Perception, 1977, volume 6, pages 541-553

Perceptual grouping in audition

Dominic W Massaro
Department of Psychology, University of Wisconsin, Johnson at Charter Streets, Madison, Wisconsin 53706, USA
Received 23 August 1976, in revised form 22 February 1977

Abstract. The present experiments evaluated the effect of relative frequency as a determinant of the figure-ground organization of sequences of auditory tones. Observers counted sequences of 20 ms tones that were presented at the same frequency or that alternated between two different frequencies. The alternating tones differed in frequency by one whole tone, seven tones, or nineteen tones. Counting accuracy increased with increases in the silent interval between the tones. When the alternating tones differed by seven or nineteen tones, counting was disrupted at rates of presentation of eight tones per second or slower. In contrast to this decrement in the counting of tones that alternated by over an octave, very little decrement was observed when the tones alternated by just one whole tone. The best subjects counted these alternating tones more accurately than the tones presented at the same frequency. The poorest subjects showed a small decrement even when the tones alternated by just one whole tone. The results were discussed in terms of determinants of figure-ground organization in auditory information processing.

1 Introduction

The Gestalt psychologists are best known for their articulation of the laws of figure-ground organization and perceptual grouping (Koffka 1935; Kohler 1929). The treatment appears to have been carried out almost entirely in terms of our visual experience, however, without much study of these phenomena in other modalities. While acknowledging the paucity of experimental data on figure-ground articulation in other senses, Koffka (1935) pointed to the obvious auditory example of hearing speech against the background of a variety of auditory noise. Koffka concluded that "the figure-ground distinction (in other senses besides vision) offers new problems... which are of great significance for the theory of behavior, but which as yet are in too embryonic a state to deserve further discussion" (p 201). Forty years later, the concern of the present paper is whether figure-ground determinants and perceptual grouping can be useful constructs in understanding auditory information processing.

One of the best examples of figure-ground organization in auditory processing is the cocktail-party phenomenon. It is possible to listen to one of a number of different speakers without much interference from the other conversations. The primary determinants of isolating the relevant speaker from the background are the speaker's voice and its spatial location (Cherry 1953; Treisman 1964). If the same discussion is heard through a single ear or on a single-channel recorder, the listener finds it much more difficult to follow the relevant message. Similarly, if the simultaneous conversations are by the same talker, one of the messages is not easy to follow. The perceived dimensions of voice pitch and spatial location can be thought of as similar to the perceived dimensions of color and spatial location in vision. The visual figure is supposedly seen more clearly and in front of the ground. Analogously, the auditory figure is heard more clearly and may be experienced as being nearer than the background 'noise'.

Spatial location and pitch are also important in the organization of auditory melodies. Spatial location has been exploited by contemporary musicians recording for stereo and quadraphonic listening. Pitch differences have been used to segregate

Each trial began with the presentation of the sequence of tones followed by a 2 s response interval. The subject made the appropriate response by hitting one of four pushbuttons labeled five, six, seven, and eight, respectively. Subjects were required to make a response on each trial. Feedback was given by a visual presentation of the digit five, six, seven, or eight for 250 ms. The intertrial interval was 500 ms. Within a given session, all experimental conditions were chosen randomly with equal probability. Subjects were instructed to count the total number of tones in the sequence, and were informed that the rate of presentation could vary and that sequences might alternate between different frequencies or be presented at the same frequency. They were explicitly told to count the total number of tones regardless of the rate of presentation and whether or not the tones alternated between frequencies.

The procedure was aimed at preventing the successful utilization of any strategy other than actually counting the tones. Four lengths were used, so knowing that the sequence had an odd or even number of tones would not be sufficient for a correct answer. All 256 possible trial types (eight sequences x eight rates x four test-sequence lengths) were equally likely to occur on any trial. This prevented subjects from learning to use the duration of a sequence as a reliable cue to the number of tones it contained. If the rate of presentation did not vary, subjects could make reliable judgments on the basis of sequence duration alone. The completely random presentation also worked against the utilization of different strategies under the different experimental conditions. Subjects could not know the rate of presentation or whether or not the tones were alternating in frequency until they heard, at least, two tones.

Two sessions of 305 trials each were given each day for five consecutive days. There was a short break between the two sessions. The first day and the first five trials of each experimental session were eliminated from the data analysis. The dependent measure was the percentage of correct judgments averaged across the four different lengths of the test sequence. The results were also pooled over the two types of sequences within each of the thirty-two experimental conditions of alternating versus same frequency x the large or small frequency difference x the eight rates of presentation. The total of 2400 observations gave about seventy-five observations per condition for each subject.

2.1.3 *Apparatus.* All experimental events were controlled by a PDP-8L computer. The tones were generated by a digitally controlled oscillator (Wavetek model 155) and were presented binaurally over matched headphones (Grason Stadler TDH-49). The feedback was given visually over displays made of light-emitting diodes (Monsanto model MDA III). Four subjects were tested simultaneously in separate sound-attenuated rooms.

2.2 *Results and discussion*

The subjects were partitioned into three groups of nine, eleven, and ten subjects, respectively, on the basis of average performance at the longest processing interval. The best subjects averaged at least 98% correct at the 275 ms processing-time condition, the performance of the middle subjects fell between 92 and 97%, whereas the poorest subjects averaged between 75 and 91% at this interval. The two remaining subjects were eliminated from the data analysis since they performed at less than 54% at the longest processing interval.

Figures 1, 2, and 3 present the percentage of correct identifications of the test sequences for the best, middle, and worst subjects, respectively. Performance is plotted as a function of processing time (the time between the onsets of successive tones), the same or alternating conditions and the small or large frequency differences.

Counting performance improved with increases in processing time but to a greater extent under the same-frequency than under the alternating-frequency conditions. Performance showed a large decrement with alternating frequencies for all three groups at rates of eight tones per second or slower. The same decrement was observed both for the large and for the small frequency differences. The results show that subjects have much more difficulty counting tones alternating in frequency relative to tones presented at the same frequency but that increasing the frequency differences by an additional two octaves provides no further decrement.

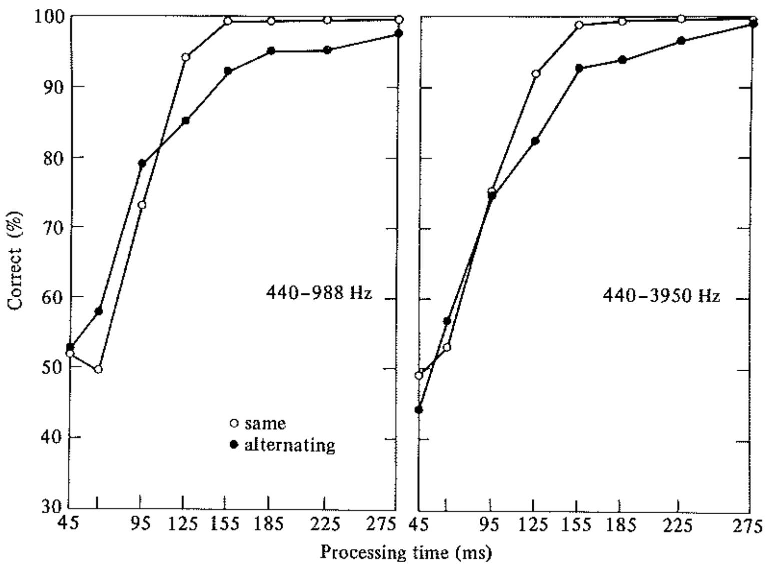

Figure 1. Percentage of correct identifications as a function of processing time for each tone for the sequences presented at one frequency or alternated between two frequencies. Results for the best nine observers.

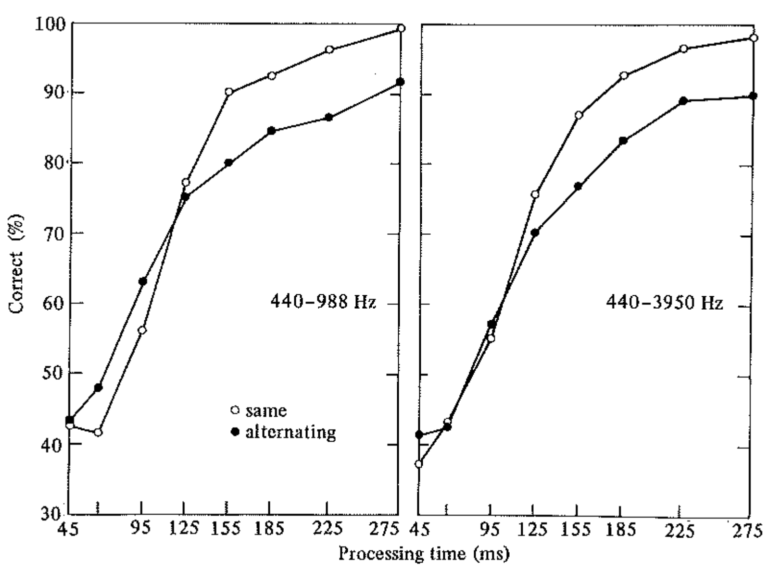

Figure 2. Percentage of correct identifications as a function of processing time for each tone for the sequences presented at one frequency or alternated between two frequencies. Results for the middle eleven observers.

No large perceptual biases were observed, in agreement with the results of Massaro (1976). The forced-choice task and feedback seems to eliminate any consistent biases that would be apparent in a task with an unrestricted set of responses and without feedback. Therefore, the negative results with respect to perceptual biases should not be interpreted to mean that no *biases* would be found if the task were structured differently. In an experiment exactly analogous to the present experiment, tones that alternated between the ears were counted less accurately than tones presented to the same ear, and no biases were observed. Guzy and Axelrod (1972) did not give feedback and found that subjects tended to underestimate the number of clicks alternating between the ears relative to the number of clicks presented to the same ear. Accordingly, we might expect that perceptual biases would also be observed with alternating frequencies if feedback is eliminated and an unrestricted response set is used.

Subjects averaged 80, 70, 68, and 70% correct on the five-, six-, seven-, and eight-tone sequences, respectively. This provides some evidence that decreasing the number of tones in a sequence facilitates counting, but also that performance on the shortest and longest sequences benefits from these sequences being at the extremes of the stimulus and response ranges. For example, if the longest sequence of eight tones is overestimated at nine tones, the subject will respond with the response eight, given that nine is not a valid response. Overestimating a sequence of six tones as one having seven tones will lead to an incorrect response since seven is a valid response alternative.

The size of the frequency differences of the alternating tones did affect performance of the best and middle subjects at the shorter processing times of 65 and 95 ms. Figures 1 and 2 show about a 7% advantage for the alternating frequencies in the 440–988 Hz condition whereas no difference between the alternating and same frequencies occurs in the 440–3950 Hz condition. The advantage at the fast rates for the 440–988 Hz tones replicates previous results found by Massaro (1976). One interpretation of this result would be that less detection masking occurs with alternating-frequency than with same-frequency tones. It is well known that detection masking, the ability of one sound to prevent the detection of the presence of another, is directly related to the similarity of the two sounds (Massaro 1972). In this case,

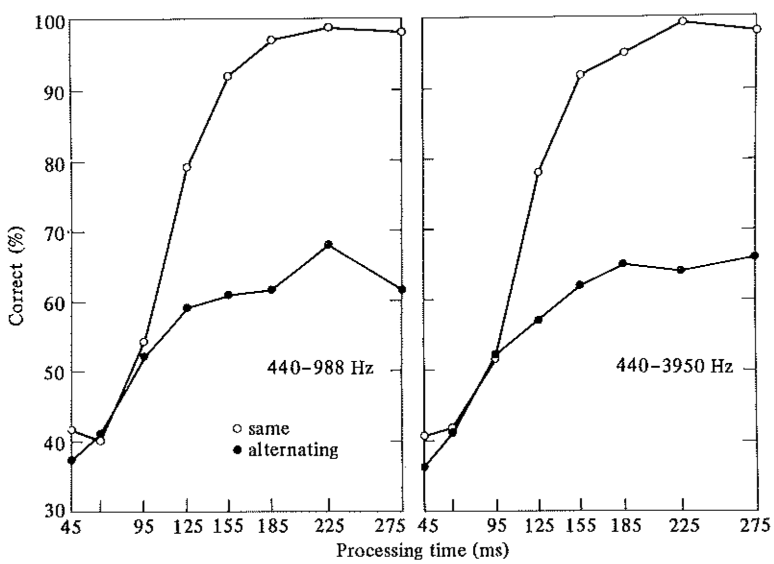

Figure 3. Percentage of correct identifications as a function of processing time for each tone for the sequences presented at one frequency or alternated between two frequencies. Results for the worst ten observers.

the alternating frequencies would show an advantage at fast rates because these tones produce less detection masking than the tones presented at the same frequency. This interpretation predicts even more of an advantage when the tones alternate by over three octaves, however, and this was not the case. No advantage was found for the alternating sequences in the 440–3950 Hz condition at the relatively fast rates.

It could be the case that the alternating up–down contour of the tones alternating in frequency can aid the counting of tones presented at relatively fast rates. In this case, backward recognition masking may be important in understanding the interaction of rate of presentation and the frequency differences of the alternating tones. The resolution of the pitch of a 20 ms tone has been shown to improve with increases in the retroactive silent interval before presentation of a second tone (Massaro 1970, 1975). The sequence of tones can be considered a series of backward-masking events in which each succeeding tone terminates the resolution of the pitch of the preceding tone. The pitch of the sounds cannot be resolved as well at the faster rates of presentation. Accordingly, it might be argued that the frequency differences at the fast rates are not heard as being as different as they would be heard at slower rates. Although the tones are alternating by slightly over an octave, the pitch differences that are actually heard at fast rates might correspond to just a couple of whole tones. The tones alternating by over three octaves are heard as being sufficiently different so that there is no advantage of the alternating condition at the fast rates.

There were significant individual differences in counting performance for both same-tone and alternating-tone sequences. When the tones were presented at the same frequency, the best subjects attained perfect performance at a processing interval of 155 ms. The middle and worst subjects required 225–275 ms of processing time to reach a perfect level of performance. These individual differences and performance levels are consonant with previous studies of the counting of tones presented at the same frequency (Garner 1951; Taubman 1950). The best subjects showed a decrement in the counting of tones alternating in frequency at processing times of 155–225 ms. The tones that alternated in frequency could be counted accurately with 275 ms of processing time per tone. The middle subjects continued to show a 10% decrement in counting the alternating tones even at the longest processing intervals. The worst subjects peaked at only 56% correct at the longest processing intervals with alternating tones but performed perfectly at these same intervals with tones presented at the same frequency.

It should be stressed that the poor counting performance of the worst subjects in the alternating condition is not due to a simple failure of motivation in the task as a whole. They counted tones perfectly when they were presented at the same frequency at the long processing intervals. It is possible that these subjects simply counted the sounds presented at one of the two frequencies and multiplied this number by two. Given that an odd number of sounds was as probable as an even number, the expected performance given this strategy would be about 50%. This simple strategy seems to be an unlikely one given that subjects showed no bias to respond with an even number of sounds. It appears that there is something inherently difficult about counting the total number of sounds that alternate between different octaves. Although all subjects showed a decrement in counting tones alternating in frequency, the size of the recovery at longer processing intervals varied with individual observers.

One might argue that subjects might favor a different strategy of counting tones that alternate in frequency from that used to count tones of the same frequency. For example, subjects might count the number of tones of the first frequency that is presented, double it, and add one if the last tone differs in pitch from the first tone. As mentioned earlier, this experiment was designed to prevent subjects from using different strategies of counting in the different conditions as much as possible.

It might be of interest, however, to present the same-frequency versus alternating-frequency conditions between different blocks of trials so that subjects could utilize different strategies under the different conditions. It is predicted that a similar decrement will be found with alternating frequencies, since my interpretation is that the present results measure a true processing deficit.

The reader may find it surprising that alternating the tones between frequencies made them more difficult to count. If perceptual grouping by frequency did occur, the subjects should have been able to count the tones in one group, then within the other, and add the two numbers. This strategy would have been analogous to that in the dichotic split-span experiments in which three digits are presented in one ear simultaneously with three digits presented to the other ear (Broadbent 1958). Subjects in this task prefer to report the digits by ear of presentation rather than by strict temporal order and are more accurate in the former than in the latter recall strategy. However, this result is not unique to the dichotic presentation, but appears to be due to a preference of the auditory system to organize simultaneous inputs by successive rather than by strict temporal order (Parkinson et al 1974; Savin 1967). For example, Treisman (1970) asked subjects to recall both dichotic and binaural six-item lists as either two triplets of successive items or as three pairs of simultaneous items. Subjects recalled about 15% more digits in successive recall than in temporal recall in both the dichotic and the binaural presentations.

Another verbal task that serves as a better analogy to the present counting experiments has been carried out by Moray (1960) and Treisman (1971). In this task subjects were asked to recall a single list of items presented to one auditory location or alternated between two locations. For example, Treisman (1971) presented six or eight digits that were alternated between the ears or were presented to both ears. Recall was significantly poorer when the digits alternated between the ears. This result is exactly analogous to the decrement observed in the present experiment when the tones are alternated between the ears relative to being presented to the same ear (cf Massaro 1975).

The first experiment showed that the counting of tones that alternate in frequency is more difficult than counting tones presented at the same frequency. Furthermore, a frequency separation of slightly over one octave was just as difficult as a frequency separation of slightly over three octaves. It is possible that the detrimental effect of a frequency separation in the counting task reaches its maximum at separations of slightly over an octave. There is a substantial amount of evidence showing that successive tones differing by an octave or more are much more difficult to process than successive tones within the same octave (Bregman and Campbell 1971; Idson and Massaro 1976). Idson and Massaro required subjects to recognize a melody of three tones when each tone was followed by another extraneous tone. If the extraneous tones were outside the octave of the melody, recognition performance was as good as a condition in which the melody was presented alone. If the extraneous tones were within the octave of the melody, however, recognition performance decreased by at least 25%. These results show that tones presented in one octave can be processed without interference from tones of another octave, whereas tones from the same octave are processed together. The psychological separation of the tones from different octaves facilitated performance in the Idson and Massaro studies since the test melody could be processed more accurately without the interference of extraneous tones outside the octave. The separation of the tone frequencies in experiment 1 interfered with counting performance because the tones from different octaves had to be processed together. According to this analysis, tones with similar frequencies are processed together, and counting performance should be as good as the single-frequency case. If tones alternate between frequencies that are less than an octave apart, we might expect much less of a decrement in counting performance than was observed in the first experiment. Experiment 2 compared the performance decrement in counting tones alternating by one whole tone to that found in counting tones alternating by over three octaves.

3 Experiment 2

3.1 *Method*

3.1.1 *Subjects.* Eleven subjects participated an hour a day for five consecutive days as an option to fulfill an introductory psychology course requirement.

3.1.2 *Procedure.* Experiment 1 was replicated exactly except that the A_4–B_5 conditions in that study were replaced with the tones A_5 (880 Hz) and B_5 (988 Hz). The design of the study allowed a direct comparison of the counting of tones alternating by one whole tone to the counting of tones alternating by slightly over three octaves (nineteen whole tones). The tones in a given sequence were presented at the same frequency on half of the trials, and alternated between two frequencies on the other half of the trials. When all of the tones were presented at the same frequency they were equally likely to be presented at A_4 (440 Hz), A_5 (880 Hz), B_5 (988 Hz), or B_7 (3950 Hz). When the tones alternated between frequencies they alternated between A_5 and B_5 on half of the trials, and A_4 and B_7 on the other half of the trials. For both sequences, there was an equal likelihood of either tone starting the sequence. Therefore, there were four kinds of alternating trials (A_5B_5..., B_5A_5..., A_4B_7..., B_7A_4...) that occurred with equal probability. The tones A_4, A_5, and B_5 were heard at 86 dB SPL, and B_7 at 75 dB SPL. All other experimental details were exactly the same as in experiment 1.

3.2 *Results and discussion*

Figures 4, 5, and 6 give the results for the best three, middle four, and worst four subjects. The subjects were partitioned into the three groups on the basis of performance at the longest processing interval. The best, middle, and worst subjects averaged between 98 and 100%, 92 and 95%, and 74 and 83%, respectively. The

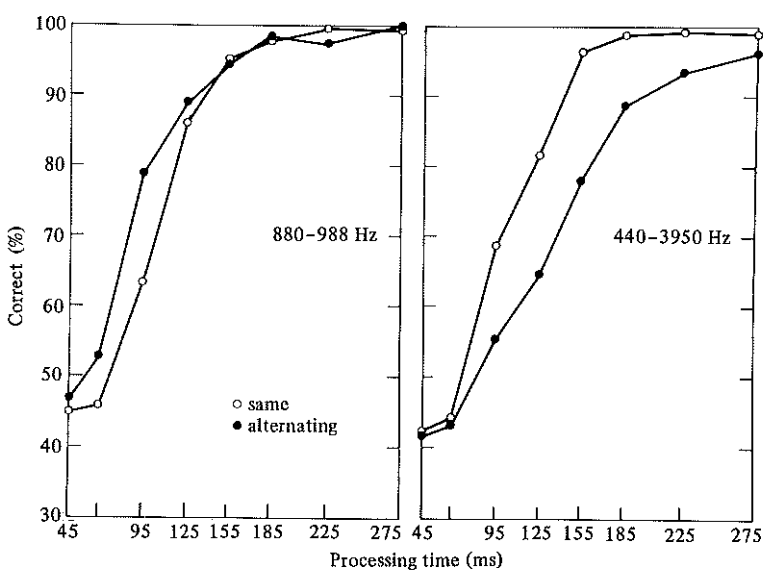

Figure 4. Percentage of correct identifications as a function of processing time for each tone for the sequences presented at one frequency or alternated between two frequencies. Results for the best three observers.

results replicated those of experiment 1 when the tones differed by slightly more than three octaves. In contrast, the best and middle subjects showed no decrement in counting the tones that alternated by one whole tone relative to counting those tones presented at the same frequency. The worst subjects showed some decrement in counting the tones that alternated by one whole tone but significantly less than the decrement observed when the tones alternated by over three octaves.

Tones that were presented are roughly ten per second (95 ms processing time) produced a surprising result for the best three subjects. These subjects showed a

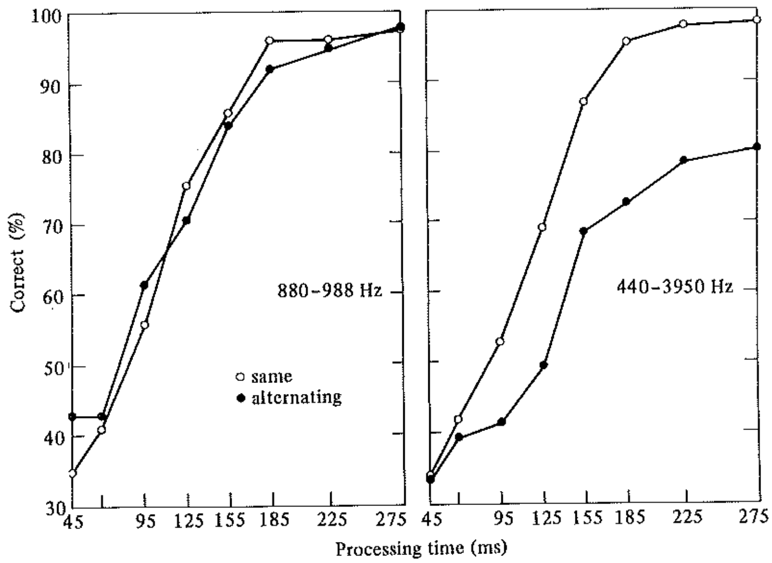

Figure 5. Percentage of correct identifications as a function of processing time for each tone for the sequences presented at one frequency or alternated between two frequencies. Results for the middle four observers.

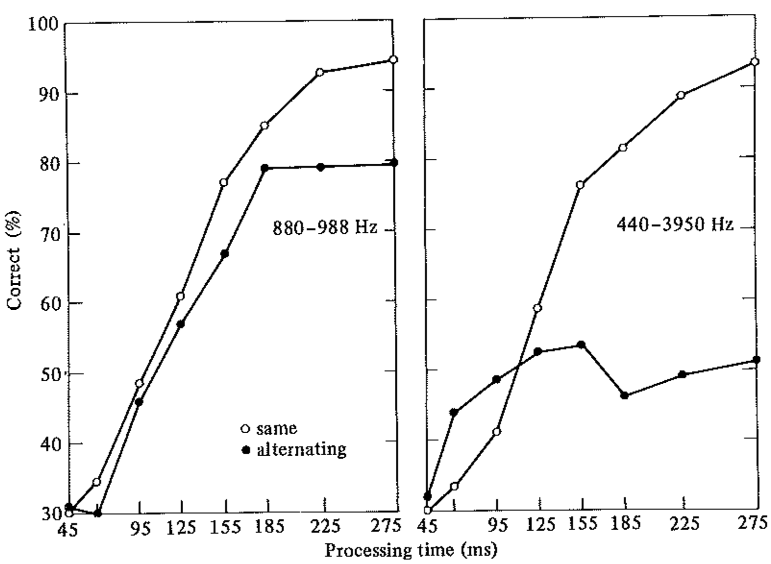

Figure 6. Percentage of correct identifications as a function of processing time for each tone for the sequences presented at one frequency or alternated between two frequencies. Results for the worst four observers.

significant 15% advantage when counting tones that alternated by one whole tone relative to the single-frequency case. In contrast, tones alternating by over three octaves produced a 13% decrement relative to the single-frequency case at this processing interval. The advantage of counting tones alternating by one whole tone may mean that these subjects were able to utilize the up–down pitch contour of the sequence of tones to facilitate performance when the pitch differences were not large. The up–down contour cannot be used to improve counting, however, with the large pitch differences produced by tones alternating by over three octaves.

The results for the middle four subjects are the most straightforward. Tones that alternated by one whole tone were counted as well as tones presented at the same frequency, at all rates of presentation. A significant decrement was found in counting the tones separated by over three octaves at processing times of 95 ms per tone and longer. The most amazing result is the performance decrement found at the slowest rate of presentation (275 ms). Although these subjects were able to count perfectly tones presented at the same frequency or tones alternating by a whole tone, they were incorrect on one out of every five trials, when the tones alternated by slightly over three octaves. Given the perfect performance in three of the four experimental conditions, it is difficult to attribute this decrement to some uninteresting variable such as lack of practice or motivation or a failure to follow instructions. That these variables cannot account for the decrement is also reinforced by the fact that all of the experimental conditions were equally likely to occur on each trial. If one of these variables were responsible for the decrement, there should have been a decrement in the other three conditions also, and this was not the case.

The worst four observers showed some decrement even when the tones were alternated by just one whole tone. This decrement was about one-third of that found when the tones alternated by nineteen whole tones (over three octaves). Although these subjects performed poorly at the slowest rates of presentation with alternating tones, they were close to perfect in counting tones presented at the same frequency.

4 General discussion

After the present study was completed, a series of related experiments were reported by van Noorden (1975). The perceptual experience of tone sequences was measured under a variety of conditions. Subjects listened to a sequence of tones that alternated in frequency for *fission* or for *temporal coherence*. Fission is the experience that some of the tones do not follow one another and appear to be unconnected and unordered. Temporal coherence is the opposite perception, that the tones appear to follow one another and to be connected and ordered in time. The major independent variables were the rate of presentation of the tones and the frequency difference between the alternating tones. If subjects listened for fission, it was heard at a smaller frequency separation than if they listened for temporal coherence. At ten tones per second, subjects listening for fission could hear fission when the tones alternated by three semitones whereas subjects listening for temporal coherence did not report the loss of temporal coherence until the tone alternated by at least three tones. Temporal coherence could be heard at larger frequency separations if the rate of presentation of the tones was decreased. Tones alternating by six semitones with 100 ms of processing time were perceived to have the same degree of temporal coherence as tones alternating by twelve semitones with 150 ms of processing time per tone.

It is difficult to compare the perceptual-experience measure used by van Noorden to the accuracy measure used in the present experiments. One cannot expect the perceptual-experience measure to predict the accuracy measure since, as van Noorden has shown, the measure is dependent on whether the subject listens for fission or for temporal coherence. We also might expect that highly practiced introspective observers could report differences in perceptual experience with stimulus changes that would not produce differences in an accuracy task. In agreement with this idea, van Noorden finds differences in perceived temporal coherence as the alternating tones are increased in frequency separation beyond seven tones, whereas we find no differences in accuracy of counting tones alternating by seven or by nineteen tones. A second difference between the studies can be seen in the processing of tones presented at roughly ten per second. Some subjects were able to count tones alternating in frequency just as accurately as tones presented at the same frequency even when the alternating tones were separated by frequencies that produced perceptual reports of fission in van Noorden's study. Given the differences between these observations, it appears that both perceptual-experience and accuracy measures are necessary dependent measures for the study of frequency separation on figure–ground organization.

In the introduction, it was proposed that integrating a sequence of alternating-frequency or single-frequency tones might be described in terms of the Gestalt laws of figure–ground organization. Although the Gestalt laws were formulated for visual scenes, it is possible to translate the laws for auditory melodies. There are a number of possible analogies between auditory and visual information processing; the goal is to find dimensions that operate similarly in the two modalities. We consider the logic and descriptive value of four such analogies in the following discussion. A visual scene is primarily organized along a spatial dimension whereas an auditory melody occurs along a temporal dimension. Accordingly, increasing the time between successive tones in a melody might be analogous to increasing the distance between the objects in a visual display. However, increasing the time between the tones appeared to have an effect on the processing of the display, rather than on the structure of the display itself. The additional time between tones was utilized to detect, integrate, and count the notes. If we increased the distance between objects in a visual display, performance might deteriorate rather than improve since additional spatial areas would have to be processed during the single look at the display. If this is the case, the analogy between the spatial dimension in vision and the temporal dimension in audition is inappropriate for the counting task.

A second possible analogy would be to equate time and frequency with the horizontal and vertical dimensions in two-dimensional space, as in the spatial representation of the musical scale. In this case, it is conceivable that it would be more difficult to count tones that alternated by over an octave relative to counting tones within the same octave. One might expect, however, that it would be more difficult to count tones alternating by over three octaves relative to counting tones alternating by just over an octave. In experiment 1, performance did not differ in these two conditions, so that the spatial representation of both the time dimension and the frequency dimension may not be appropriate.

The third analogy between the auditory and visual cases would be to equate the auditory-frequency dimension with the wavelength of light. In this case, subjects would count patches of color of the same or alternating hues. The patches could be presented simultaneously in a linear array with varying distances between patches. In this case, spatial extent could not be used as a cue for the number of patches. The time between successive tones in a melody could be represented by the total exposure time of the visual display. Analogous to the auditory case, counting performance should improve with increases in exposure (processing) time and there should be a decrement in counting patches that alternated in hue relative to those presented at the same hue. Another possibility would be to present the patches successively at the same location in space, and to vary the time between presentations.

The fourth analogy would be to equate frequency proximity of a sequence of tones with spatial proximity of a sequence of visual forms. For example, Bregman and Achim (1973) devised a visual analog of Bregman and Campbell's auditory-stream segregation. A dot was moved in discrete jumps across positions on a vertical axis. The dot was made to alternate between the positions on two regular trajectories. The regular trajectories were defined to enhance grouping by proximity and by good continuity. If the dots were presented at relatively slow speeds, the subjects saw just one dot moving in irregular motion. With faster rates of presentation, however, two dots were seen to move in separate and regular trajectories. In this case, spatial proximity has an effect similar to proximity in tone frequency, and the rate of presentation has an equivalent role in both modalities. Bregman and Achim were successful in demonstrating similar rules of perceptual grouping in the visual and auditory modalities. Finding analogous results in the visual and auditory modalities would argue for formulating relatively abstract laws of perceptual organization rather than looking for unique properties of the visual or auditory sensory system.

Acknowledgements. This research was supported by Public Health Service Grant MH-19399. James Bryant and Michael Olson assisted in testing the subjects. Wendy L Idson provided valuable discussion and comments.

References

Bregman A S, Achim A, 1973 "Visual stream segregation" *Perception and Psychophysics* **13** 451-454
Bregman A S, Campbell J, 1971 "Primary auditory stream segregation" *Journal of Experimental Psychology* **89** 244-249
Broadbent D E, 1958 *Perception and Communication* (New York: Pergamon Press)
Cherry E C, 1953 "Some experiments on the recognition of speech, with one and with two ears" *Journal of the Acoustical Society of America* **25** 975-979
Dowling W J, 1973 "The perception of interleaved melodies" *Cognitive Psychology* **5** 322-337
Garner W R, 1951 "The accuracy of counting repeated short tones" *Journal of Experimental Psychology* **41** 310-316
Guzy L T, Axelrod S, 1972 "Interaural attention shifting as response" *Journal of Experimental Psychology* **95** 290-294
Harvey L T, Treisman A M, 1973 "Switching attention between the ears to monitor tones" *Perception and Psychophysics* **14** 51-59
Idson W L, Massaro D W, 1976 "Cross-octave masking of single tones and musical sequences: the effects of structure on auditory recognition" *Perception and Psychophysics* **19** 155-175
Koffka K, 1935 *Principles of Gestalt Psychology* (New York: Harcourt Brace)
Kohler W, 1929 *Gestalt Psychology* (New York: Liveright)
Massaro D W, 1970 "Preperceptual auditory images" *Journal of Experimental Psychology* **85** 411-417
Massaro D W, 1972 "Preperceptual images, processing time, and perceptual units in auditory perception" *Psychological Review* **79** 124-145
Massaro D W, 1975 *Experimental Psychology and Information Processing* (Chicago: Rand-McNally)
Massaro D W, 1976 "Perceiving and counting sounds" *Journal of Experimental Psychology: Human Perception and Performance* **2** 337-346
Moray N, 1960 "Broadbent's filter theory: postulate H and the problem of switching time" *Quarterly Journal of Experimental Psychology* **12** 214-220
Noorden L P A S van, 1975 *Temporal Coherence in the Perception of Tone Sequences* dissertation, Technical School of Eindhoven, Eindhoven, Netherlands
Parkinson S R, Knight M V, DeMaio J C, Connors J F, 1974 "Channels and order of report in dichotic memory" *Perception and Psychophysics* **16** 235-241
Savin H, 1967 "On the successive perception of simultaneous stimuli" *Perception and Psychophysics* **2** 479-482
Taubman R E, 1950 "Studies in judged number: I. The judgment of auditory number" *Journal of General Psychology* **43** 167-194
Treisman A M, 1964 "Verbal cues, language, and meaning in selective attention" *American Journal of Psychology* **77** 206-219
Treisman A M, 1970 "Perception and recall of simultaneous speech stimuli" *Acta Psychologica* **33** 132-148
Treisman A M, 1971 "Shifting attention between the ears" *Quarterly Journal of Experimental Psychology* **23** 157-167