The perceived tempi of coherent and streaming tone sequences: II

GERT TEN HOOPEN and SIMONE AKERBOOM

University of Leiden, Leiden, The Netherlands

Previous research demonstrated, by means of a reaction time technique, that there is no difference between the perceptual onset asynchronies (POAs) of nonalternating tone sequences and of frequency-alternating tone sequences, even if the latter are streaming. This finding implies that listeners are capable of extracting the tempo of the total alternating sequence, even though successive tones belong to different streams. The present study investigated whether subjects are also capable of establishing the tempo of one stream. To this end, listeners were required to match the rate of one stream by adjusting the rate of a nonalternating comparison stream (a pulse sequence). In addition, they were required to match the total rate of the streaming sequence. It turned out that the total tempo could be matched precisely and that the tempo of one stream could be approximated very closely.

It is a well-established fact that sequences whose tones alternate between ears (interaural sequences) have a slower apparent repetition rate than their nonalternating counterparts (monaural sequences) (Axelrod & Guzy, 1968; Axelrod, Guzy, & Diamond, 1968). Recently, this amount of subjective tempo difference between interaural and monaural sequences (having the same objective rate) has been established by several experimental techniques. Ten Hoopen, Vos, and Dispa (1982) demonstrated (1) that the difference in *perceived* onsets between interaural and monaural tones is 24 msec relative to the stimulus onset asynchrony (SOA) and (2) that this difference is invariant with the value of SOA. Ten Hoopen, Vos, and Dispa coined the concept of perceptual onset asynchrony (POA), which is the *perceived time* between successive tone percepts in auditory memory. Although the SOAs of interaural and monaural tone sequences are the same, the respective POAs differ by 24 msec.

The most straightforward technique to reveal this POA difference is the "stop reaction time" (stop-RT) paradigm. Listeners are required to respond to the offset of sequences comprising an unpredictable number of tones. The difference in stop-RT between two different presentation modes (e.g., interaural and monaural) reflects the difference in POA.

Ten Hoopen, van Meurs, and Akerboom (1982) applied this technique also to sequences of tones that alternated between frequencies in order to investigate

whether the POAs of these sequences differed from the POA of monotonous tone sequences. Such an investigation interested us because alternating tones between frequencies can give rise to different percepts, depending on the SOA and frequency interval between successive tones (see Bregman, 1978a, for an overview).

When tones alternate between frequencies not separated by an interval of more than about two semitones, the sequence is perceived as a coherent whole. When tones alternate relatively fast between much larger frequency intervals, two separate streams of tones can be heard (one high and the other low), of which only one can be "put in the foreground" or attended to. Van Noorden (1975) has meticulously investigated which percept (coherence or streaming) arose as a function of SOA and frequency interval combination.

Ten Hoopen, van Meurs, and Akerboom (1982) posed two questions. First, is the POA of frequency-alternating, but coherent, tone sequences the same as the POA of nonalternating tone sequences or not? In other words, is the subjective tempo of coherent frequency-alternating tone sequences the same as that of monotonous sequences or not? The second question was more complicated. When a frequency-alternating tone sequence is perceived as streaming, is the perceived tempo "governed" by the sequence as a whole or only by the stream to which attention is directed? If the latter is true, a POA that corresponds to twice the SOA should be found. If the former is true, a POA that corresponds to the SOA itself should be found.

Ten Hoopen, van Meurs, and Akerboom tackled these problems by means of their stop-RT paradigm and found that there was no difference between the POAs of nonalternating (monotonous), coherent, and

This research was supported by a grant from the Netherlands Organization for the Advancement of Pure Research (ZWO). We are most grateful to Edward C. Carterette for his help in improving the English of this report. Requests for reprints should be sent to G. ten Hoopen, Unit of Experimental Psychology, Hooigracht 15, 2312 KM Leiden, The Netherlands.

streaming tone sequences. Hence, the answer to the first question was evident: frequency-alternating, but coherent, tone sequences have the same subjective tempo as monotonous sequences.

The answer to the second problem was also clear. Apparently, the tempo of the streaming sequence is governed by the POA between adjacent tones, although these (physically) adjacent tones belong to different streams. Thus, one might get the impression from this study that listeners are not capable of perceiving and establishing the subjective tempo of only the attended stream (which should correspond to half of the objective tempo of the whole sequence). It is quite possible, however, that the stop-RT paradigm does not measure such a capability of the listener. Informal listening to demonstrations taped by McAdams and Bregman (1979) and van Noorden (1975) indeed suggests that one can perceive the tempo of a separate stream. A good example is the repeating tone pattern ABA. If the frequencies of A and B are separated far enough, streaming is heard and directing attention to stream A yields a much faster tempo than directing attention to stream B.

The present study was an attempt to test whether listeners are capable of establishing the tempo of one stream by means of procedures more convenient than the stop-RT paradigm.

EXPERIMENT 1

In this experiment, the subjects were required (1) to match the rate of a pulse sequence with the rate of coherent tone sequences and (2) to match the rate of a pulse sequence with the rate of one streaming tone sequence.

Method

Subjects. Eighteen male and female undergraduate students served as subjects. They were paid for their services. None reported having any hearing deficit.

Stimuli and Apparatus. Coherent tone sequences comprising enveloped (2-msec rise and fall time) sine tones of 45-msec duration, alternating over one semitone (972-1,029 Hz), were generated by means of Coubourn Instruments modules under control of a microcomputer. Three sequences, with SOAs between tones of 150, 185, and 220 msec, respectively, and each sequence lasting 4 min were taped with an Uher magnetic tape recorder (Royal 561). The sound pressure level was measured by attaching the shell of the headphones (AKG Cardan K 140 S) to an artificial ear (Brüel & Kjaer, type 4152) mounted with a condenser microphone (Brüel & Kjaer, type 4144). The precision sound-level meter (Brüel & Kjaer, type 2203) registered 65.5 ($\pm .25$) dBA, 64.5 ($\pm .50$) dBA, and 63.5 ($\pm .50$) dBA (fast mode) for the sequences with SOAs of 150, 185, and 220 msec, respectively.

Streaming tone sequences comprised sine tones alternating over 20 semitones (561-1,782 Hz), but in other characteristics were identical to the coherent sequences. Pulse sequences comprised positive square waves lasting 40 msec. The sound pressure level of the pulse sequences was the same as with the tone sequences. The duration of the response of the headphone to a pulse was measured by attaching the phones to the artificial ear and dis-

playing the response of the condenser microphone on a digital storage oscilloscope (Gould Advance, type OS 4000). The activity started immediately after the onset and ended within 10 msec after the offset of the pulse.

Design and Procedure. Each subject was required to make two matches of the repetition rate of the pulse sequence with the rate of each of the three coherent sequences. Before the subject started, the experimenter set the SOA of the pulse sequence either at 100 msec or at 600 msec.

In the case of the streaming sequences, the streaming phenomenon was first made explicit to the subject, who was asked to attend to one of the streams (either high or low), according to preference. The subject was instructed to match the pulse rate exclusively with the tempo of this attended stream. This was repeated twice for each of the three sequences. The starting SOAs of the pulse sequences were either 100 msec or 600 msec.

The experimental sequences were presented diotically by means of the Uher recorder and AKG headphones. Using a switch, the subject could listen to either the experimental (coherent and streaming) sequences or to the pulse sequence over the same headphones. The subject could vary the rate of the pulse sequences by means of a rotating SOA knob, whose scale could not be seen. By switching back and forth between the experimental sequence and the pulse sequence, the subject could adjust the pulse rate until satisfied with the match. On average, a subject switched back and forth about six to eight times.

Results and Discussion

The adjusted SOAs of the pulse sequence, averaged over replications and subjects, are displayed in Figure 1 and Figure 2 according to type of experimental sequence (coherent vs. streaming) and SOA. The best-fitting linear regression lines are inserted. As one can see in Figure 1 by comparing the line of actual match with the line of "veridical" match, the subjects came quite close to the veridical,

Figure 1. Matched stimulus onset asynchrony (SOA) of the adjustable pulse sequence is plotted (ordinate) as a function of the SOA of coherent tone sequences (abscissa). The line through the circles represents a veridical match, and the line through the triangles, the actual match (Equation: $-20 + 1.07 \text{ SOA}$; r^2 for least squares fit is .997).

Figure 2. Matched stimulus onset asynchrony (SOA) of the adjustable pulse sequence for one stream of streaming tone sequences is plotted (ordinate) as a function of the SOA of streaming sequences (abscissa). The line through the open circles represents a veridical match, and the line through the triangles, the actual match (Equation: $-21 + 1.96 \text{ SOA}$; r^2 for least squares fit is .994).

although they slightly overestimated the rate of the coherent sequences.

It is quite evident from Figure 2 that the subjects matched twice the SOA between tones, that is, the tempo of the attended stream rather than the tempo of the physical sequence.

An analysis of variance verifies that there was a significant difference between the adjusted SOAs for the streaming and coherent conditions [$F(1,17) = 765.27$, $p < .0001$]; and a significant interaction between coherence vs. streaming and SOA [$F(2,34) = 49.97$, $p < .0001$] reflects the fact that the slope of the linear regression line in Figure 2 is much steeper than the slope in Figure 1. It should be stated, however, that the match with the tempo of the attended stream is not perfect. The subjects adjusted an SOA that was about 25-30 msec shorter

than the duration between the tones in one stream. An explanation might be that the kind of adjustment procedure utilized prevented the listener from giving an adequate estimate. If the switching back and forth between the streaming sequence and the pulse sequence was done too quickly by some subjects, then streaming might not have been fully developed, because, as is known, it takes a few cycles before one gets into the streaming mode (Bregman, 1978b). This possibility has been pointed out by van Noorden (Note 1). The implication of such an explanation, however, is that some of the individual adjustments should approximate one SOA, but this was not the case. Another explanation might be that the adjustment procedure is of a successive, instead of simultaneous, nature. Listeners have to rely on memory for the tempi in order to accomplish the task. It might be that unattended background tones still exerted some influence by increasing the subjective rate of the tones in the attended stream.

The main conclusion from this experiment, however, is that subjects are capable of approximating the tempo of one stream by means of the tempo-matching task, a capability not revealed by the stop-RT paradigm of our previous study (ten Hoopen, van Meurs, & Akerboom, 1982).

One question remained: if required to, were subjects also able to match the total rate of a streaming sequence? Experiment 2 explored this question.

EXPERIMENT 2

The subjects were required (1) to match the rate of a pulse sequence with the rate of coherent tone sequences, (2) to match the rate of a pulse sequence with the rate of one stream of streaming tone sequences, and (3) to match the rate of a pulse sequence with the total rate of streaming tone sequences, that is, to match the SOA of a pulse sequence with the SOA between physically adjacent tones of a streaming sequence.

Method

The method was the same as that of Experiment 1, except that a different group of 18 subjects was used. In addition to receiving the coherent sequences, they received each of the three streaming sequences four times, twice with the instruction to match the rate of a single stream and twice with the instruction to match the total rate.

Results and Discussion

The adjusted SOAs of the pulse sequence, averaged over replications and subjects, are displayed in Figure 3 and Figure 4 according to type of sequence, instruction, and SOA. Figure 3 displays the means of the coherence conditions through which the best-fitting linear regression line is drawn. Comparison with the line of "veridical" match shows

Figure 3. Matched stimulus onset asynchrony (SOA) of the adjustable pulse sequence is plotted (ordinate) as a function of the SOA of coherent tone sequences (abscissa). The line through the circles represents a veridical match, and the line through the triangles, the actual match (Equation: $-2 + .99 \text{ SOA}$; r^2 for least squares fit is .997).

that the subjects' line nearly perfectly matches the rate of the coherent sequences. These results correspond rather well with those of Experiment 1 (cf. Figure 1).

Figure 4, lower panel, illustrates the ability of the subjects to match the total rate of the streaming sequences. Hence, although the sequences are streaming, subjects are capable of establishing the duration between physically adjacent tones. This is the same result that was found with our stop-RT paradigm.

As was found in Experiment 1, subjects, when instructed to, are able, to a considerable degree, to match the tempo of the attended stream. Figure 4 (upper panel) shows that the subjects in Experiment 2 approach twice the SOA even better than did the subjects in Experiment 1 (cf. Figure 2).

GENERAL DISCUSSION

The joint conclusions from the present study and our previous one (ten Hoopen, van Meurs, & Akerboom, 1982) are: observers are quite capable of matching the rate of coherent sequences. They are also quite capable of matching the rate of each stream of a segregated sequence. These capabilities are consonant with auditory stream formation theory (McAdams & Bregman, 1979). Coherent sequences are mentally interpreted as a continuous whole; that is, they appear to emanate from *one* source. It is a fundamental property of the auditory system that the rate of such a sequence can be judged and reproduced accurately: otherwise, speaking or playing music would be im-

Figure 4. Lower panel: Matched stimulus onset asynchrony (SOA) of the adjustable pulse sequence for the total rate of streaming tone sequences is plotted (ordinate) as a function of the SOA of streaming sequences (abscissa). The line through the circles represents a veridical match. The line through the triangles represents the actual match (Equation: $0 + 1.0 \text{ SOA}$; r^2 for least squares fit is .974). Upper panel: Matched stimulus onset asynchrony (SOA) of the adjustable pulse sequence for one stream of streaming tone sequences is plotted (ordinate) as a function of the SOA of streaming sequences (abscissa). The line through the circles represents a veridical match, and the line through the triangles, the actual match (Equation: $-72 + 2.34 \text{ SOA}$; r^2 for least squares fit is 1.00).

possible. Why the auditory system should also be capable of establishing the overall rate of a streaming sequence remains unclear, however. The streams appear to emerge from two different sources and to have no relation to each other. Yet, as our stop-RT paradigm and the matching task demonstrated, observers show the ability to "weave" the separate streams into the "tempo processor" and to establish the interlacing time grid. This fact seems to be in conflict with studies in which listeners had to judge the temporal order of sounds. Judgment of order was much better when the sounds belonged to the same stream than when the sounds belonged to different streams (Bregman & Campbell, 1971; Warren & Obusek, 1972). However, this latter task requires that the *names* of the sounds be correctly ordered in time, which is evidently possible only if the names belong to the same stream. Thus, in judging the overall rate of the segregated sequence, the names are not picked up by the tempo processor; that is, information about the sounds other than that required for identification is sufficient for judging rate. The fact that only one stream of a segregated sequence can be attended to and yet the overall rate of the sequence can be established implies that the "parsing processor" and the "tempo processor" have different sites in the auditory system. These processors are, however, strongly related, for the parsing processor needs, among other things, the output from the tempo processor in order to interpret the auditory scene. The present experiments show, however, that the parsing processor does not "destroy" or recode the original temporal information contained in the sequence of tones. The temporal information remains available in parallel with the percepts generated by the parsing processor. A question for further research is whether the present patterns of results can

also be found when the streams do not comprise simple tones but more complex sounds or sounds that differ along several dimensions.

REFERENCE NOTE

1. Van Noorden, L. P. A. S. Personal communication, March 2, 1982.

REFERENCES

- AXELROD, S., & GUZY, L. T. Underestimation of dichotic click rates: Results using methods of absolute estimation and contrast stimuli. *Psychonomic Science*, 1968, 12, 133-134.
- AXELROD, S., GUZY, L. T., & DIAMOND, I. T. Perceived rate of monotonic and dichotically alternating clicks. *Journal of the Acoustical Society of America*, 1968, 43, 51-55.
- BREGMAN, A. S. The formation of auditory streams. In J. Requin (Ed.), *Attention and Performance VII*. Hillsdale, N.J.: Erlbaum, 1978. (a)
- BREGMAN, A. S. Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4, 380-387. (b)
- BREGMAN, A. S., & CAMPBELL, J. Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 1971, 89, 244-249.
- MCADAMS, S., & BREGMAN, A. S. Hearing musical streams. *Computer Music Journal*, 1979, 3, 26-43.
- TEN HOOPEN, G., VAN MEURS, G., & AKERBOOM, S. P. The perceived tempi of coherent and streaming tone sequences. *Perception & Psychophysics*, 1982, 31, 256-260.
- TEN HOOPEN, G., VOS, J., & DISPA, J. Interaural and monaural clicks and clocks: Tempo difference versus attention switching. *Journal of Experimental Psychology: Human Perception and Performance*, 1982, 8, 422-434.
- VAN NOORDEN, L. P. A. S. *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Institute of Perception Research, Eindhoven, The Netherlands, 1975.
- WARREN, R. M., & OBUSEK, C. J. Identification of temporal order within auditory sequences. *Perception & Psychophysics*, 1972, 12, 86-90.

(Manuscript received March 24, 1982;
revision accepted for publication July 7, 1982.)