The perceived tempi of coherent and streaming tone sequences

GERT ten HOOPEN, GERARD van MEURS, and SIMONE AKERBOOM
University of Leiden, Unit of Experimental Psychology, Leiden, The Netherlands

Previous research has demonstrated that a sequence of tones that is alternated between ears is stretched out in echoic memory, as compared with monaural sequences. Although the stimulus onset asynchronies (SOAs) were the same in the interaural and monaural conditions, the perceptual onset asynchronies (POAs) differed by 24 msec. The present study investigated whether an analogous perceptual phenomenon exists for frequency-alternating tone sequences. It turned out that the POAs of coherent and even of streaming tone sequences were precisely the same as with nonalternating tone sequences.

If an isochronous sequence of auditory signals is presented interaurally (i.e., each consecutive signal arrives at the other ear), such a sequence has a slower apparent tempo than a monaural sequence (Axelrod & Guzy, 1968; Axelrod, Guzy, & Diamond, 1968). We devoted a series of studies to this phenomenon and were finally able to establish that the amount of the subjective tempo difference between interaural and monaural sequences was 24 msec (ten Hoopen & Vos, 1979a, 1979b, 1980, 1981; ten Hoopen, Vos, & Dispa, in press). Although the value of 24 msec can be arrived at by three different methods, we will discuss only the most powerful one (see ten Hoopen et al., in press, for the complete discussion of all three methods).

Listeners were required to respond as fast as possible to the unpredictable end of a tone sequence. The so-called stop-reaction time (stop-RT) comprises two durations. The first duration is the perceptual onset asynchrony (POA) between the tones. This is the criterion time that the subject sets for himself after each tone, and, if no further tone arrives during that time, he initiates his response, a process taking an additional amount of time—the second duration. If it is assumed that response initiation time is invariant with the presentation mode (interaural vs. monaural), then one can interpret differences between stop-RTs as differences between POAs.

From our ten Hoopen et al. (in press) study, it turned out that the difference between interaural and monaural POA was 24 msec. In addition, it turned out that this difference was invariant with the stimulus onset asynchrony (SOA) between the tones, at least in the range of 125-250 msec SOA. From further

studies at our laboratory (ten Hoopen, van der Schaaf, Vos, & Akerboom, Note 1; ten Hoopen, Akerboom, Verduin, Vos, & van der Schaaf, Note 2), it appeared that this invariance holds even for a range of 40-2,100 msec SOA. In other words, no matter how fast or slow the objective tempo of the sequence was, the subjective difference between interaural and monaural tempo remained 24 msec. Schaefer (Note 3), who actually developed the stop-reaction-time paradigm, found a 26-msec difference and also established invariance of this tempo difference with SOA.

Figures 1 and 2 represent our conception of the difference between interaural and monaural POA. The essence of this representation is that interaural events are stretched out in auditory memory for 24 msec more than monaural events, although for both conditions the SOA is the same. Hence, each time that a
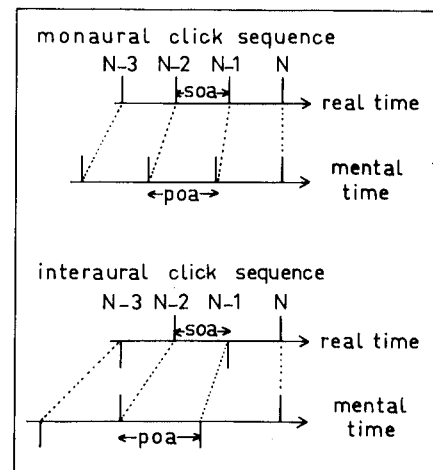
Figure 1. Monaural and interaural sequences with the same stimulus onset asynchrony (SOA) between tones (in real time), but having different perceptual onset asynchronies (POA) (in mental time). When listening "back" in echoic memory, interaural tones appear more widely spaced.
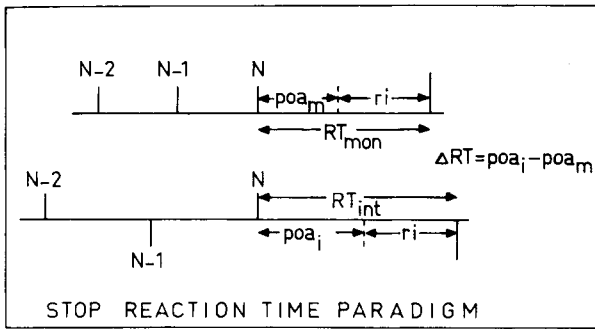
Figure 2. Illustration of the stop-reaction-time paradigm, which shows that the difference in time between detecting the end of interaural and monaural tone sequences ($RT_{int} - RT_{mon}$) represents the difference between interaural and monaural perceptual onset asynchronies ($POA_i - POA_m$).

new tone arrives, the sequence of interaural tone percepts in echoic memory is pushed back 24 msec more than in the monaural condition. At present, we can make only vague guesses about the purpose of such a perceptual mechanism.[1]

Thus, we decided first to investigate whether the analogous phenomenon exists with regard to other dimensions of auditory space. In the last decade, an increasing number of studies have investigated the perceptual effects of alternating tone sequences between different frequencies (van Noorden, 1975; also see Bregman, 1978, and McAdams & Bregman, 1979, for overviews). The main interest has been directed toward the phenomenon of auditory streaming. If a tone sequence is alternated between frequencies, one of two possible sensations can be experienced. Either the sequence is heard as a coherent whole or it is heard as two separate streams of tones, one of high tones and one of low tones. Which sensation arises depends on the SOA and frequency interval between the tones and, for some combinations, also on what the listener wants to hear. Van Noorden (1975) has systematically investigated which combinations of

Figure 3. Stylized redrawing of van Noorden's (1975) Figure 2.9. See text for explanation.

SOA and frequency interval cause the percepts of coherence and streaming (called fission by him). It is worthwhile to depict a redrawing of van Noorden's Figure 2.9, shown here as Figure 3.

Below the fission boundary, the listener always hears coherence. Above the temporal coherence boundary, the listener always hears streaming (i.e., fission). Between the two boundaries, it depends on the intention of the listener whether he hears coherence or streaming.

In analogy to the location-alternating situation described above, one might ask what the perceived tempi of coherent and streaming tone sequences are relative to nonalternating ones. For coherent sequences, this is merely a question of investigative nature: Are frequency-alternating (but coherent) tone percepts more stretched out in auditory memory, like location-alternating tone percepts, or not? For streaming tone sequences, two interesting hypotheses can be coined about the perceived tempo. We recall that if a listener is required to respond to the unpredictable end of a tone sequence, the stop-RT is indicative of the POA between the tones. The stop-RT paradigm allows us to infer whether the tempo of a streaming sequence is dictated only by the foreground stream (to which attention is directed) or by foreground and background streams together. In the former case, a POA should be found that corresponds to twice the SOA. In the latter case, a POA should be found that corresponds to one SOA.

METHOD

Subjects

Eighteen subjects (students and staff members of the psychology department) participated in the experiment. All subjects attended two sessions. Left- and right-ear thresholds were determined for pure tones of 500 and 2,000 Hz. The rejection criteria were either a loss of 20 dB or a difference between the two thresholds for the tones of more than 10 dB.

Stimuli and Design

Sequences of pure tones were generated with Coulbourn Instruments modules controlled by a microprocessor. All tones had a duration of 45 msec and thus had a tonal character. In order to avoid loudness changes due to frequency changes, the center frequency for all sequences was 1,000 Hz; in this frequency area, the equiloudness curves are relatively flat (see, e.g., Roederer, 1975, Figure 3.12). The number of tones in a sequence varied randomly from 41 to 50. SOAs varied from 90 to 140 to 190 msec. Each sequence was followed by a silent response interval of 3 sec.

All sequences were recorded on magnetic tape by a Revox stereo tape recorder (Type B 77) at a speed of 19 cm/sec.

A block consisted of 10 sequences of tones. Within a block, the presentation mode and the SOA were kept constant; only the number of tones in the sequences varied. The number of tones (41 to 50) in each of the 10 sequences was randomized.

Apart from the training sequences, the subject was presented with 12 experimental blocks at each of the two sessions: six streaming blocks, three coherence blocks, and three nonalternating blocks.

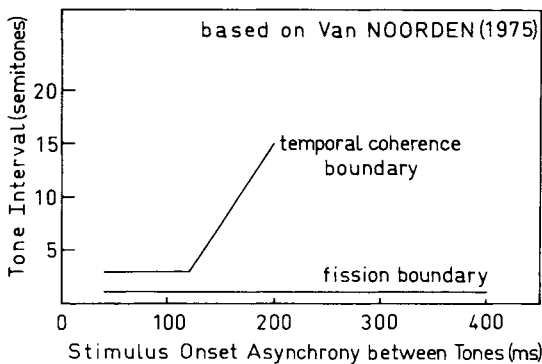In the streaming blocks, the sequences consisted of alternating

high (1,782 Hz, 10 semitones above the center frequency of 1,000 Hz) and low (561 Hz, 10 semitones below the center frequency of 1,000 Hz) tones. The interval of 20 semitones in combination with each of the three SOAs is in van Noorden's area of "inevitable fission" (i.e., streaming; see Figure 3). For each of the three SOAs, two streaming blocks were constructed—one in which all sequences started with the higher tone, and one in which all sequences started with the lower tone. Thus, a total of six [3 SOAs × 2 (first tone high/low)] streaming blocks were constructed. The fact that six streaming blocks instead of three (as in the coherence and nonalternating presentation modes) were presented will be explained in the results. The three coherence blocks consisted of alternating high (1,029 Hz, half a semitone above the center frequency of 1,000 Hz) and low (972 Hz, half a semitone below the center frequency of 1,000 Hz) tones. The interval of 1 semitone in combination with each of the three SOAs results in "inevitable coherence" that is below the fission boundary (see Figure 3). The three nonalternating blocks consisted of sequences of 1,000-Hz tones.

Procedure and Apparatus

There are six possible orders for the three presentation modes (i.e., streaming, coherence, and nonalternating). Three subjects were assigned to each of these six permutations. Within the presentation modes, the orders of the three SOAs were balanced as far as possible. This resulted in a unique presentation order of the 12 experimental blocks for each of the 18 subjects. In the second session, the presentation orders for the 18 subjects were reversed.

Subjects were tested individually in a Siemens sound-attenuated booth. After the determination of the hearing thresholds of both ears (with a Peekel pure tone screening audiometer, Type D66), the subject sat down at a table mounted with a push button. The tapes were played back with the Revox recorder over AKG headphones (Type Cardan K140S).

The instruction required the subject to react as fast as possible with a push-button response (dominant index finger) when the sequence stopped. After several examples in the nonalternating and coherence presentation modes, the phenomenon of streaming was introduced. It was demonstrated to the subject that in the streaming mode it would no longer be possible to keep track of each of the successive tones and that either the higher or the lower tones would appear in the foreground, with the other tones in the background. After these introductory remarks, six streaming sequences were given (two of each SOA) in order to find out whether the subject had a particular listening preference for either the high or the low stream. If the subject had such a preference, he was asked to listen to all the streaming blocks in the two experimental sessions according to this preference. If the subject had no clear preference, he was asked to listen to all streaming blocks with the high tones in the foreground (i.e., to the high stream). After these instructions, the 12 experimental blocks were presented. Between the sixth and the seventh blocks, there was a 5- to 10-min rest.

The first tone of each sequence triggered a timer (Gould advance timer counter, Type TC311) by means of a voice-operated relay (Grason-Stadler, Type E7300A-1). As soon as the subject gave his push-button response, the timer stopped, and the time was printed out (Printina Gay alphanumerical printer, Type CP).

Because the latency of the response was measured from the onset of the first tone in the sequence, each latency was reduced by its corresponding sequence duration [$(N-1)SOA + 45$ msec, where N is the number of tones in the sequence and 45 msec is the duration of a tone].

RESULTS

The first session was regarded as a practice and training session, and therefore only the data of the second session were analyzed.

In the streaming condition, the subjects were instructed to focus their attention on the stream that they preferred: the stream of either high or low tones. The last tone of a sequence was either a tone of the attended stream or a tone of the nonattended stream. Sequences that ended with a tone of the attended stream were called congruent; sequences that ended with a tone of the nonattended stream were called incongruent.

The six streaming blocks were divided into three streaming-congruent and three streaming-incongruent blocks. This division could only be established post hoc, since it was dependent on the preference of the listener, the odd- or evenness of the number of tones in the sequences, and whether the first tone of the sequence was high or low.

We divided the streaming sequences into congruent and incongruent blocks because this might possibly be a discriminating factor in the stop-RT task. However, in a separate analysis on the streaming data (congruent-incongruent, SOA, and number of tones as fixed factors; all repeated measures), we found no difference between the congruent and incongruent stop-RTs (346 and 341 msec, respectively). Since streaming-congruent and streaming-incongruent stop-RTs did not differ significantly, we decided to drop this distinction in the further analysis and to average over congruent and incongruent RTs.

We then performed a 3 (presentation mode) × 3 (SOA) × 10 (number of tones) analysis of variance (presentation mode, SOA, and number of tones as fixed factors; all repeated measures). The effect of presentation mode did not reach significance [$F(2,34) = 2.69$, $p > .08$]. The mean stop-RTs for streaming, coherent, and nonalternating sequences were 343, 338, and 336 msec, respectively. The effect of the number of tones in the sequences was not significant either ($p > .71$). Not surprisingly, the effect of SOA was very significant [$F(2,34) = 672.62$, $p < .00001$]. The interaction between presentation mode (streaming, coherence, nonalternating) and SOA was not significant ($p > .60$; see Table 1).

The interaction between SOA and number of tones was significant [$F(18,306) = 1.95$, $p < .05$], and so was the second-order interaction between presentation mode, SOA, and number of tones [$F(36,612) = 1.64$, $p < .05$]. These significant interaction effects could not, however, be interpreted systematically.

Table 1
Mean Stop Reaction Times (in Milliseconds), Dependent on
Presentation Mode and Stimulus Onset Asynchrony
(SOA, in Milliseconds)

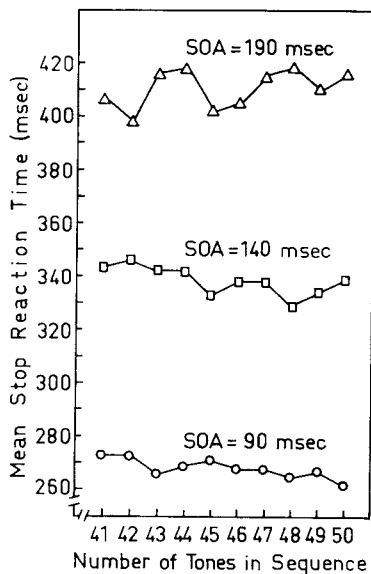| SOA	Streaming	Coherence	Nonalternating
90	274	267	264
140	343	340	333
190	412	409	410

Figure 4. Mean stop-reaction times dependent on number of tones in the sequence and stimulus onset asynchrony (SOA).

The SOA × number of tones interaction is depicted in Figure 4.

DISCUSSION

The results are very clear-cut. There is no difference between the POAs of nonalternating and alternating (but coherent) tone sequences. Informal listening tests are in accordance with this result: No tempo difference can be heard. Thus, we have to conclude that frequency-alternating (but coherent) tone sequences are spaced out in auditory memory to the same extent as nonalternating tone sequences.

With regard to the streaming tone sequences, the second hypothesis that the POA corresponds to one SOA was supported by our results, because there was no difference between the POAs of the streaming and nonalternating tone sequences. In our opinion, a rather important conclusion can be drawn from this result. Although attention can be directed toward only one part of the whole sequence—namely, the high *or* the low stream—this locking of attention does not prevent the picking up of the nonattended tones into the "tempo machine," and hence the listener is able to judge the POA between temporally adjacent tones, although they belong to different streams.

So, it seems that, although there is an auditory mechanism that for some reason forms two (or more) streams from one physical tone sequence, there must be another mechanism that at the same time is able to extract information from the tone sequence as a whole in order to establish the POA between adjacent tones..

As can be seen from Table 1, an SOA increment of 50 msec results in a stop-RT increment of about 70 msec for all three presentation modes. One should, however, not be tempted to infer absolute values for POAs from these data. On the one hand, it is quite conceivable that the larger the SOA is, the larger the just noticeable difference (jnd) for the corresponding POA. On the other hand, it is quite conceivable that the repetition rate of the tones affects the response initiation process in the sense that faster repetition rates (shorter SOAs) speed up response initiation time more than do slow repetition rates. The stop-RT paradigm is only valid in detecting differences in POAs between different presentation modes at the same level of SOA.

In a recent study by Massaro (1977), subjects had to count the number of tones in nonalternating and frequency-alternating sequences. It turned out that the number of tones was less accurately counted if the sequence alternated between frequencies that differed with 14 semitones (which implies a streaming sensation at most of the SOAs utilized by Massaro). Counting accuracy was, however, not further deteriorated if the frequencies alternated over still larger intervals (38 semitones). If the frequencies alternated over an interval of 2 semitones, no deterioration of counting accuracy could be established, in comparison with the nonalternating condition.

Thus, it appears to be a warranted conclusion of Massaro that it was not the frequency alternations as such, but rather the sensation of streaming, that hampered counting performance.

In our ten Hoopen et al. (in press) study, we demonstrated that the speed at which tones in a sequence are counted is paced by the POA (i.e., not by the objective tempo but by the subjective tempo of the sequence). This is precisely why the number of interaurally presented tones is undercounted relative to the monaural condition, and the degree to which this occurs actually is one of the methods by which the POA difference of 24 msec between interaural and monaural sequences can be established.

If the POA explanation for the number-counted results also holds in the case of frequency-alternating tone sequences, one should expect no difference between the streaming and nonalternating conditions, because the present stop-RT experiment showed convincingly that the POAs did not differ. Yet, Massaro found deteriorated counting performance of streaming tone sequences. Thus, it seems that the POA explanation, valid with regard to ambient space, may not be so with regard to frequency space.

However, it should be noted that Massaro (1977) reported only the percentages of tone sequences correctly counted in number—that is, counting accuracy and not the mean numbers of tones counted, indicative of the counting speed. In view of our POA concept, it might be quite conceivable that the mean

numbers counted at the frequency-alternating and the nonalternating conditions do not differ because the respective POAs (pacing the counting speed) are the same, whereas counting accuracy does differ between both conditions. The reason for the discrepancy between the behavior of these two dependent variables might stem from the fact that it is difficult to correctly judge the temporal order of adjacent tones because they belong to different streams (Bregman & Campbell, 1971). The accuracy of counting—heavily dependent on the precise labeling of the consecutive tones by the consecutive rank numbers—might be strongly interfered with by the poor perception of order.

Let us summarize the main finding of this study and contrast it with that of our ten Hoopen et al. (in press) study. In the case of alternating tones between frequencies, streams might be formed, but, at the same time, a tempo mechanism computes the same POA as for nonalternating tone sequences. In the case of alternating tones between the ears, the tempo mechanism adds about 24 msec to the POA, as compared with nonalternating tone sequences. It should be a point of further research whether, at the same time as this POA stretching, the ear-alternating tones also stream or that such a "streaming by ear" is even the cause of the POA stretching. Some evidence for streaming by ear has been given by Judd (1979), although some frequency separation between the ear-alternating tones seems to be necessary to facilitate streaming.

Another point of further research should be the following: The present stop-RT results showed that listeners are capable of correctly establishing the POA between adjacent tones, although these tones belong to different streams. Such a result, however, does not exclude the possibility that listeners, at the same time, are also capable of judging the tempo of the stream to which they are attending. In a tempo-matching task, they should then adjust twice the POA.

REFERENCE NOTES

1. ten Hoopen, G., van der Schaaf, T. W., Vos, J., & Akerboom, S. P. The echo store viewed through a time window. Unpublished manuscript, Unit of Experimental Psychology, University of Leiden, The Netherlands, 1981.
2. ten Hoopen, G., Akerboom, S. P., Verduin, C. J., Vos, J., & van der Schaaf, T. W. The danger of taking ratios or how to obscure the interaural tempo delay: Comments on Huggins (1974) "On perceptual integration of dichotically alternated pulse trains." Manuscript submitted for publication, 1981.
3. Schaefer, F. Gerichtete und verteilte Aufmerksamkeit bei der

Einschätzung richtungsalternierender Clicks. Paper presented at the 21st meeting of the Experimentell arbeitender Psychologen, Heidelberg, 1979.

REFERENCES

AXELROD, S., & GUZY, L. T. Underestimation of dichotic click rates: Results using methods of absolute estimation and constant stimuli. Psychonomic Science, 1968, 12, 133-134.

AXELROD, S., GUZY, L. T., & DIAMOND, I. T. Perceived rate of monotic 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.

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.

JUDD, T. Comments on Deutsch's musical scale illusion. Perception & Psychophysics, 1979, 26, 85-92.

MASSARO, D. W. Perceptual grouping in audition. Perception, 1977, 6, 541-553.

McADAMS, S., & BREGMAN, A. S. Hearing musical streams. Computer Music Journal, 1979, 3, 26-43.

ROEDERER, J. G. Introduction to the physics and psychophysics of music. New York: Springer-Verlag, 1975.

TEN HOOPEN, G., & VOS, J. Attention-switching and grouping in counting interaurally presented clicks. Acta Psychologica, 1979, 43, 283-297. (a)

TEN HOOPEN, G., & VOS, J. Effect on numerosity judgment of grouping of tones by auditory channels. Perception & Psychophysics, 1979, 26, 374-380. (b)

TEN HOOPEN, G., & VOS, J. Attention switching is not a fatigable process: Methodological comments on Axelrod and Guzy (1972). Journal of Experimental Psychology: Human Perception and Performance, 1980, 6, 180-183.

TEN HOOPEN, G., & VOS, J. Attention switching and patterns of sound locations in counting clicks. Journal of Experimental Psychology: Human Perception and Performance, 1981, 7, 342-355.

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, in press.

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.

NOTE

1. Yet, the finding that the interaural POA is 24 msec longer than the monaural POA explains several results reported in the literature: the underestimation of interaural tempi, the undercounting of the number of interaurally presented tones, the delay in monitoring probe tones in interaural tone sequences, the larger jnd for perceiving anisochrony of interaural tone sequences, and the diminished recall performance of interaural lists of words.