Role of predictability of sequence in auditory stream segregation

MARILYN FRENCH-ST. GEORGE and ALBERT S. BREGMAN
McGill University, Montreal, Quebec, Canada

This study evaluates the role of item predictability in auditory temporal coherence. Thirteen normal-hearing subjects were required to hold together long tonal sequences as single strings of notes. Temporal and spectral predictability of successive notes in a sequence varied as a function of experimental condition. As the frequency separation of the notes in the sequence increased, the subjects found it more difficult to hold the sequence together as a single stream. There was no significant difference in subjects' abilities in performing this task as a function of experimental condition. That is, the predictability of successive notes appeared not to have a role in temporal coherence.

This study was designed to evaluate the effect of serial predictability on auditory streaming—specifically, its effect on the temporal coherence boundary. Van Noorden (1975) defines this boundary as the largest frequency interval, between two alternating tones, at which an observer can still hear one alternating string of notes. When the frequency interval between alternating notes exceeds a critical value (3-15 semitones, depending on the stimulus rate), the listener hears two streams of notes: one stream containing the higher pitched notes and another stream containing the lower pitched notes.

The probability that temporal coherence will be heard is also inversely related to the duration of the tonal sequence (van Noorden, 1975). Once again, sequences tend to segregate on the basis of pitch proximity. In van Noorden's terminology, there are two perceptual boundaries between segregated and integrated percepts. The first, the temporal coherence boundary, occurs when the listener is trying to hear a single coherent stream. The second, the fusion boundary, occurs when the listener is trying to hear the streams as segregated. Only the temporal coherence boundary, in van Noorden's data, shows clear effects of frequency separation of subsets of tones and of the rate of the sequence.

It has been well documented that listeners' accuracies at "hearing out" components of a tone sequence are significantly influenced by the temporal and spectral structure of the sequence (Dowling, 1973; Dowling, Lung, & Herrbold, 1987; Handel, Weaver, & Lawson, 1983; Jones, Boltz & Kidd, 1982). However, the extent to which the same factors influence listeners' abilities at holding a string of notes together as a single stream has not been the topic of vigorous investigation.

Van Noorden (1975) has investigated the perceptual segregation of an isochronous sequence of tones in which

the frequency jump from one tone to the next was random. He found that the coherence of the sequence declined in proportion to the average size of the frequency jump, just as it did in predictable sequences when there was strict alteration of high and low notes. That is, there was no effect of spectral predictability on temporal coherence. Simplicity has dictated the use of isochronous predictable sequences in most investigations of the temporal coherence boundary. Consequently, the effect of onset-to-onset temporal irregularities on the coherence phenomenon is not known. The experiment described here was designed to serve in the further investigation of the extent to which temporal and spectral predictability influence the temporal coherence boundary.

METHOD

Stimuli

Pure tone stimuli were digitally synthesized on a DEC PDP-11/34 computer running the MITSYN software package (Henke, 1980) with a sampling rate of 10 kHz.[1] Signal outputs were amplified, low-pass filtered at 5 kHz by a Rockland 851 filter, and recorded onto reel-to-reel tape. Equality of signal levels was verified with the tape recorder's VU meter.

Table 1 shows the two ensembles of high- and low-frequency tones (A and B). The nearest tones within an ensemble are separated by one semitone. The semitone difference between the (semitone) center of the tones on the A ensemble and that of the corresponding B ensemble is given in the semitone separation column.

Four tonal sequences of 30-sec duration were created from each of the five groups of tones. Each 30-sec sequence was preceded by a gliding warning signal of 770-msec duration. For each semitone separation, adjacent tones in a sequence were drawn alternately from the A and B ensembles without replacement (e.g., A1 B4 A3 B2 A2 B3 A4 B1). Each tone was 40-msec in duration, which included a 10-msec rise/fall time.

The sequences were as follows:
1. Isochronous with repeating frequency pattern. Successive eight-note sequences were identical (e.g., A1 B4 A3 B2 A2 B3 A4 B1), and the sequence was continuously repeated for the full 30 sec. The stimulus onset asynchrony (SOA) was 91.2 msec.

384

Table 1
High- and Low-Frequency Four-Tone Ensembles by
Semitone Separation of Ensemble Centers

| | | Ensemble | | | | | | | |
| | | A | | | | B | | | |
Group	Separation	A1	A2	A3	A4	B1	B2	B3	B4
1	1	880	932	988	1048	932	988	1048	1108
2	2	880	932	988	1048	988	1048	1108	1176
3	3	830	880	932	988	988	1048	1108	1176
4	4	830	880	932	988	1048	1108	1176	1244
5	5	784	830	880	932	1048	1108	1176	1244

Note—All frequencies are in hertz. Separation is given in semitones.

2. Isochronous with irregular frequency pattern. Successive tones were drawn alternately from Ensembles A and B as in Sequence 1 above. However, choice of a particular tone within the A and B ensembles was random without replacement within a given frequency separation. Successive eight-tone sequences were concatenated until a 30-sec sequence was achieved.

3. Nonisochronous with repeating frequency pattern. Choice of successive tones was the same as in Sequence 1 above. However, onset-to-onset time for successive tones varied while retaining a mean of 91.2 msec across the eight successive tones. Successive SOAs (within an eight-tone sequence) were randomly chosen (without replacement) from a pool consisting of 65.6, 65.6, 91.2, 91.2, 116.8, 116.8, and 116.8 msec. Successive eight-tone sequences were concatenated until a sequence of 30-sec duration was achieved.

4. Nonisochronous with irregular frequency patterns. Within an eight-tone sequence, the mean SOA was 91.2 msec. Choice of onset-to-onset interval was identical to that in Sequence 3 above, whereas the choice of frequencies was governed by the rules for Sequence 2 above.

Experimental Design

The experiment may be described as consisting of a two-factor repeated measures design. The first factor, frequency separation of the high- and low-frequency ensembles, had five levels: 1, 2, 3, 4, and 5 semitones. The second factor, experimental condition, had four levels: regular, unpredictable pitch sequence, unpredictable rhythm, and unpredictable pitch sequence and rhythm.

The subjects received each experimental condition on a separate day. The order of conditions was randomized across subjects. For a given sequence type, six blocks of trials were presented. Each block contained one sequence from each of the frequency separations. Order of presentation was randomized within each block.

Subjects

Thirteen people with hearing levels better than or equal to 20 dB HL at the octave frequencies 250–4000 Hz participated. All subjects received a modest honorarium for their participation.

Task

Tone sequences were presented binaurally via TDH39 headphones at 65 dB SPL. The subjects' responses were recorded on an IBM XT computer. The subjects were told that each trial consisted of a warning signal immediately followed by a 30-sec sequence of tones. The subjects were instructed to press "G" on the computer keyboard as soon as they heard the sequence start. This started an internal timer in the IBM computer. The subjects were then told they should try to hold the sequence together as a string of tones and to indicate when this was achieved by pushing the "SHIFT" key. If the sequence split into two or more co-occurring streams, if they had a foreground/background percept, or if tones "popped" out of the sequence, they were to release the "SHIFT" key and

try to pull the sequence together again. As soon as this was achieved, they were to press "SHIFT" once again. This process continued until the 30-sec sequence was completed. Pressing and releasing the "SHIFT" key activated time markers. Thus the total time an individual was able to hold onto a single-string percept over the 30-sec sequence could be calculated. Prior to each condition, each subject was presented with nine training trials. The stimuli presented during these trials represented the extreme percepts—that is, strings that could easily be held together (1-semitone separation) and streams that could not be held together (5-semitone separation). The subjects were required to meet a pretest criterion of over 20 sec of holding a single-stream percept when the frequency separation was one semitone, and less than 10 sec for the 5-semitone separation.

RESULTS

Figure 1 illustrates the time (averaged over six trials per condition per frequency separation) during which participants heard a single-string percept as a function of frequency separation between the A and B tone ensembles for each condition.

The time for which subjects could maintain a string percept diminished as the frequency separation increased
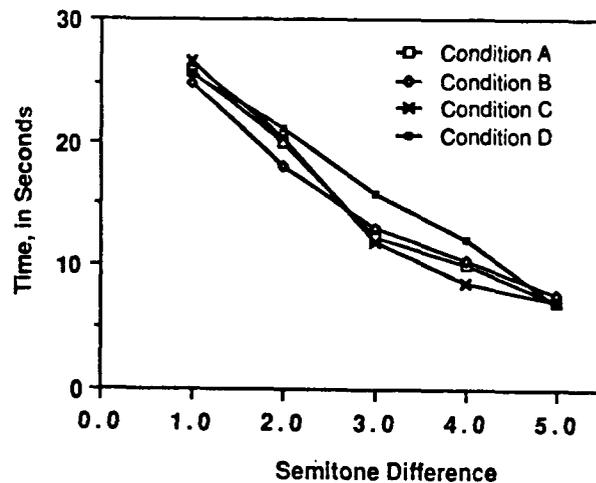
Figure 1. Mean duration of single-stream percept as a function of predictability of sequential tones. Condition A: Isochronous with repeating frequency pattern. Condition B: Isochronous with irregular frequency pattern. Condition C: Nonisochronous with repeating frequency pattern. Condition D: Nonisochronous with irregular frequency patterns.

Table 2
Mean Durations (in Seconds) of Single-String Percept
as a Function of Frequency Separation (in Semitones)

		Frequency Separation		
	2	3	4	5
Mean Duration 25.7	19.9	13.2	10.1	7.0

Note—Means connected by underline are not significantly different from one another.

from 1 to 5 semitones under all conditions. When the frequency separation was 1 semitone, the average duration of the single-string percept was at least 25 sec. When the frequency separation increased to 5 semitones, the average duration of the single-tone percept was less than 7.5 sec.

Analysis of variance revealed a single main effect for semitone difference [$F(4,240) = 82.96, p < .0001$]. No significant effect was found for condition, and there was no interaction between frequency separation and condition. Follow-up Tukey (WSD) evaluation of differences in mean durations of the single-string percept as a function of frequency separation is illustrated in Table 2.

DISCUSSION

The results obtained in this experiment parallel those of previous research. In all four conditions, the mean number of seconds (out of 30) during which listeners could hear the sequence as a single stream becomes less as the frequency separation between the higher and lower tones became greater. Furthermore, predictability (temporal or spectral) did not help listeners hold the sequence together as a coherent one. That is, Condition 1 (predictable rhythm and predictable pitch pattern) was not held as a single stream for a longer duration than was any of the other conditions. However, it should be noted that when the total stream was predictable (Condition 1), the sub-streams (high and low) were also predictable. It is conceivable that in this condition, the predictability of the sub-streams helped them be heard as separate units (despite listeners' attentions being directed toward the global sequence), thereby counteracting the effects of the regularity of the global sequence. Experiments in which the regularity of the overall sequence and that of the within-substream sequence are varied independently are needed to clarify this issue.

It could be argued that because the frequency separation changed on each trial, listeners did not become suf-

ficiently familiar with the regular sequence to take advantage of its regularity. In conditions with large frequency separations, the sequence might have split before the listener had a chance to learn it. This argument, however, is not altogether convincing, since, for the intermediate separations, listeners were still unable to hold the predictable sequence together longer than in the other conditions, despite their having sufficient time to learn the pattern before segregation took place.

A final criticism that might be made of this experiment is that the range in ambiguity of frequency and timing was too small to demonstrate the stronger effects of isochronous, predictable patterns. Further experiments are required to address this issue.

In conclusion, these results suggest that when listeners' attentions are directed away from "hearing out" sequence elements, item-to-item predictability in both the spectral and the temporal domains play no role in the inevitable segregation that occurs with long-duration alternating tonal sequences. It appears that grouping of acoustic material is affected directly by Gestalt-like factors of proximity in frequency and time.

REFERENCES

Dowling, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, **14**, 37-40.

Dowling, W. J., Lung, K., M.-T., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception & Psychophysics*, **41**, 642-656.

Handel, S., Weaver, M. S., & Lawson, G. (1983). Effect of rhythmic grouping on stream segregation. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 637-651.

Henke, W. L. (1980). *MITSYN: A coherent family of command-level utilities for time signal processing* [computer program]. Available from W. L. Henke, 133 Bright Street, Belmont, MA 02178.

Jones, M. R., Boltz, M., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. *Perception & Psychophysics*, **32**, 211-218.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Technische Hogeschool Eindhoven, Eindhoven, The Netherlands.

NOTE

1. The MITSYN package was revised and extended by Daniel Solomon, Department of Psychology, McGill University.