & Meddis, 1996). In addition, the model is able to reproduce Gestalt auditory grouping on the basis of frequency *proximity* (Beauvois & Meddis, 1995). A summary of the model features follows, and a diagram showing the construction of one model channel is given in Figure 5.

1. The acoustic signal (represented by a number sequence) is first subjected to a peripheral frequency analysis, which establishes "channels" characterized by a bandpass frequency response to stimuli.

2. The output of each bandpass filter (the amplitude of the input signal attenuated by the filter function) is fed into a simulation of a group of inner hair-cell/AN (auditory nerve) synapse units.

3. Each model channel then subdivides into two pathways.

3.1. A temporal fine-structure path preserves all aspects of the AN output, in order to enable the signal to be processed by higher levels and to preserve all AN information for pitch extraction purposes.

3.2. A temporal-integration path temporally integrates the AN output ( $\tau = 3$  msec).

4. The temporally integrated AN output is then sent through two further pathways.

4.1. An excitation-level path adds a cumulative random element to the temporally integrated AN output and

then subjects it to a slower temporal-integration process ( $\tau = 70$  msec).

4.2. An amplitude-information path preserves the temporally integrated AN output for determining the "loudness" of the signal.

5. The outputs of the excitation-level paths in the different model channels are compared, in order to see which one has the highest excitation level. The channel with the highest excitation level is then defined as the "dominant channel."

6. The activity in all of the amplitude-information paths, except the one belonging to the dominant channel, is then attenuated by a factor of 0.5.

7. The final stage of the model sums the outputs of the attenuated and nonattenuated amplitude-information paths across channels in order to give the "loudness" output of the system.

When an ABAB sequence segregates into two streams, a figure/ground percept in terms of loudness is produced, with the attended stream being louder than the unattended stream (van Noorden, 1975). Similarly, the model assumes that segregation occurs when one tone dominates the model output in terms of "loudness"—that is, when a critical output difference between tones A and B is exceeded. To assess this, the average "loudness" output while tone A is on is compared with the average "loudness" output while tone B is on over each second, from the beginning to the end of the stimulus. If the output difference between tones A and B exceeds a critical value, they are considered to have segregated into separate

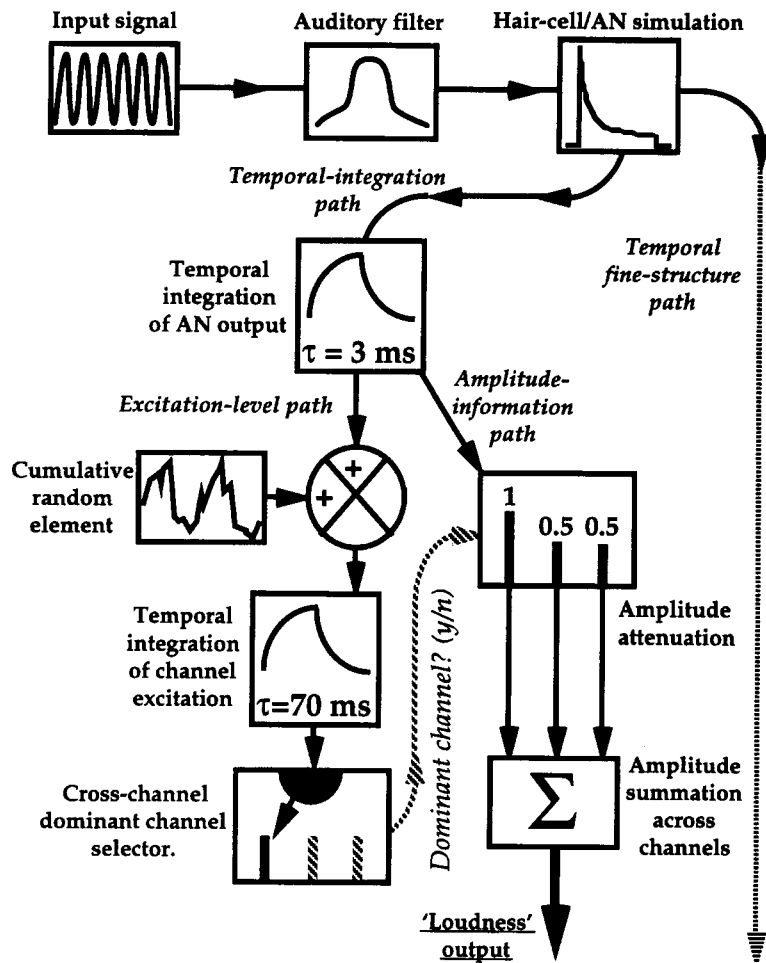


Figure 5. Diagram showing the construction of one model channel.

streams over the previous second. That is, the model “hears” one tone as being consistently “louder” than the other. If not, a coherent percept is reported.

When the tones of an ABAB stimulus are presented to the model, a gradual buildup of excitation occurs in the 70-msec temporal integrators of the tone-A and tone-B channels. Normally, the dominant channel (in terms of excitation level) oscillates between the tone-A channel (when A is active) and the tone-B channel (when B is active), and each tone is relayed faithfully through its own channel without attenuation. This regular oscillation between the tone-A and tone-B channels creates roughly equal levels in the system output for both tones A and B, resulting in a model output of temporal coherence. However, the oscillation between the tone-A and tone-B channels will be destabilized by the action of the cumulative random element that is (1) unique to each channel and (2) proportional to the channel activity, which is at its highest when a tone is being presented. This destabilization of channel oscillation occurs because the random element, when passed through a slow temporal-integration process, creates separate random walks (in terms of exci-

tation level), which differ from one channel to another. If the action of the random element should cause the accumulation of unusually large amounts of excitation in one channel, this channel will remain the dominant channel, even when it is not being actively stimulated. In this situation, the action of the channel-attenuation mechanism will produce unequal levels of A and B in the system output, resulting in a model output of segregation.

The responses of the Beauvois and Meddis (1996) model to the experimental stimuli can be predicted by considering the destabilizing effect of the random element on channel oscillation. With respect to global-TD, the overall channel activity, and consequently the random element, will increase as global-TD increases. Inasmuch as an increase in the random element increases the destabilization of channel oscillation, we would therefore expect the model to give more segregation reports as global-TD increases. With respect to  $\Delta$ TD, the difference between the activity levels in the tone-A and tone-B channels will increase as  $\Delta$ TD increases, with the random element being strongest in the channel corresponding to the tone with the longest TD. This situation will again

increase the destabilization of channel oscillation and, consequently, increase the number of segregation reports occurring in the model output. However, this increase in segregation will be further accentuated by the different lengths of the A and B tones, which will create unequal levels in the system output for A and B. These level differences will increase as  $\Delta TD$  increases, and the system output will be biased in favor of whichever tone had the longest TD. Consequently, more segregation reports will occur in this situation, as compared with a situation in which both tones had the same TD. We would therefore expect the model to give more segregation reports as  $\Delta TD$  increases, and we would expect increases in  $\Delta TD$  to produce greater segregation effects than would increases in global-TD. Note that these predictions of the model output in response to the experimental stimuli agree with the subject responses found in the experiment: coherence decreased as both global-TD and  $\Delta TD$  increased, and  $\Delta TD$  produced stronger segregation effects than did global-TD.

Beauvois and Meddis (1996) optimized the model parameter values so that the model gave the best fit to a range of experimental data. However, for this particular model simulation, the model parameter ( $M_r$ ) controlling the cumulative random element was increased from 0.006 to 0.009, in order to heighten any effect that TD might have on the random element and, consequently, the model output. When  $M_r = 0.009$ , the model is still able to simulate the results of other experimental studies, albeit with lower overall coherence responses, as com-

pared with cases in which  $M_r = 0.006$ . This decrease in overall coherence occurs because of the increased destabilization of channel oscillation created by the increase in the random element controlled by  $M_r$ .

The stimuli presented to the model were exactly the same as those presented to the 17 subjects in the experiment described above. Each  $\Delta f/TD_A/TD_B$  combination was presented to the model 250 times, and the number of coherence responses for the last second of each sequence was totaled for each combination and converted to a percentage. Figure 6 shows the percentage of coherence responses as a function of  $TD_B$ , with  $TD_A$  as parameter, and is the model equivalent to the subject responses shown in Figure 2. Figure 6 indicates that, although the overall coherence levels are somewhat reduced, as compared with those of listeners, the model gives essentially the same pattern of responses, with both the main effects of global-TD and  $\Delta TD$  being present. That is, coherence decreases as both global-TD and  $\Delta TD$  increase, with  $\Delta TD$  producing stronger segregation effects than does global-TD. To verify these trends and compare them with those shown by listeners, Figure 6 is replotted in Figure 7, in order to show the percentage of coherence responses as a function of  $\Delta TD$  ( $TD_B - TD_A$ ), with  $TD_A$  as parameter. A comparison of Figures 3 and 7 indicates that, although the model shows a similar decline in coherence as  $\Delta TD$  increases, it shows a much smaller effect of global-TD (when  $\Delta TD = 0$  msec) than listeners.

One possible explanation for the model's inability to fully reproduce the effect of global-TD may lie in spec-

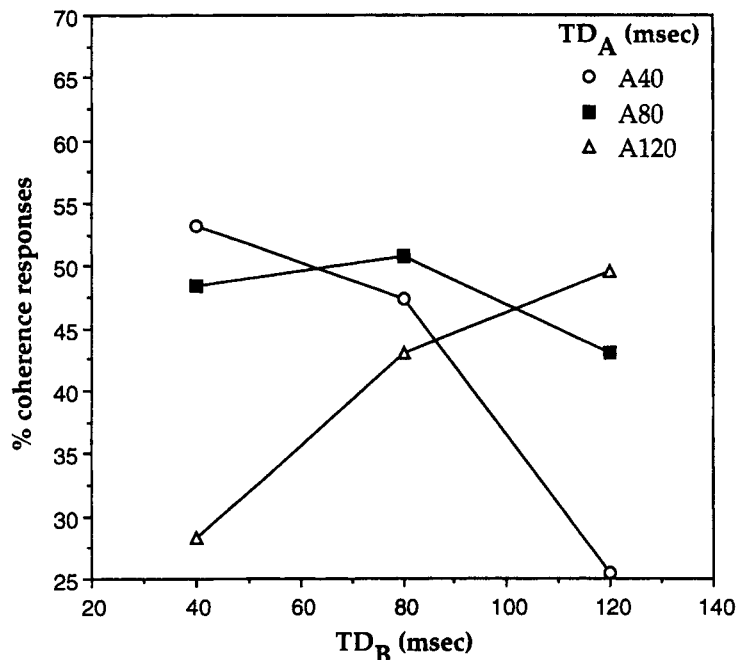


Figure 6. Percentage of coherence responses given by the model for the last second of the experimental stimuli as a function of  $TD_B$ , with  $TD_A$  as parameter.

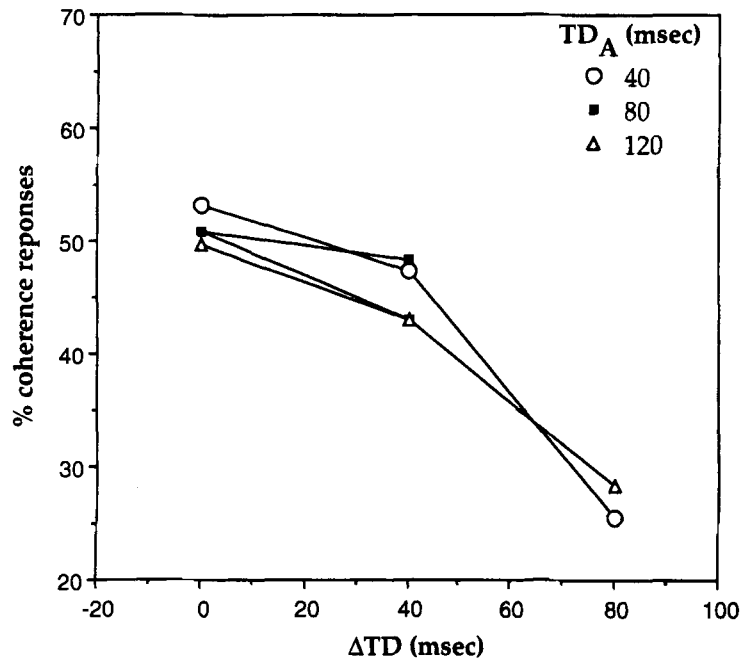


Figure 7. Percentage of coherence responses given by the model for the last second of the experimental stimuli as a function of  $\Delta TD$  ( $TD_B - TD_A$ ), with  $TD_A$  as parameter.

tral (i.e., peripheral-channeling) factors. Beauvois and McAdams (1998) found that the formation of auditory streams is strongly affected by the spread of stimulus excitation on the basilar membrane, with coherence increasing as the overlap in the excitation created by the A and B tones increases. These findings suggest that, because the A and B tones in this experiment create broader spreads of spectral energy on the basilar membrane (spectral splatter) as TD decreases, greater excitation overlap and, consequently, more coherence will occur when TD decreases. Spectral splatter, therefore, may be responsible for the greater increase in coherence scores shown by listeners, as compared with the model, when global-TD decreased. Beauvois and McAdams' findings further suggest that the secondary effect of TD that was found by Hartmann and Johnson (1991) was caused by the envelope shapes they used for their TD stimuli, which were designed to minimize spectral splatter, and that a greater TD effect would have been observed if this measure had not been carried out.

If spectral splatter *does* contribute to the global-TD effect found in this study, then, because the model, at present, cannot reproduce splatter effects, it will be unable to reproduce the full effect of global-TD shown by listeners. Alternatively, higher level processes not simulated by the model may be involved in the global-TD effect. However, considering that the model is simple and low-level in nature, it is not surprising that some limitations are present in its ability to simulate complex auditory-streaming phenomena. In conclusion, the results of the model simulation suggest that spectral (peripheral-channeling) fac-

tors may be involved in the global-TD effect found in the experiment. Furthermore, the inability of the Beauvois and Meddis (1996) model to fully simulate these effects offers a challenge to other auditory-streaming models to reproduce these findings.

## GENERAL DISCUSSION

Previous studies (e.g., van Noorden, 1975) have highlighted the strong effect of TRT on stream formation. However, this study has shown that when TRT remains constant, global-TD and  $\Delta TD$  have significant effects on stream formation. This suggests that stream segregation phenomena hitherto ascribed to the effect of TRT may be due to other temporal intervals associated with ABAB stimuli. Bregman (1990) describes four such intervals (see Figure 8): the interval between the onsets of consecutive tones (TRT), the silent interval between consecutive tones (SI), the interval between the onsets of tones in the same frequency range (On-On), and the interval between the offset and onset of tones in the same frequency range (Off-On).

Bregman (1990) asked the question, which intervals are important? Some evidence for answering this question is provided by van Noorden (1975), who investigated the effect of global-TD by keeping TRT and the On-On interval constant and varying SI and the Off-On interval. This procedure is analogous to presenting listeners the  $TD_A/TD_B$  combinations 40/40, 80/80, and 120/120, used in this study. Van Noorden found that segregation increased with global-TD, indicating that the SI



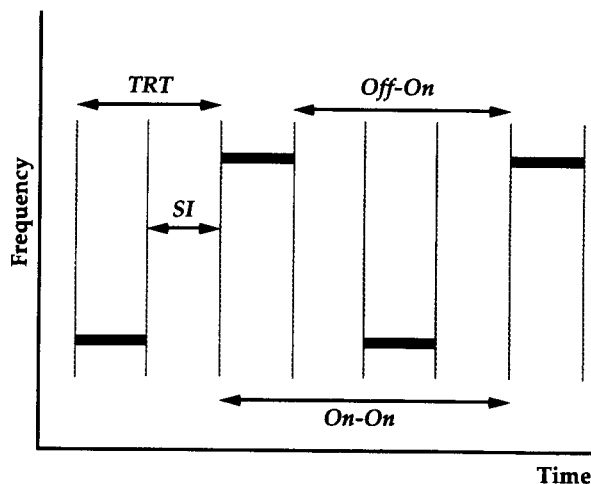


Figure 8. Temporal intervals associated with ABAB pure-tone stimuli (redrawn from Bregman, 1990; see text).

and the Off-On intervals determine the listener's percept. This finding is supported by the similar effect of global-TD reported in this study. However, in an experiment investigating the fission boundary—the  $\Delta f$  value below which a temporally coherent percept is always heard—van Noorden found that when TRT and the On-On interval were fixed, the fission boundary remained at a constant  $\Delta f$ , even when tones A and B overlapped in time. That is, the listener's percept was independent of SI. Similar results were found by Dannenbring and Bregman (1976) for nonoverlapping tones.

Van Noorden (1975) concluded from the results of his global-TD and fission boundary experiments that the Off-On interval was the most important temporal interval for determining the listener's percept. This conclusion is further supported by a study by Bregman, Ahad, & O'Reilly (1998), which presented listeners with ABA-ABA stimuli in which TRT and the On-On interval remained constant but the Off-On interval and SI were varied. Their results indicated that segregation increased as global-TD increased and that this trend, which remained independent of  $\Delta f$ , held even when the stimulus tones overlapped in time (i.e., when SI was negative).

The evidence given above not only implies that the effect of global-TD reported in this study is due to the length of the Off-On interval, but also suggests that the auditory system decides which percept is heard on the basis of within-stream temporal intervals. If this is the case, we would expect strong segregation effects to occur when listeners are presented with stimuli that create different within-stream temporal intervals. This situation, in fact, occurred when listeners were presented with the  $\Delta TD$  stimuli used in this study, as the effect of  $\Delta TD$  is to create two different Off-On (within-stream) intervals, one for the A tones and one for the B tones. Inasmuch as the experimental results reported here indicate that  $\Delta TD$  produces far stronger segregation effects than does global-TD, the weight of evidence presented here strongly sug-

gests that the lengths of the Off-On (within-stream) temporal intervals are what determine the listener's percept.

The model simulation, although limited in its ability to reproduce the full effect of global-TD, was able to successfully reproduce both the main and the relative segregation effects of global-TD and  $\Delta TD$ , thereby supporting the validity of the low-level Beauvois and Meddis (1996) model. The model itself uses the two principles of stochastic operation (the cumulative random element) and temporal integration (low-pass filtering) to account for stream segregation phenomena. The model's random element reflects the probabilistic nature of spike generation in the auditory nerve and elsewhere in the brainstem. When this random element is passed through neural systems characterized by low-pass filtering, random-walk phenomena are produced with properties that simulate stream segregation phenomena. Therefore, the results of the model simulation suggest that a combination of low-level neurophysiological processes and spectral (peripheral-channeling) factors are responsible for the effects of global-TD and  $\Delta TD$  on stream formation. This idea has ramifications for auditory scene analysis theory.

Bregman (1990) views stream formation as the result of the grouping performed by a pre-attentive primitive segregation mechanism that operates on the basis of Gestalt principles of grouping. For example, the Gestalt principle of *proximity* states that nearer elements are grouped together in preference to those that are spaced further apart. According to Bregman, Gestalt theory would suggest that it is the *proximity* of the tones in a time-by-frequency space that determines stream formation. That is, the silences that separate the tones (SI and the Off-On interval) are the important factor in determining the listener's percept. The experimental findings discussed here support the Gestalt approach, by indicating that the Off-On interval is the crucial temporal factor in stream formation. However, the results of the model simulation would suggest that this Gestalt auditory grouping by temporal *proximity* could equally be understood in terms of low-level neurophysiological processes and peripheral-channeling factors. Furthermore, the model's ability to simulate various stream segregation phenomena and Gestalt auditory grouping by both temporal *and* frequency *proximity* would suggest that Bregman's primitive segregation mechanism could be considered to be an umbrella term for describing the massed action of low-level neurophysiological processes and peripheral channeling, the effect of which is to partition the incoming sensory evidence into streams on a Gestalt basis. If this is the case, further research may well discover the peripheral/low-level processes responsible for other Gestalt auditory-grouping principles.

## REFERENCES

- BEAUVOIS, M. W., & MCADAMS, S. (1998). *An excitation-pattern model of auditory-stream formation*. Unpublished manuscript, IRCAM, Paris.  
 BEAUVOIS, M. W., & MEDDIS, R. (1995). Computer simulation of Gestalt auditory grouping by frequency proximity. In L. S. Smith &

- P. J. B. Hancock (Eds.), *Neural computation and psychology* (pp. 155-164). London: Springer-Verlag.
- BEAUVOIS, M. W., & MEDDIS, R. (1996). Computer simulation of auditory stream segregation in alternating-tone sequences. *Journal of the Acoustical Society of America*, **99**, 2270-2280.
- BREGMAN, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.
- BREGMAN, A. S., AHAD, P. A., & O'REILLY, J. (1998). *Effects of tone duration on auditory stream segregation*. Unpublished manuscript, McGill University.
- DANNENBRING, G. L., & BREGMAN, A. S. (1976). Stream segregation and the illusion of overlap. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 544-555.
- HARTMANN, W. M., & JOHNSON, D. (1991). Stream segregation and peripheral channeling. *Music Perception*, **9**, 155-184.
- LINDEMANN, E., PUCKETTE, M., VIARA, E., DE CECCO, M., DECHELLE, F., & SMITH, B. (1991). The architecture of the IRCAM Musical Workstation. *Computer Music Journal*, **15**(3), 41-49.
- VAN NOORDEN, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Institute for Perception Research, Eindhoven.

#### NOTE

1. It could be argued that although the  $TD_A/TD_B$  combinations 40/40, 40/80, and 40/120 differ in  $\Delta TD$ , they also differ in overall duration ( $TD_A + TD_B = 80, 120, \text{ and } 160$  msec, respectively) and therefore cannot be used to examine the effect of  $\Delta TD$ , as some effect of overall duration may have been present. The argument developed in the main text is that  $\Delta TD$  produces stronger segregation effects than does global-TD, the overall duration having no effect on listeners' responses. If this ar-

gument is correct, then, if the overall duration stayed constant, we would expect more segregation to occur when  $TD_A \neq TD_B$  ( $\Delta TD$ ) than when  $TD_A = TD_B$  (global-TD). Similarly, if  $\Delta TD$  stayed constant, we would expect no differences in mean coherence scores to occur as the overall duration changed.

To see whether these assumptions were correct, the differences between mean coherence responses were examined for specific sets of  $TD_A/TD_B$  combinations shown in Figure 2, in order to see whether they were significant by single-degree-of-freedom planned contrasts. In the first set, the overall duration stayed constant at 120 msec (40/80 and 80/40), 160 msec (40/120, 80/80, and 120/40), and 200 msec (80/120 and 120/80). In the second set,  $\Delta TD$  stayed constant at 40 msec (40/80, 80/40, 80/120, and 120/80). In the first set, when overall duration stayed constant at 160 msec, the 40/120 and 120/40 combinations produced significantly more segregation than did the 80/80 combination [ $F(1,64) = 111.3, p < .0005$ ]. Inasmuch as none of the other contrasts for this set proved significant and the 40/120 and 120/40 combinations have a greater  $\Delta TD$  (80 msec) than does the 80/80 combination (0 msec), the results indicate that  $\Delta TD$  produces stronger segregation effects than does global-TD, even when the overall duration remains constant. In the second set, in which  $\Delta TD$  stayed constant at 40 msec, the 80/40 combination only just produced significantly more segregation than did the 120/80 combination [ $F(1,64) = 4.3, p = .0421$ ]. Inasmuch as none of the other contrasts for this set proved significant and the effect found here is weak, the results indicate that the effect of  $\Delta TD$  remains independent of the overall duration. For the purposes of the argument in the main text, the overall duration of the  $TD_A/TD_B$  combinations can therefore be disregarded.

(Manuscript received July 29, 1996;  
revision accepted for publication July 14, 1997.)