

Cross-octave masking of single tones and musical sequences: The effects of structure on auditory recognition

WENDY L. IDSON and DOMINIC W. MASSARO
University of Wisconsin, Madison, Wisconsin 53706

A series of experiments explored the role of structural information in the auditory recognition process, within the context of a backward recognition masking paradigm. A masking tone presented after a test tone has been found to interfere with the perceptual processing of the test tone, the degree of interference decreasing with increased durations of the silent intertone interval between the test and masking tones. In the current studies, the task was modified to utilize three-tone sequences as the test stimuli. Six test sequences were employed (LMH, LHM, MLH, MHL, HLM, HML), where L, M, and H represent the lowest, middle, and highest frequencies in the melody. The observers identified these six possible sequences when the three tones of the test sequence were interleaved with three presentations of a single masking tone. All three tones of the test sequence were drawn from the same octave, while the masking tones could be drawn from any of three octaves, symmetrical around the octave containing the test tones. Under these conditions, interference occurred primarily from masking tones drawn from the same octave as the test tones. Masking tones drawn from other octaves were found to produce little, if any, interference with perception of the test tones. This effect was found to occur only for the identification of tonal sequences. Substantial masking of single-tone targets occurred with masking tones drawn from octaves other than that containing the targets. The results make apparent the use of structural information during auditory recognition. A theoretical interpretation was advanced which suggests that, while single tones are perceived on the basis of absolute pitch, the presence of auditory structure may allow relational information, such as exact pitch intervals or melodic contour, to facilitate perception of the tonal sequence.

In recent years, a good deal of attention has been devoted to the study of human information processing capacities. From a variety of perspectives, experimental psychologists have attempted to delineate the structures and processes involved in extracting information from a stimulus. Yet, within this perspective, relatively little effort has been devoted to exploring the nature of the information contained within the stimulus itself (however, see Garner, 1970, 1974). This is surprising, in that the attributes of a stimulus may be a critical variable in determining the manner in which that stimulus is processed (Gibson, 1950, 1966). In this context, it seems worth exploring the possible effects that the stimulus may have upon the processing system, as well as the operations that the system performs upon the stimulus.

One phenomenon in which the properties of the stimulus seem to be particularly striking is the apparent independence of tones in different frequency channels, variously called primary auditory stream segregation (Bregman & Campbell, 1971), rhythmic fission (Dowling, 1968), or a trill threshold (Miller &

Heise, 1950). Essentially, the various terms all refer to the fact that, at rapid presentation rates, a sequence of tones drawn from sufficiently different frequency ranges appears to split into separate channels, with no apparent correlations between them. The phenomenon has been used extensively by composers, such as J. S. Bach, to make a single instrument playing interleaved melodies in different ranges appear as two independent melodies (Ortmann, 1926).

Miller and Heise (1950) subjected the phenomenon to an empirical analysis. They alternated two 100-msec sine-wave tones at a rate of 10/sec, gradually increasing the frequency difference between the two tones. Using tones over a frequency range of approximately 150 to 7,000 Hz, they found what they call a "trill threshold." At a frequency difference of at least 3 semitones, the tones come to sound like two independent and simultaneous tones, rather than two continually alternating tones. In a further investigation, Heise and Miller (1951) asked subjects to adjust the frequency of a target tone within an auditory pattern to the point at which the target first sounded disjointed from the rest of the pattern. Within the context of this task, it appeared as if the frequency at which a tone split from the sequence as a whole was a function of the nature of the auditory pattern, as well as the absolute frequency differences. That is, the frequency at which the target tone appeared to split from the pattern was determined not

This research was supported by Public Health Service Grant MH-19399. The authors would like to thank Michael Cohen, David Klitzke, Lola Lopes, and Willard Thurlow for helpful discussions. Requests for reprints should be sent to Wendy L. Idson, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

only by the directly adjacent tones but by those tones preceding or following it at a distance.

Though the Miller and Heise (1950; Heise & Miller, 1951) findings were provocative, little work was done concerning the question for a number of years. Renewed interest derived from its implications for the perception of temporal order. Norman (1967) investigated the problem from this perspective. He presented subjects with two tones of disparate frequencies, alternating in a 10-tone sequence. Each tone had a 100-msec duration, and tones were separated by 30-msec blank interstimulus intervals. Though the exact frequencies of these tones were not reported, Norman does indicate that they fell within the Miller and Heise (1950) "trill threshold"—presumably within a 3-semitone separation. A probe tone of 30 msec was inserted between two of the background tones in the middle of the sequence. The probe tone had a frequency either between that of the background tones or much higher or lower than these tones. The principal finding was that when the frequency of the probe fell between those of the background tones, subjects were quite accurate in judging whether the probe had followed a high or a low tone. However, when the frequency of the probe was either much higher or much lower than the background tones, the subjects could not report its position in the sequence. As an additional finding, Norman (1967) observed that the effect did not occur when a single pair of background tones was presented. Subjects, in this case, could accurately order the two tones.

Bregman and Campbell (1971) attempted to organize other findings on the perception of temporal order into the framework of frequency splitting. Earlier work (Warren, Obusek, Farmer, & Warren, 1969) had indicated that subjects required much longer stimulus durations to judge the order of unrelated sounds than they did for similar sequences of related sounds. Warren et al. (1969) found that unpracticed subjects were unable to identify the order of a hiss, buzz, high tone, and low tone presented in a repeated sequence until the durations of each stimulus were extended to 700 msec. Though subsequent work employing practiced subjects obtained accurate identifications with durations as short as 200 msec, the finding was still surprising in that both speech sounds and tones within melodies—which can have durations as short as 70 and 50 msec, respectively—are clearly perceived in their correct temporal relations (Warren, 1974).

Bregman and Campbell (1971) felt that Warren's work was closely related to that of Miller and Heise (1950) and Norman (1967). They renamed the effect "auditory stream segregation," and defined it as the splitting of co-occurring auditory events into separate channels defined by the perceptual relationship among stimuli within a channel. They reasoned that

the repetition of a sequence of unrelated sounds would organize itself into separate channels, each composed of repetitions of common elements. Judging order of stimuli within the actual stimulus group, but across the perceptual channels, would therefore be quite difficult, as Warren et al. reported.

In an initial test of the phenomenon, Bregman and Campbell (1971) presented subjects with sequences of six 100-msec sine-wave tones, alternating between two different frequency ranges (2,500, 2,000, 1,600 Hz and 550, 430, 350 Hz). The subject's task was to report the order in which the tones occurred, after listening to the sequence until they felt confident of their judgments. The primary result was that subjects were highly accurate in reporting the order of the high tones relative to each other and of the low tones relative to other low tones, but were unable to relate the tones across frequency groups to report the actual sequence order. Moreover, subjects tended to group the tones, for a written report of order, by frequency group rather than by their actual temporal order.

These findings were extended in a second study (Bregman & Campbell, 1971) using a short-term memory recognition task. The subject heard 5 sec of repetition of a standard sequence of three 100-msec tones, separated by 100-msec blank intervals. Immediately following the third tone of the standard, a comparison sequence composed of six 100-msec tones was presented, the additional tones being inserted into the blank intervals of the standard sequence. The comparison sequence was also repeated for 5 sec. The tones used were drawn from the same frequency range as those employed in the first experiment. The principal independent variable was whether the tones comprising the standard sequence were all drawn from the same stream or were drawn from different streams. The subjects' task was to judge whether the order of the standard tones contained in the comparison sequence was the same or different from the order of these tones in the standard sequence. They found high accuracy for subjects' performance of same/different judgments when all of the standard tones were from the same stream and poor judgments when the standard tones were drawn from the two different streams, suggesting again an inability to report order across frequency channels. It should be noted that by filling the comparison sequence with the tones not used in the standard, Bregman and Campbell (1971) confounded the structure of the standard sequence with the structure of the comparison sequence. When a within-stream standard was used, the added tones in the comparison were drawn from the other stream, while in the cross-stream standard, the added tones were drawn from both streams. It is difficult to distinguish whether the inferior performance in the cross-stream condition was due to poorer perception

of order information in the standard or to greater interference from the added tones in the comparison sequence.

The essential Bregman and Campbell (1971) finding has recently been replicated in a somewhat different paradigm. Fitzgibbons, Pollatsek, and Thomas (1974) employed sequences of tones composed of two high (2,093 and 2,394 Hz) and two low (440 and 494 Hz) sine-wave tones. They inserted one of three silent gaps (20, 40, or 80 msec) into one of three possible locations in some of the sequences (following the first, second, or third of the four tones). On each trial, the subject heard two repetitions of one of the sequences, and his task was to determine whether the twice-presented sequence had a gap in some prespecified location. The primary finding was that subjects were highly accurate in determining the location of a gap which occurred within either of the frequency groups (high or low), but were extremely poor at detecting a gap between the two frequency groups. If subjects were unable to keep track of the temporal order of tones which alternated between disparate frequency ranges, they would also find it difficult to notice a silent gap between them.

Bregman and Campbell (1971) have argued that stream segregation results from a process analogous to Neisser's (1967) formulation of preattentive mechanisms. Frequency is seen as one of many possible dimensions along which the stimulus field could be organized, rather than as an absolute structural channel. According to this line of reasoning, perceptual splitting could be obtained with variables other than frequency as the organizational dimension. Similarly, if an organizational framework stronger than frequency channels is imposed on stimuli, which would otherwise group into independent ranges, then perceptual splitting should not occur as a result of the influence of the dominant organization.

Two classes of experiments have been performed to test the hypothesis that perceptual splitting can be prevented by imposing a higher level of organization upon the stimuli. First, Bregman and Dannenbring (1973) introduced frequency transitions between alternating high and low tones in an attempt to impose a higher order structure which would prevent stream segregation. In both a standard-comparison task similar to that employed by Bregman and Campbell (1971) and in direct judgments by observers of the number of streams present in a sequence, transitions between tones decreased the amount of perceptual splitting obtained. In the second class of studies, stream segregation has been investigated using speech stimuli. The motivation for this research was the observation that, unlike other auditory sequences, speech does not segregate into independent streams based upon such acoustic characteristics as frequency. The chief hypothesis that

has been advanced to explain the absence of segregation is that transitional cues prevent perceptual splitting by preserving the temporal order of the individual acoustic segments. Studies directed towards this question have taken two tacks: removing the frequency transitions which normally accompany successive phonemes in order to induce segregation or, inversely, inserting transitions to eliminate segregation. Both manipulations have demonstrated the crucial role played by transitions. When transitions are eliminated, accurate perception of the order of speech segments becomes quite difficult (Cole & Scott, 1973) and perceptual splitting occurs (Dorman, Cutting, & Raphael, 1975). When transitions are reintroduced, the segregation effect is eliminated (Dorman et al., 1975). Taken with the Bregman and Dannenbring (1973) finding, these results suggest that perceptual splitting on the basis of frequency ranges can be overcome by the presence of a higher level of organization, though it is not clear that this finding in turn supports the preattentive-mechanism formulation.

The stream segregation effect has also been demonstrated using musical sequences. Dowling (1973) attempted to investigate experimentally—yet more directly—the melodic fission phenomenon in music. He presented subjects with two familiar melodies, interleaved in time. A method of limits design was employed, the two melodies starting with a maximum overlap between frequency ranges. The two melodies were made progressively more distinct, one of the melodies being transposed up by a semitone on every fourth presentation. The subject's task was to name one of the melodies as soon as possible, and then to name the second melody whenever he could. Subjects were unable to identify the melodies until one had been sufficiently transposed that their frequency ranges were no longer overlapping. Though the amount of separation necessary for identification varied somewhat with the individual melodies employed, the median degree of separation needed was 1.2 semitones. In a second experiment, Dowling (1973) used a short-term memory recognition task. A standard melody was compared to a comparison melody, which was interleaved with a second melody. The subject's task was to determine whether the comparison melody was the same as or different from the standard melody. Subjects performed much better in the task when the comparison and background melodies were played in different frequency ranges than when the ranges overlapped. In additional studies, Dowling (1973) found that while knowledge of the identity of the target melody aided identification, knowledge of the identity of the background melody did not. On trials on which the target melody was not presented, no erroneous identifications occurred. This argues that knowledge of the identity of the target melody facilitated the

recognition of that melody, rather than simply altering the subjects' decision criterion.

Dowling (1973), in line with Norman (1967), offers a somewhat different theoretical interpretation of the stream segregation effect. In contrast with the preattentive mechanisms approach suggested by Bregman and his associates, Dowling and Norman have argued that the effect is an attentional phenomenon. Tones within different frequency ranges constitute different perceptual channels, analogous to the sensory channels postulated by Broadbent (1958). Though the frequency range that constitutes a channel has never been precisely defined in the literature, Norman (1967) hypothesized that it might correspond to the critical bandwidth. The critical band can be defined as the frequency range around a given tone, such that tones falling outside of this range are independent of the tone while tones falling within the range are not. A variety of measures—such as loudness constancy, musical consonance, and masking thresholds—have been found to vary with the critical band (Scharf, 1971). In line with this, Norman (1967) has suggested that within the critical band, tones fall within the same frequency range or stream while outside of the critical band they do not. No other specific suggestions as to the critical range have been made. Studies concerned with the effect have simply used large separations, generally well over an octave, though Dowling (1973), Miller and Heise (1950), and Van Noorden (Note 1) have reported the effect at the smaller separation of 1 to 3 semitones.

Whatever the frequency range of the channels may be, it has been suggested that a limited-capacity processor can attend to only one such channel at a time. When tones falling within different channels are presented alternately, at rapid rates, attention must be shifted repeatedly between channels. If the rate of alternation between the tones is more rapid than that at which the attentional mechanism can follow, breakdowns will occur in accurate perception (Dowling, 1973). Within this context, the effects associated with stream segregation correspond closely to those which occur with rapid alternation of items between the two ears (cf. Axelrod & Guzy, 1968; Massaro, 1975; Treisman, 1971; Massaro & Idson, Note 2).

Though the attentional hypothesis can handle much of the data presented thus far, it rests upon the assumption that order errors result from time limitations on the limited-capacity processor. A study by Deutsch (1972b) casts some doubt upon this assumption. Deutsch presented subjects with the tune "Yankee Doodle," in which the notes were dispersed randomly throughout three octaves. The subjects were unable to recognize the melody under these conditions, even though it was perfectly clear when played entirely in any one of the three octaves. The presentation rate in this study was either 3 tones/sec

or 1 tone/sec, dependent upon the tempo of the melody. At rates of 1 tone/sec, an attentional mechanism would surely be able to follow the tune across octaves if a time switching limitation were the predominant factor at work. Deutsch argues that her results derive from the operation of a separate analytic mechanism for the abstraction of successive intervals in a melody (cf. Deutsch, 1969), not from stream segregation. However, her argument rests primarily upon the differences between her presentation rates and those which were used in the studies cited earlier. As the lower limit on presentation rate for obtaining the segregation effect has not been reliably established, this reasoning seems somewhat circular, and the data can be handled equally by a melodic fission explanation.

As can be seen, there is, as yet, too little evidence upon which to base an explanation of the phenomenon. Moreover, despite the clear theoretical distinction between a preattentive (Bregman & Campbell, 1971) and an attentive (Dowling, 1973; Norman, 1967) explanation of the stream segregation effect, differences concerning the empirical consequences of the two predictions are somewhat less clear. For example, Bregman and his associates argue that one strong prediction of the preattentive mechanisms approach would be that it should be possible to obtain segregation based upon stimulus dimensions other than frequency. Yet Dowling (1973), arguing from a different orientation, has also suggested that the effect can be obtained on the basis of various types of stimulus information.

It is apparent that some basic information is required before more advanced exploration or interpretation can be undertaken. The only factor that has been concretely established is that at rapid presentation rates a sequence of tones—separated widely in frequency—will appear to split into two independent channels. However, neither the degree of separation needed to elicit the effect nor the necessary frequency ranges have been reliably determined. Similarly, little has been done with what is probably a critical variable, the structure of the stimulus sequence. The Norman (1967) finding that the effect cannot be obtained with a single tone pair suggests the importance of stimulus structure in this context.

The present study seeks both to elicit such information and to use the stream segregation phenomenon to examine the role of structure in the auditory recognition process. Previous research on auditory recognition has employed a backward recognition masking paradigm (cf. Massaro, 1975). In this task, the observer is to identify a target tone followed, after a variable silent interval, by a masking tone. The typical finding in such studies (e.g., Massaro, 1970) is that the presence of the masking tone interferes with perception of the target, within a critical interval of separation between two tones.

Recognition of the target has been found to improve monotonically with increased durations of the silent intertone interval between target and mask, out to an asymptotic value of approximately 250 msec.

On the basis of such data, the argument has been made that perception of a stimulus takes time (Massaro, 1972a). Information is extracted continuously from a tone over the course of approximately 250 msec. If a second or masking tone is presented before this process is completed, the information needed to perceive the tone will be disrupted and identification of the target will be based upon that information which has been extracted prior to onset of the mask. The negatively accelerated monotonic function typically obtained in backward recognition masking can be seen as representing the extraction of successively greater amounts of information from the target over the time available for processing. Backward masking effects have been found to be modality-limited. When the target and mask are drawn from separate modalities—such as a tone and a light—no interference with recognition of the target occurs from the mask (Massaro & Kahn, 1973). In the current study, it was hypothesized that if the stream segregation effect represents the grouping of stimuli into separate perceptual channels, then these channels should operate in a manner analogous to separate modalities in a backward-masking task. That is, a mask drawn from a different octave than the target should not interfere with perception of the target tone. Such a finding—constituting the elimination of a well-substantiated effect—would argue strongly for an important role of stimulus information on the nature of the auditory recognition process. Consequently, a backward recognition masking task was employed, in which the mask could fall into the same octave as the target, the octave higher, or the octave lower. In light of the Norman (1967) finding that the stream segregation effect does not occur with single tone pairs, the task was modified and a sequence of test tones interleaved with repeated presentations of a masking tone.

EXPERIMENT I

Method

Subjects. The subjects were 10 University of Wisconsin undergraduates, 6 of whom were fulfilling a course requirement and 4 of whom received \$1.50/h for their services. One subject was dropped from the analysis for failing to respond on over 25% of the trials. On the average, the other subjects missed responding on under 3% of the trials.

Apparatus and Stimuli. Four subjects could be tested simultaneously in separate sound-insulated rooms. All experimental events were controlled by a PDP-8/L computer. The tonal stimuli were generated as sine waves by a digitally controlled oscillator (Wavetek Model 155) and were presented binaurally over matched headphones (Grason-Stadler Model TDH-49).

The three test tones were chosen so as to be 3 semitones apart on an equal-tempered scale (international pitch: A = 435 Hz); they corresponded to the notes A₄ (435 Hz), C₅ (517 Hz), and D#₅

(615 Hz). Six masking stimuli were drawn from the same scale, two falling within the same octave as the test tones, two falling within the octave higher, and two falling within the octave lower. The masks within the same octave were selected so as to be adjacent to C and separated from both A and D# by a single semitone, corresponding to B₄ (488 Hz) and C#₅ (548 Hz). The higher octave masks were displaced in pitch exactly one octave higher than the within-octave masks, B₅ = 976 Hz and C#₆ = 1,096 Hz. Similarly, the lower octave masks were displaced in pitch exactly one octave lower than the within-octave masks: B₃ = 244 Hz and C#₄ = 274 Hz. All tones had a duration of 100 msec. The test tones and within-octave masks were presented at 80 dB SPL, the lower octave tones were presented at 82.5 dB SPL, and the higher octave tones at 78 dB SPL, to equate for subjective loudness.

Procedure. The experiment was conducted on 5 consecutive days. Each day was divided into two 25-min sessions. On the first practice day, Session 1 was further subdivided into a series of absolute learning trials. The task on the absolute learning trials was to identify the pitch of the test tones presented in isolation, by associating the labels A, B, and C to the notes A, C, and D#, respectively. Observers received 110 absolute learning trials, the first 10 being unscored practice trials. On each trial, a single tone was presented. Presentation of the three tones was random and each tone was programmed to occur with equal probability within a session. During a 1.5-sec response interval, the observer pressed one of three buttons labeled A, B, and C to indicate his response. Feedback was provided as to the correct tone by a 500-msec presentation of the letter associated with the tone on a visual display of light-emitting diodes (Monsanto Model MDA-III). Trials were separated by 1-sec blank intertrial intervals.

The sequence learning trials were provided to enable the subjects to learn the sequence of tones which would appear in the test trials. Observers received 220 such trials, the first 20 being unscored practice trials. The subjects were not aware that any of the trials were considered practice trials. On each trial, the subject heard a sequence of the three test tones, separated by 500-msec intertone intervals. The sequences were constructed from all possible permutations of the three test tones, yielding six test sequences: ABC, ACB, BAC, BCA, CAB, CBA. The sequences were presented in random order and were programmed to occur equally often within a session. Following presentation of the sequence, subjects were given 2 sec in which to respond by pressing one of six buttons labeled with the sequence names (i.e., ABC). Feedback was then provided by a 1-sec visual display of the sequence name, listing the labels for the tones in left-to-right order, corresponding to their order in the sequence.

Session 2 of Day 1 and both sessions on all subsequent experimental days consisted of 250 test trials, the first 30 being unscored practice trials. Again, the subjects did not know that any of the trials were practice trials. On each trial, one of the six possible test sequences was presented, with a 350-msec interval between the offset of one test tone and the onset of a following test tone. On one-quarter of the trials, this interval between successive test tones was left blank (no-mask condition). On the remaining trials, one of the six possible masking tones was inserted in the interval following each test tone. The test tone and its following mask were separated by a variable blank intertone interval of 40, 80, 120, or 160 msec. The interval between successive test tones was constant over changes in the intertone interval between each test tone and its following mask. Both the masking-tone frequency and the intertone interval were constant within a trial but varied between trials. The two masks within each octave were treated as one level of the masking variable, giving four levels: lower octave, higher octave, within octave, and no mask. All 96 experimental conditions (6 sequences by 4 intertone intervals by 4 masking conditions) were completely random and were programmed to occur with equal probability within a given session. Following presentation of the sequence, each subject had 3 sec in which to respond by pressing one of six buttons with the correct sequence label. Feedback was provided for 1 sec in the same manner as for the sequence learning trials. The intertrial interval was 1 sec.

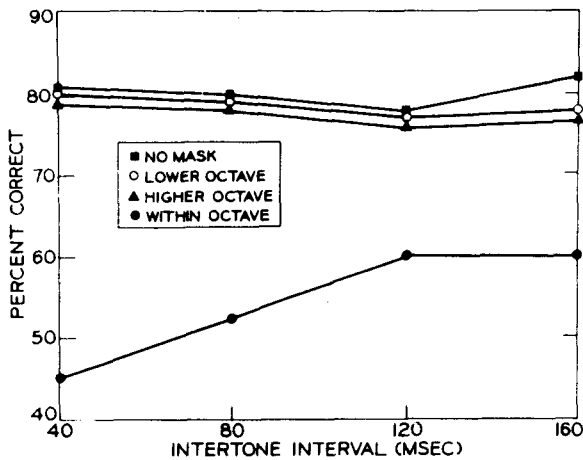


Figure 1. Percentage of correct identifications of the test sequence as a function of the masking condition and the duration of the silent intertone interval between the test and masking tones, in Experiment I. Changes in the no-mask condition across the intertone interval is a dummy variable since no masking tones were presented.

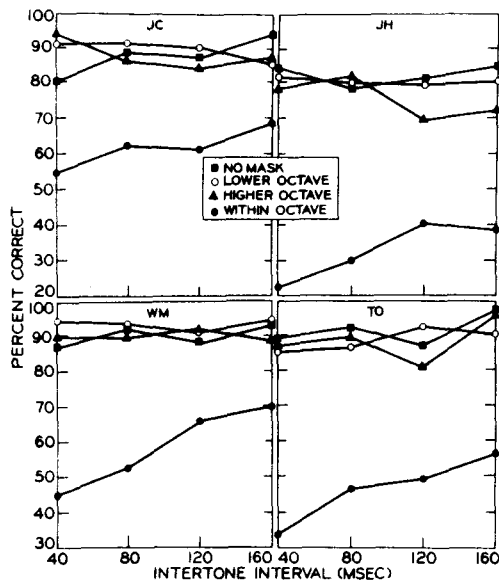


Figure 2. Percentage of correct identifications of the test sequence as a function of the masking condition and the duration of the silent intertone interval, for subjects J.C., J.H., W.M., and T.O., in Experiment I.

Results

The dependent measure is the percentage of correct responses. First, an analysis of variance was carried out with days (2-5), subjects, sequences, masking conditions, and intertone intervals as factors. The main effect of days and all interactions involving days as a term were nonsignificant in the analysis. Consequently, to increase the reliability of the individual subject scores, the response frequencies were pooled over Days 2-5 and reanalyzed.

Figure 1 presents the percentage of correct identifications of the test sequence as a function of both the silent intertone interval and the frequency of the masking tone. The figure shows that identification performance was substantially worse in the within-octave masking condition than in any of the other conditions. Performance with the within-octave mask asymptoted at 60%, as compared with an average accuracy of 80% for the other three conditions. It can also be seen that performance did not differ between the higher and lower octave masking conditions. Moreover, subjects were equally accurate in identifying the test sequence with masks outside the octave containing the test tones as they were in the no-mask condition. An additional result, which is apparent in Figure 1, is that performance improved with increased durations of the intertone interval only in the case of the within-octave mask. Performance was essentially constant across all intervals for both masks outside the octave. The no-mask condition is plotted as a function of the intertone interval in Figure 1. This condition constitutes a dummy variable in which a constant interval of 350 msec was left blank between successive test tones. The results in the no-mask condition simply represent four sets of independent observations, collected under the same no-mask condition, in four separate cells corresponding to the intertone intervals of the three masking conditions. The no-mask condition also entered the analysis as a level of the masking variable.

An analysis of variance on subjects, sequences, masking conditions, and intertone intervals confirmed these conclusions. The analysis revealed significant effects for the type of mask, $F(3,24) = 25.09$, $p < .001$, the intertone interval, $F(3,24) = 4.01$, $p < .025$, and the interaction of these two variables, $F(9,72) = 10.78$, $p < .001$.

Though there were considerable individual differences in overall level of performance, the same effects were present for each individual subject. The results for the four subjects (J.C., J.H., W.M., and T.O.) agreeing with the group data most strongly are presented in Figure 2.

Examination of the data on the six test sequences indicates that the effects described above were present for each of the sequences. Differences in identifying the test sequences appear to be largely the result of the greater ease with which subjects could identify sequences ABC and CBA—in which there was a consistent rise or fall in frequency. Performance on sequences BAC, BCA, CAB, and ACB differed from one another by at most 2%, while they differed from CBA by at least 6% and from ABC by at least 12%. As can be seen in Figure 3, this difference in performance on individual sequences occurred only in the no-mask and in the outside-of-the-octave masking conditions; no advantage was found for sequences

ABC and CBA under the within-octave masking condition. Supporting these conclusions, the analysis of variance revealed significant effects for the type of sequence, $F(5,40) = 4.635$, $p < .005$, the Sequence by Intertone Interval interaction, $F(15,120) = 2.782$, $p < .005$, the Sequence by Masking Condition interaction, $F(15,120) = 4.474$, $p < .001$, and the Sequence by Intertone Interval by Masking Condition interaction, $F(45,360) = 1.137$, $p < .001$.

Discussion

The results indicate that processing of the test sequence was disrupted by the presence of the within-octave masking tones. Performance in this condition was considerably worse than that in the no-mask condition—in which the interval between test tones was left blank. Accuracy with the within-octave mask improved with increases in the silent intertone interval, asymptoting at 120 msec. It has been demonstrated (Massaro, 1972b) that the rate of perceptual processing of an auditory stimulus during the silent intertone interval is equivalent to the rate of processing during presentation of the test tone itself. Thus, with a 120-msec intertone interval, subjects had 220 msec of effective time in which to process each individual tone. These findings are in close agreement with those cited earlier (Massaro, 1970, 1972a), indicating that perceptual processing continues after stimulus offset, with total processing time requiring roughly 250 msec.

However, strikingly different results were obtained with both the higher and lower octave masks. In both cases, level of performance was unaffected by the intertone interval. Performance was essentially constant across intertone intervals, subjects being highly accurate even with intervals as short as 40 msec. Moreover, performance was equally as accurate for masks outside of the octave containing the test tones as it was for the no-mask condition. Subjects were apparently able to process the test tone completely despite the presence of the masking tone, indicating that the mask failed to interfere with perception of the test tone. It can be concluded from this that, in some sense, tones drawn from different octaves are independent—falling into distinct channels at a stage of processing as early as perception.

Having demonstrated the power of the stream segregation effect within the backward-recognition masking paradigm, the question must now be raised as to what mechanism can be said to be responsible for the effect. Two different theoretical frameworks can be offered. The first would suggest that the stream segregation phenomenon demonstrates an as-yet-undefined role of stimulus information during auditory recognition. This implies that, in the Gestalt sense, the whole of the sequence is more than the sum of its component tones. That is, in analyzing a

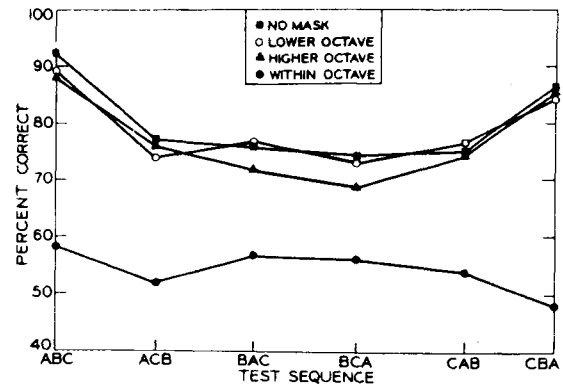


Figure 3. Percentage of correct identifications of each test sequence as a function of the masking condition, in Experiment I.

sequence, the recognition process proceeds in a manner different from that involved in processing the individual tones. This explanation is appealing in light of the complete elimination of interference from the outside-of-the-octave masks—in direct opposition to a substantial body of literature on auditory recognition (Massaro, 1975).

However, a stimulus-oriented explanation of the effect lies outside the class of theoretical models generally proposed in this area. As such, it is worth considering the fit between the current data and the prominent attentional explanations. Though Dowling (1973) and Norman (1967) do not offer explicit process analyses of the stream segregation effect, their theoretical arguments map nicely into a Broadbentian model which assumes that distinct frequency ranges correspond to independent perceptual channels, with no cross-talk occurring between channels. If a mask drawn from a different octave than a test tone falls into a different perceptual channel, then the masks would not be processed until after all three test tones had been processed. As a result, the masks drawn from different octaves than the test tones would not be expected to interfere with the processing of the test tones as in backward masking experiments. Both of these results were, in fact, obtained. The outside of the octave masks produced curves which were virtually identical to that for the no-mask condition, with performance at 80% correct over all four interstimulus intervals. The within-octave masks, however, produced a strong masking function, performance increasing monotonically across intervals.

Despite the close correspondence between the results obtained and the major predictions derived from a Broadbent model, there is one aspect of the data which may run counter to a somewhat less obvious prediction of the model. While it has been argued that tones of sufficiently different frequencies fall into different perceptual channels, the nature of a channel has not been explicitly defined. The central question here concerns the manner in which

information is assigned to a channel. One possible solution would suggest, with Broadbent, that belonging to a channel is an intrinsic attribute of a stimulus. That is, all tones falling within a particular frequency range would, *a priori*, belong to the same channel. Alternatively, it could be argued that the auditory structure itself defines the stimulus channels, in terms of the frequency range spanned by the tones in the sequence. Essentially, this range could define an auditory space as relevant to the task. The limited-capacity processor would then selectively attend to only those tones falling within this space. The central implication of such a proposal would be the concept that channels are synthesized over time from the incoming information. Such a view is somewhat more appealing, as it would allow the divergent estimates of the size of a frequency channel to be reconciled on the basis of differential task demands.

The primary prediction to be derived from such a theory is that as attentional effects could come into play only subsequent to channel formation, at very short interstimulus intervals no effects of attention should be found. That is, in order to synthesize a channel, the recognition process must extract a certain amount of information from a stimulus. If a masking tone is presented prior to the extraction of this information, no channel will yet have been synthesized, in which case no effects of attention would be expected (Massaro & Idson, Note 2). The critical prediction of such a model for Experiment I would be that masking should be obtained from outside of the octave masks at short interstimulus intervals, as the channels would not yet have been developed and the filter would have no basis for excluding the masking tones from processing. Contrary to this prediction, however, even at intervals as short as 40 msec the outside-of-the-octave masks did not produce a performance decrement relative to the no-mask condition.

The constant level of performance under the higher and lower octave masks across interstimulus intervals might be taken as evidence against the channel synthesis explanation of this effect. However, an alternative interpretation is possible. That is, it could be argued that the test-tone duration and the interstimulus intervals were too long to elicit the expected effect. As the test-tone durations were 100 msec at the shortest interstimulus intervals, the recognition process still had 140 msec of effective processing time. It seems reasonable to suggest that this would constitute sufficient time in which to synthesize a channel. If such were the case, the absence of an interference effect from outside of the octave masks at the shortest intervals would be expected, as the mask would be segmented from the octave channel even at this point in time.

Experiment II was designed to investigate the tenability of the concept of synthesis of independent

perceptual channels over time. It was predicted that if such a channel mechanism was in fact at work, then interference might logically result if the subject were given substantially less time in which to organize the stimulus input. Consequently, Experiment II was intended as a replication of Experiment I, utilizing shorter test-tone durations of 50 msec and shorter intertone intervals of 20, 40, 100, and 160 msec. If the results of Experiment I can be attributed to the presence of an auditory channel that develops over time, then interference from the outside of the octave masks should occur at the shortest intertone intervals, which is comparable to the interference obtained with within-octave masks, while at the longer intertone intervals the results of Experiment I should be replicated. If no interference occurs from outside of the octave masks at even the shortest intertone intervals, then it is probable that a channel-synthesis concept of attention is not an adequate explanation for the data.

EXPERIMENT II

Method

Subjects. The subjects were nine University of Wisconsin undergraduates who received course credit for their participation in the study.

Stimuli. The stimuli in Experiment II were modified from those used in the earlier studies. Pilot work indicated that, with test tones of a 50-msec duration, the frequencies employed in Experiment I made the task too difficult to produce greater than chance performance in the within-octave mask condition. Subjects performed in this condition with less than 15% accuracy—chance being 16% for the six-alternative task.

The test-tone frequencies were set farther apart to increase the discriminability of the individual tones, making the task somewhat easier. The test tones were now separated by 4 semitones; they corresponded to the notes A_4 (435 Hz), $C\#_5$ (548 Hz), and F_5 (732 Hz) and were labeled A, B, C for purposes of identification, again in order of ascending pitch. Six masking-tone values were employed—spanning a three-octave range symmetrical around the octave containing the test tones. The higher of the two masking tones within a given octave was increased in frequency to keep the masks equidistant from the test tones: within-octave masks, B_4 (488 Hz), $D\#_5$ (615 Hz); lower octave masks, B_3 (244 Hz), $D\#_4$ (307 Hz); higher octave masks, B_5 (976 Hz), $D\#_6$ (1,230 Hz). The two masks within a given octave were treated as a single level of the masking variable—giving four levels, including the no-mask condition, in which the 350-msec interval was left blank between test tones.

Procedure. The procedure for the current experiment was identical to that of Experiment I, with three exceptions. First, shorter interstimulus intervals of 20, 40, 100, and 160 msec were employed. Second, 50-msec test tones were substituted for the 100-msec test tones of Experiment I. The masking tones were maintained at 100 msec. The interstimulus interval between successive test tones remained 350 msec. Third, Session 2 of Day 1 and both sessions of the 4 subsequent days had 320 trials each—the first 20 being unscored practice trials.

Results

As in Experiment I, there was no effect of days and data were pooled in the analysis over Days 2-5. Figure 4 presents the percentage of correct identifications of the test sequence as a function of both the duration of the silent intertone interval and

the frequency of the masking tone. It can be seen that a substantial difference in performance was obtained among the masking conditions. The within-octave masks produced performance at a level of only 30%-35% correct, while both the higher and lower octave masks yielded performance at approximately 65% correct. Contrary to the results of Experiment I, the outside-of-the-octave masks did not reach a level of performance as high as that of the no-mask condition—approximately 75% correct. However, this difference appears to be due to the data of two subjects. For seven of the nine subjects, the outside-of-the-octave masks produced performance equal to the no-mask condition. Data for three subjects showing this effect are presented individually in Figure 5. For the eighth subject, the higher and lower octave masks were more difficult than the no-mask condition, though both of the outside-of-the-octave masks were in turn easier than the within-octave mask. For the ninth subject, all three masking conditions were equally difficult—producing performance at a level only slightly better than chance for the six alternative tasks, while the same subject identified the sequence correctly 56% of the time in the no-mask condition.

These conclusions are supported by analysis of variance. The analysis revealed a significant main effect for the type of mask, $F(3,34) = 20.59$, $p < .001$. Planned comparisons among the levels of the masking variable revealed a highly significant difference between the within-octave mask and the other three masking conditions, $F(1,24) = 16.66$, $p < .001$. Supporting the conclusion that the difference between the no-mask and outside-of-the-octave masking conditions was a function of data from two subjects only, a comparison of the no-mask condition against the higher and lower octave masking conditions showed the difference between these conditions to be insignificant.

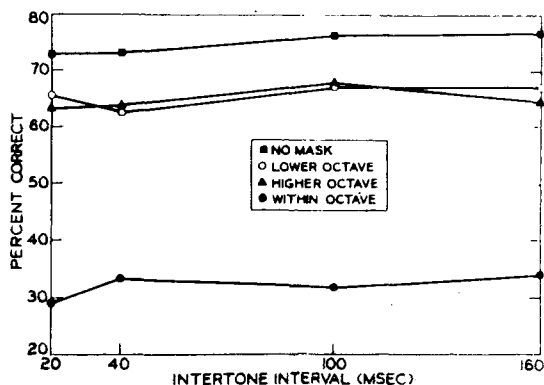


Figure 4. Percentage of correct identifications of the test sequence as a function of the masking condition and the duration of the silent intertone interval between the test and masking tones, in Experiment II. Changes in the no-mask condition across the intertone interval is a dummy variable since no masking tones were presented.

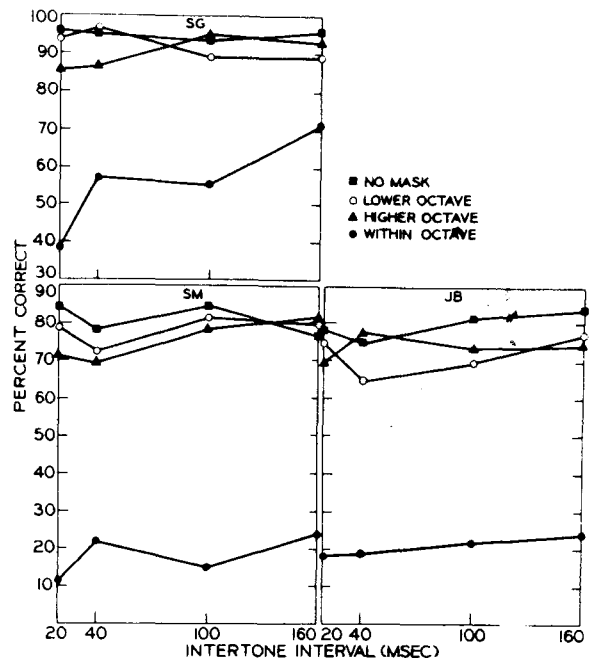


Figure 5. Percentage of correct identifications of the test sequence as a function of the masking condition and the duration of the silent intertone interval, for subjects S.G., J.B., and S.M., in Experiment II.

As can be seen in Figure 6, performance with sequences CBA and ABC—in which there is a consistent rise or fall in pitch—was superior to performance on the other four sequences by at least 20%, for the no-mask, higher octave mask, and lower octave mask conditions. For the within-octave mask, however, performance on sequences CBA and ABC was comparable to that found for the other four sequences. These observations are also supported by the analysis of variance. Significant effects were found for the type of sequence, $F(5,40) = 20.56$, $p < .001$, and the Sequence by Mask interaction, $F(15,120) = 3.797$, $p < .001$.

As can be seen in Figure 4, the effects of intertone interval differ from those found in Experiment I. The higher and lower octave masks still gave performance which was essentially constant over intertone intervals and was quite similar to the no-mask condition. For the within-octave mask, however, increasing the duration of the silent intertone interval failed to produce the better performance typical of a masking function. Although the main effect of intertone interval was significant in the analysis, $F(3,24) = 3.625$, $p < .05$, visual inspection of Figure 4 indicates that the effect of intertone interval was quite small. The significance level obtained is probably a function of the statistical power of the analysis rather than being a psychologically meaningful result. This interpretation is supported by the absence of a significant interaction between the masking condition and the intertone interval ($F < 1$), which would be expected if masking were actually occurring.

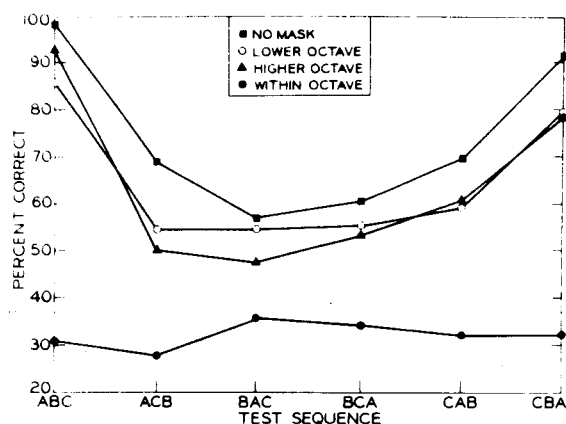


Figure 6. Percentage of correct identifications of each test sequence as a function of the masking condition, in Experiment II.

Examination of the individual subject data indicated that for several of the subjects the within-octave condition was still too difficult to produce above-chance performance. Individual mean percentage correct for the within-octave condition, summed over intertone intervals, reveals that performance was essentially at chance for five of the nine subjects (.191, .209, .109, .173, .199) and just above chance for a sixth subject (.328). The pronounced floor effects for these subjects may have worked to flatten the masking functions in the group data. Of the subjects whose performance was substantially better than chance (.719, .564, .437), two revealed within-octave masking functions. Figure 5 shows a masking function under the within-octave condition for subject S.G. but not for subjects S.M. and J.B., who are at chance performance in this condition.

Discussion

The results of Experiment II can be taken as further support for the strength of the stream segregation effect, and the argument that this effect is perceptual in nature. However, the primary purpose of Experiment II was to assess the viability of an explanation of perceptual splitting in terms of attentional channels, which must be synthesized over time. Contrary to the prediction of a channel-synthesis model, that within-octave masking would occur at short intertone intervals, in Experiment II, when the subject had only 70 msec from target onset to synthesize the channel, the relative distance between the within-octave and the cross-octave masking curves was as great as at the longest intertone intervals. Taken together, the results of Experiments I and II argue against a channel synthesis model of attention. In order for channels to be completely formed within 70 msec of stimulus onset, the amount of information which need be extracted to form the channels would have to be quite minimal. It is

difficult to see how a relatively sophisticated organization could be produced with so little information. Consequently, though it is intuitively appealing, the channel-synthesis notion must be tentatively rejected.

Before exploring an alternative theoretical structure, one possible difficulty with Experiment II deserves mention. That is, changes in the durations of the tones and the intertone intervals were accompanied by simultaneous changes in the frequency separation between test tones. While the adjustments in frequency value were necessary in order to avoid floor effects which would have obscured the data, any cross-octave masking elicited might have been due to the larger frequency separations having induced the inclusion of a larger pitch range within a channel. However, the close consonance between the results of Experiments I and II argues that the phenomenon is sufficiently robust to be independent of the particular frequency range employed, so long as the relationships between tones remain comparable. This point should be kept in mind, as similar adjustments in frequency values were employed in the later experiments also.

Having presented evidence against at least one version of the attentional explanation usually advanced for the stream segregation effect, the alternative explanation suggested earlier can be considered more seriously. Rather than the phenomenon reflecting an attentional limitation of the processing system, it may be taken to indicate a qualitative difference in the way in which sequential auditory information is processed. To indulge momentarily in an analogy, the stimuli used in Experiments I and II are quite similar to the figure-ground demonstrations found in visual perception. The observer's task in this case would be to extract the sequence from the background created by the masking tones. The utility of this analogy is that it demonstrates a shift in emphasis from the negative question of the manner in which irrelevant material is excluded from processing to the positive question of the organization of the relevant information during processing.

The essence of such an argument would be to suggest that the relationship between successive test tones within a sequence provides information which allows for accurate sequence identification—despite the presence of the outside-of-the-octave masking tones—which an individual tone could not by itself provide. This relational judgment may require less information allowing the readout of sufficient information to identify the sequence prior to the onset of the mask. In taking such a tack, it becomes necessary to ask two primary questions. First, what is the nature of the between-tone information critical for processing the sequence? An explicit assumption of the view delineated above would be that this critical

information is something more abstract than the simple addition of the information present in each tone. In addition, it must be asked why a within-octave mask should disturb this information, when cross-octave masks have no effect on it, a major implication of this approach being that the masking stimulus plays a qualitatively different role in a task involving sequence identification than it does in the recognition of a single tone.

In order to answer these questions, it is important to first consider the nature of an auditory sequence. Essentially, such a sequence can be regarded as a melody or melodic phrase. A variety of evidence suggests that there is a qualitative difference between the perception of a single pure tone and that of a melodic sequence. In perceiving melodies, listeners appear to abstract information other than the absolute pitch of each successive tone (Dowling, 1972). The fact that melodies can be transposed into different keys, and yet still be clearly recognized, argues that dimensions other than—or at least in addition to—absolute pitch information are being utilized. This line of reasoning has been verified empirically. Dowling (1972; Dowling & Fujitani, 1970), in the work reviewed above, has found that disturbing the melodic contour of a melody—the sequence of rises and falls in pitch—is more disruptive of melodic recognition than is disturbing the absolute pitch information. He concludes from this that the temporal dimension takes precedence over the pitch dimension in the perception of music.

A somewhat similar—though more formal—proposal has been made by Deutsch (1969, 1973a). She has suggested that the critical dimensions of analysis differ for pure tones and melodies. In perceiving a pure tone, the processing system operates upon two sources of information—the absolute pitch of the tone and its position in the octave. In perceiving a melody, however, the tone is also analyzed in relation to the adjacent tones in the sequence, by extracting the interval—or frequency ratio—between successive test tones. According to her model, a tone is first analyzed in terms of its absolute pitch, or a dimension of tonal height. Following this, the tone is analyzed—in parallel—along two dimensions. Along one dimension, there is convergence of information from exact octave multiples of the tone. The result of analysis along this dimension is the storage, in an auditory short-term memory, of a representation of the tone in terms of its position in an abstract octave. Both sources of information—the tone chroma and the tone height—represent the tone in auditory short-term memory. If the tone is part of a sequence, then analysis proceeds along the second dimension, whose function is to abstract information about successive intervals between test tones. In the perception and memory for a single tone, pitch seems to be the primary source of information. For

melodies, however, successive interval information—which allows judgments of relative pitch—appears to be primary.

Deutsch's (1969) theory receives support from her demonstration of a difference between perception of pure tones and melodies. In perceiving pure tones, there appears to be a large degree of convergence of information from exact octave multiples of the tone (Deutsch, 1972a, 1973b, 1974). However, quite different results are found for melodies. In the study cited earlier, Deutsch (1972b) found that subjects were unable to follow a tune if the notes were displaced randomly across three octaves. This technique preserves the tone chroma—or the relative pitch of the tone within an abstract octave. If there were a large convergence of information across octaves in melodic perception also, it would be expected that subjects could perform the task. Their complete inability to do so suggests that this type of pitch information is much less important for melodies than for single tones. As the tune was easily identified when played entirely in any of the three octaves, absolute pitch does not seem to be the critical factor; another dimension must be providing the necessary information. Deutsch argues that it is exact interval information which is disrupted by displacing tones within three octaves, and interprets her results as supporting the role of successive interval abstraction in the perception of melodies. However, these findings could also result from the use of information concerning melodic contour. In dispersing the tones across octaves, Deutsch did not preserve the directions of pitch change in the original melody. On approximately half the trials, it might be expected that the direction of pitch change would be in the direction opposite to that found in the melody. For example, if A_4C_5 occurred in the original melody, yielding an upward shift in pitch, displacement of these two tones to A_6C_4 would in fact give a downward pitch shift. Thus, both interval and contour information was disrupted by the major manipulation of the study.

The difference in the critical information necessary for processing a melody, as opposed to a single tone, has major implications for the current research. In perceiving a melodic phrase, information about the relationship between tones is used to identify both the sequence and its component tones. Two specific sources for such relational information have been suggested—melodic contour (Dowling, 1972) and exact pitch intervals (Deutsch, 1969). The data from the present research indicates that both types of information may be utilized. If the melodic contour of a sequence is sufficient for its perception, then the exact interval between test tones will not be extracted. If, however, accurate identification cannot be based upon the sequence of pitch changes, then exact interval information will be employed. On the basis of the Deutsch (1972b) results reviewed above, it might

be further suggested that the processing system is limited in terms of the space over which the exact intervals can be computed, the span of an octave serving as a first approximation of the relevant range. In contrast, there appears to be no reason why melodic contour information should be so limited, as large separations would, in fact, increase the discriminability of the contour. Further support for the use of contour and interval information derives from work on the perception of auditory sequences, which has found superior recognition to result for sequences involving a unidirectional pitch change (Divenyi & Hirsch, 1974). Moreover, an even greater facilitation occurred for sequences in which both the direction of pitch change and the exact interval between tones in the sequence was preserved (Divenyi & Hirsch, 1975).

The use of relational information—whether contour or interval—might allow the recognition process to analyze the sequence while extracting less information from each individual tone. As the time course of perceptual processing is directly related to the amount of information which must be extracted from the stimulus (Massaro, 1972a), the relevant information could be read out much sooner in the case of a relational judgment. This suggestion seems reasonable in that melodies can be perceived with notes of durations as short as 50 msec (Warren & Obusek, 1972).

Within the framework of the proposed relational judgment theory, the differential role played by the masking tone in the typical backward recognition masking studies and in Experiments I and II becomes apparent. In an absolute masking task, the mask functions to disrupt the readout of the information present in the preperceptual representation of the test tone, eliminating the absolute pitch information necessary for identification of a single tone. In contrast, if it is relational information which is critical in perceiving melodies, a masking tone should be disruptive only to the extent that this relational information cannot be read out prior to onset of the mask. This is not to argue that the mask fails to disrupt the preperceptual representation of that tone, but simply to suggest that all of the relevant information may be extracted prior to the onset of the masking tone.

This line of reasoning would lead to the interpretation that the within-octave masks disrupted the subject's ability to make relational judgments, while the cross-octave masks did not. Such a conclusion does not seem unwarranted. Assume that the subjects in Experiments I and II were making their judgments on the basis of intervals/contours rather than on the basis of absolute-pitch information. The within-octave masks would be sufficiently close to the test tones that it would be difficult to exclude them from these judgments.

Rather, judgments would have been made relative to the within-octave masking tones—as well as between the test tones. With the exception of the rare instance in which the mask fell between the frequencies of the successive test tones—allowing two successive judgments to produce an accurate relationship between test tones—using the mask to make relative judgments would have had a detrimental effect upon performance. In the case of the outside-of-the-octave masks, the frequencies were so distant from those of the test tones that it would be relatively easy to exclude the masking tones from consideration, allowing the subject to make relational judgments between the test tones across the intervening mask.

This framework is capable of providing a good fit to the data actually obtained in Experiments I and II. Two aspects of the proposed relational judgment theory are relevant here. First, it has been argued that all of the information needed to make such a judgment can be extracted from the tone within a period of time as short as 70 msec from onset of the tone. This contrasts greatly with the 250-msec estimates of the time course of processing required to identify a tone on the basis of its absolute pitch. This implies that, in making relational judgments, less information is required than that needed to make absolute pitch judgments. The argument is substantiated by the finding that performance with the outside of the octave masks was equally good across all interstimulus intervals. If the recognition process were abstracting information after 70 msec, then performance would be expected to improve with increased durations of the intertone interval.

The second relevant aspect of the theory concerns the nature of the information used in making relational judgments. It was suggested that either, or both, exact pitch intervals or melodic contours are employed during the perception of an auditory sequence. A number of findings from Experiments I and II support this idea. First, the question might be raised as to why—if all the relevant information is abstracted prior to the onset of the mask—interference of the within-octave masks on the perception was found. As was suggested briefly, the key point in this context is the frequency separation of the test-tone/masking-tone interval. If exact intervals are being used, then the degree of masking-tone interference would be dependent upon the extent to which the test-mask interval is perceptually viable. If the frequency separation is too great—exceeding that of an octave—the recognition process will be unable to compute the interval. In this case, the mask could be discarded from analysis and then the appropriate interval between test tones perceived. This is precisely what would occur for the higher and lower octave masks. For the within-octave case, however, the test-mask interval would be viable and would be computed, obscuring the relationship between

successive test tones. Similarly, were melodic contour employed, the disparate frequencies of the outside of the octave masks would allow them to be excluded from analysis, while such an exclusion would be impossible for the within-octave masks.

The curves yielded by the different masking conditions support these predictions strongly. Not only were the curves for the lower and the higher octave masks invariant across interstimulus intervals, but performance under these conditions was comparable to that found in the no-mask condition. If the critical relational information were left intact by cross-octave masks, then performance equal to that found in a condition in which the masking interval was simply left blank would be expected. In contrast, the within-octave masks produced substantial decrements in performance, arguing that the mask was interfering with the information critical for analyzing the target. This result is consonant with the suggestion that insertion of the masking tones, in the within-octave condition, confused the relationship between successive test tones.

It is most revealing, in this context, to consider the differential effects of the within-octave and cross-octave masks on the perception of individual test sequences. In both experiments, performance was substantially better with sequences ABC and CBA than any of the other sequences, under all masking conditions, with the exception of the within-octave condition. In these sequences, there is a consistent rise or fall in pitch across the three test tones. As a result, for sequences ABC and CBA, contour information would be sufficient for accurate identification of the sequence. In all sequences except ABC and CBA, however, the recognition process must determine the interval between successive tones. Were this not the case, then sequences with identical patterns of rising and falling pitch—such as ACB and BCA—would be indistinguishable. The high level of performance on these sequences suggests that subjects could accurately discriminate such pairs as ACB and BCA, implying that the extent of pitch difference was also analyzed. As exact interval extraction is presumably more difficult than analysis of melodic contour, an explanation based upon the use of differential information would expect superior performance on sequences ABC and CBA.

If this suggestion is valid, then in the within-octave condition, in which the melodic content of the sequence is severely disrupted, no superiority should be found on ABC and CBA. This is, in fact, what happened. In both Experiments I and II, ABC and CBA were recognized no better than the other four sequences, under the within-octave masking condition. Note also that increasing the frequency separation between tones within the test sequence in Experiment II increased the superiority of ABC and CBA over the other four sequences in the no-mask

and outside-of-the-octave masking conditions. This finding supports the suggestion made earlier that increased frequency differences emphasize melodic contours leading to superior recognition. These results would appear to argue strongly that the information which the within-octave masks disrupt is interval or contour information, indicating again that it is this type of information upon which melodic recognition is based.

While this explanation appears capable of handling the data from the current research quite well, one objection might be raised against it. That is, if interval judgments are the critical factor in recognizing melodies and if this information is assumed to be extracted from each test tone within 70 msec, why then should performance in the within-octave masking condition have increased with increased durations of the interstimulus interval? A post hoc explanation—for which there is no relevant data from Experiments I and II—would be to argue that when interval judgments proved ineffective in analyzing the sequence, the subjects switched to a strategy of processing the tones individually, on the basis of absolute pitch information. In this case, the task would be essentially the same as that in the absolute masking experiments, to recognize a single tone on the basis of its pitch. Under such a strategy, the interstimulus interval would function—as it typically does—to increase the time available for analysis of the auditory image, resulting in more accurate performance with increased durations of the interstimulus interval.

It is by now apparent that a relational judgment theory can handle the data from the current studies quite well. It might also be noted that this proposed role of an auditory structure in producing the perceptual splitting effect is quite compatible with many of the earlier findings, in the sense that an auditory structure or pattern may be necessary for stream segregation to occur. Such a view is in line with the Heise and Miller (1951) finding that the frequency at which a target tone split from the pattern was a function, in part, of the pattern itself. Moreover, it would suggest an explanation for why Norman's (1967) subjects were able to order tones of widely different frequencies only when a single presentation of the pair was used. In order to provide further support for the relational judgment theory, predictions based upon this theory—but not directly addressed in the previous studies—were sought. The fundamental prediction which immediately suggested itself was that if the absence of cross-octave masking results from a processing mode unique to melodic stimuli, then these effects should not be present when single-tone stimuli are employed. That is, if only a single tone and its mask are presented on a trial, absolute pitch information would be required to identify the tone. If the assumption that it is the use of

relational information over absolute pitch information which eliminates cross-octave masking is correct, then, when absolute pitch information is required, this advantage should be eliminated. In this case, masking would be expected to result from tones drawn from any octave.

Some evidence for the prediction that cross-octave masking will occur with single-tone stimuli derives from an earlier study (Massaro, 1970, Experiment V), which required the identification of a single test tone in a backward recognition masking paradigm. Performance improved with increases in the silent interval between the test tone and the mask, and the frequency of the mask had little effect upon the masking functions obtained. These results are in direct opposition to those found in Experiments I and II, and support the relational judgment model presented above. Although the Massaro study is suggestive that cross-octave masking will occur with single-tone stimuli, more direct evidence is required. Experiment III was designed to elicit such evidence. In this experiment, single test tones and masks were presented on each trial, for an identification of the test tone. As in the earlier studies, the masking tones could be drawn from any of three octaves, either the same octave as the test tones, the octave higher, or the octave lower. If the relative judgment theory is correct, then equal masking should result from all three classes of mask, as any auditory mask would interfere with the acoustic information (present in the auditory image) critical for making an absolute pitch judgment.

In addition to providing a test of the relative judgment hypothesis, Experiment III may provide additional evidence relating to the attentional theory of stream segregation. On the basis of Experiments I and II, a channel synthesis model of attention was rejected. However, the Broadbentian (1958) concept of built-in channels still remains a viable alternative. Essentially, the Broadbentian model argues that tones drawn from different frequency ranges belong a priori to different channels. As a result, segregation by octave channels should occur immediately, in a manner analogous to segregation of information arriving at the two ears. The principal prediction that could be derived from such a model is that the presence of an auditory structure would not be necessary to elicit cross-octave masking, even in an absolute masking paradigm in which only a single test tone and its mask are employed. Consequently, if the built-in channel approach is correct, the results of Experiment III should mirror those of Experiments I and II, with no masking resulting from outside of the octave masks. As the predictions of a relational judgment hypothesis and a Broadbentian attentional model are in direct opposition, the results of Experiment III should prove informative for both of these proposed theories.

EXPERIMENT III

Method

Subjects. Subjects were nine University of Wisconsin undergraduates, six of whom were paid \$1.50/h for their services and three of whom received course credit for an introductory course in psychology. One of the paid subjects was dropped from the study after Day 2, as he was still performing at a chance level at that time.

Stimuli. The test tones for Experiment III were changed in frequency from those used in Experiment II, as pilot work had indicated that the task was too easy to elicit masking in any condition. Consequently, the test tones for the current study were chosen so as to be separated by only a single semitone, decreasing the frequency difference between tones: A₄ (461 Hz), C₅ (517 Hz), and D₅ (580 Hz). These tones were again labeled A, B, C—in order of ascending pitch—for purposes of identification. The duration of the test tones was 50 msec. The same 100-msec masking tones were used as for the previous experiment (within octave, B₄ = 488 Hz, C₅ = 548 Hz; lower octave, B₃ = 244 Hz, C₄ = 274 Hz; higher octave, B₅ = 976 Hz, C₆ = 1,096 Hz). The two masks within a given octave were treated as a single level of the masking variable, giving four levels for this variable—higher octave, lower octave, within octave, and no mask. Four additional intertone intervals were included, yielding eight total intervals—0, 20, 40, 80, 100, 120, 160, and 200 msec.

Analysis of the data following Day 2 revealed that the task was still too easy for two of the subjects—both were performing at a level of over 98% accuracy under all experimental conditions. Consequently, for these two subjects only, the frequency difference between the test tones was further decreased to a 20-Hz separation: A = 478 Hz, B = 498 Hz, C = 518 Hz. In order to keep a symmetrical arrangement of the test tones around the masks, new masking tone values were also employed (within octave, 488 and 508; higher octave, 976 and 1,016; lower octave, 244 and 254), being separated from the test tones by 10 Hz in either direction. These changes resulted in a level of performance for the two highly accurate subjects which was comparable to that of the subjects receiving the semitone separations.

Procedure. The experiment was conducted on 7 consecutive days, each day being divided into two sessions.

On Day 1, the first session of 110 absolute learning trials was identical to that for the absolute learning trials of Experiment II. Session 2 of Day 1 and both sessions of all subsequent days consisted of 370 test trials—the first 20 being unscored practice trials. On every trial, one of the three possible test tones was presented, followed—after a variable intertone interval—by one of six possible masks. On one-quarter of the trials, no masking tone was presented. Subjects had 2 sec in which to respond by pressing one of three buttons with the correct tone label. One second of feedback was then provided in the same manner as for the absolute learning trials. The intertrial interval was 1 sec. All 96 experimental conditions (3 tones by 4 masks by 8 intertone intervals) were completely random, and were programmed to occur with equal probability within a session.

Results

Data for five subjects were analyzed for Days 3-7 only, to accommodate the frequency changes made for two of the subjects. Two other subjects were tested at a later date, and their data were analyzed for Days 2-5. There were no appreciable practice effects which would prevent pooling data across different days.

Figure 7 presents the percentage of correct identifications of the test tone as a function of both the duration of the silent intertone interval and the frequency of the masking tone. The figure shows that performance was poorer in any of the conditions in

which a masking tone was presented than in the no-mask condition.

As in the earlier studies, the no-mask condition constitutes a dummy variable since no masking tone was presented. The no-mask condition also entered the analysis of variance as a level of the masking variable. It can also be seen that the obtained masking functions were qualitatively similar for masks drawn from all three octaves. The higher octave mask did appear to be somewhat more difficult than either the lower or the within-octave masks, though the difference was relatively small—approximately 5%. Figure 7 also demonstrates that performance in the masking conditions improved with increased durations of the silent intertone interval. These findings support the contention that the masking tones interfered with perception of the test tones. The masking functions approached asymptote at the 200-msec intertone interval, or with a total processing time of 250 msec. Though performance comparable to that obtained in the no-mask condition was not reached, the difference in performance between the no-mask and the masking conditions decreased from 22% at a 0-msec intertone interval to 5% for the lower and within-octave masks and 12% for the higher octave mask at the 200-msec intertone interval. These conclusions were supported by an analysis of variance on subjects, tones, masking conditions, and intertone intervals. The analysis revealed significant main effects for masking condition, $F(3,21) = 14.68$, $p < .001$, intertone intervals, $F(7,49) = 14.40$, $p < .001$, and the interaction of these variables, $F(21,147) = 2.56$, $p < .05$.

Although there were large individual differences in overall level of performance, the masking effect was present for all subjects except one, who performed at a level only slightly higher than chance. Functions for two individual subjects are presented in Figure 8. The same effects were also present for each individual test tone. Figure 9 presents d' values—which index stimulus discriminability—for each of the three test tones as a function of both the duration of the silent intertone interval and the frequency of the masking tone. The d' value for test tone A is determined from the hit rate of the probability of response A given the stimulus A and the false alarm rate of the probability of the response A given to the stimulus alternatives B and C (Massaro, 1975). The d' values were computed for individual tones on the basis of percentages calculated from the raw frequency data, which was pooled over subjects and days. Essentially the same functions were obtained for all three test tones, though tones A and C were somewhat easier than tone B. This equivalence was confirmed by the absence of significant effects in the analysis of variance for the type of test tone, $F(2,14) = 1.26$, $p > .25$, for the Test Tone by Masking Condition interaction, $F(6,42) =$

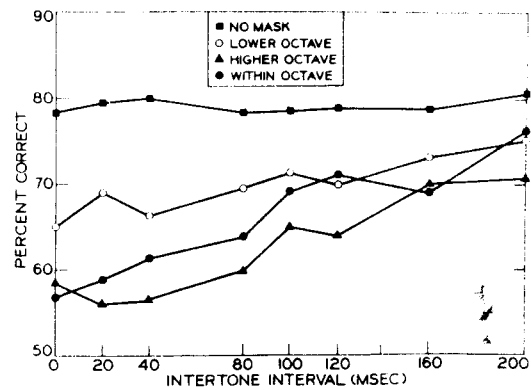


Figure 7. Percentage of correct identifications of the test tone as a function of the masking condition and the duration of the silent intertone interval between the test and masking tone, in Experiment III. Changes in the no-mask condition across the intertone interval is a dummy variable since no masking tone was presented.

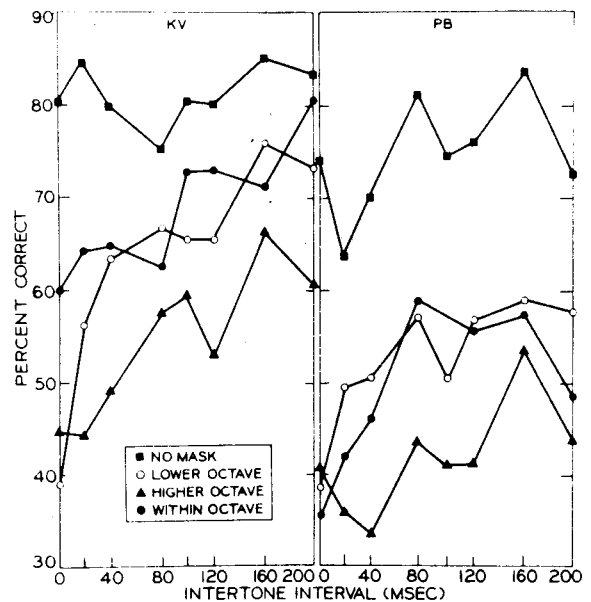


Figure 8. Percentage of correct identifications of the test tone as a function of the masking condition and the duration of the silent intertone interval, for subjects K.V. and P.B., in Experiment III.

1.34, $p > .25$, the Test Tone by Intertone Interval interaction, $F(14,98) = 1.322$, $p > .2$, and the Test Tone by Masking Condition by Intertone Interval interaction, $F < 1$.

Discussion

The principal results from Experiment III indicate that for a single-tone stimulus, masking occurred with masks drawn from any of the three octaves. The presence of any of the three masking tones resulted in significantly poorer performance than that which was obtained in the no-mask condition—in which the interval following the test tone was simply left blank.

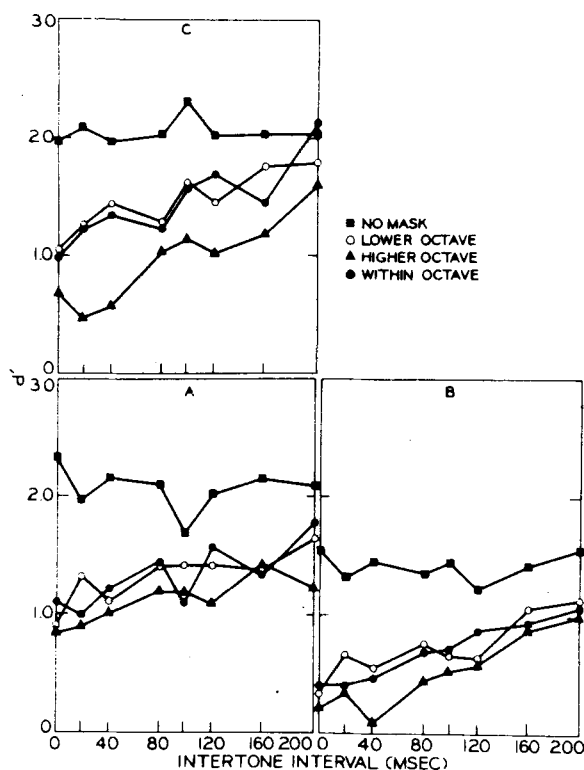


Figure 9. Discriminability values (d') for each test tone as a function of the masking condition and the duration of the silent intertone interval, in Experiment III.

Moreover, performance improved with increased durations of the intertone interval—indicating that presentation of the mask interfered with perceptual processing of the test tones. Though the higher octave mask interfered more with perception of the test tones than either of the other masks, the shapes of the masking functions are qualitatively similar and the absolute differences in performance small—arguing that essentially the same process was occurring with masking from any of the three octaves. Moreover, this same effect of greater interference from higher frequency masks has been found previously with both nonspeech (Divenyi & Hirsch, 1975; Watson, Wroton, Kelly, & Benbassat, 1975) and speech (Massaro & Cohen, in press) stimuli.

The results of Experiment III provide strong support for the idea that an auditory structure is required to eliminate cross-octave masking and, by implication, to elicit the perceptual splitting effect. When an auditory structure is present, cross-octave masking does not occur. With a single-tone stimulus, however, masking will result from a masking tone of any frequency.

The differences in the results of Experiments I and II from those of Experiment III argue not only for the relational judgment hypothesis, but even further against an attentional formulation. It was argued earlier that channel formation could theoretically

occur in one of two ways. Either channels are synthesized over time or belonging to a particular channel is an intrinsic attribute of the stimulus. Together, Experiments I and II provide evidence against the synthesis concept. Experiment III provides evidence against the concept of built-in channels. If belonging to a given channel was an intrinsic attribute of a tone, then by the interpretation given above, the same results should have been obtained in Experiment III as were found in Experiments I and II. As the same frequency relationships between the test tones and masks were retained in Experiment III, if a tone and its mask were assigned to different channels in Experiments I and II so should they have been assigned to different channels in Experiment III. Since masking was, in fact, obtained with all three octave masking tones in the third study, the absence of cross-octave masking effects in Experiments I and II cannot be attributed to the absolute frequencies of the tones. Together, the results of the current research provide evidence against the attentional explanation usually advanced for the stream segregation phenomenon. A channel mechanism—in the sense in which both Bregman and Campbell (1971) and Broadbent (1958) use the term—would appear to be untenable. It is difficult currently to envision what alternative formulation for an attentional channel would not require either time in which to organize the channel or the absence of cross-octave masking for single tones.

Experiment III confirmed the importance of an auditory structure in eliciting the stream segregation effect, supporting the notion that pure tones and melodies are processed in different manners. However, it has yet to be shown that the distinction between these two modes of processing lies in the utilization of interval or contour information, as opposed to absolute pitch. Though the results from the individual sequences in Experiments I and II are suggestive of this, it would be desirable to provide more direct evidence.

Experiment IV seeks to elicit this evidence by eliminating the possibility of utilizing absolute pitch information in processing the sequences. In Experiment IV, the frequencies of the test and masking tones were chosen randomly, from trial to trial, within a 400-Hz range. The subject's task was to identify the tones within a sequence on the basis of their pitch relationships. They were instructed to call the highest tone in the sequence H, the middle tone M, and the lowest tone L—identifying the sequence as a whole by a combination of these three values. As the frequencies varied from trial to trial, over a wide range, it would not be possible to identify the tones on the basis of absolute pitch; rather, frequency relationships would have to be extracted. If, in fact, contour and interval information is critical in perceiving a sequence, then performance in Experiment IV should mirror that in Experiments I

and II. If absolute pitch information were critical, however, subjects should be less able to identify the sequences, given that the frequency differences between the three test tones were of the same magnitude as in the first two experiments.

EXPERIMENT IV

Method

Subjects. The subjects were 13 University of Wisconsin undergraduates who received credit towards an introductory course in psychology for their participation in the 5-day experiment.

Stimuli. The 100-msec test tones in Experiment IV were chosen randomly within a range of 350-750 Hz. On each trial, one of the six possible sequences (HLM, HML, MLH, MHL, LHM, or LMH) was chosen randomly. A frequency value between 530 and 570 Hz was also chosen randomly on each trial. This value was assigned to the first tone of the sequence chosen. Thus, regardless of whether the initial tone was the high, low, or middle tone for that trial, it assumed a value between 530 and 570 Hz. As a result, hearing the first tone of the sequence alone conveyed no information as to whether that tone was high, low, or middle for the sequence. The frequency values for the second and third tones of the sequence were computed from the frequency of the initial tone \pm multiples of 90 Hz. That is, the high and low tones of a sequence were always separated by 180 Hz, while both of these tones were separated from the middle frequency by 90 Hz. A 90-Hz frequency separation was chosen in order to make the sequences from Experiment IV roughly comparable in frequency separation between test tones to those of Experiment I, which also used 100-msec test tones. After the frequency values for all three tones were computed, the ranges of the sequences were adjusted. On half the high-initial trials, randomly determined, 50 Hz was added to each tone in the sequence. On half the low-initial trials, randomly determined, 50 Hz was subtracted from the tones in the sequence. This was done in order to ensure overlap between the frequency ranges of high, low, and middle tone initial sequences.

After all computations of the test tones had been completed, the within-octave mask was computed with reference to the frequency of the first tone in the sequence, being ± 50 Hz from the frequency of this initial tone. The higher octave masks were chosen by computing a value for the within-octave mask and multiplying it by two. Similarly, the lower octave masks were computed as half the frequency value of the within-octave mask chosen on a particular trial. As an example of the stimulus sequences, assume sequence MLH was chosen. If the initial frequency value of 545 Hz was selected for the M tone, the values of the H and L tones would be 635 and 455 Hz, respectively. A within-octave mask of 545 ± 50 Hz would be computed. If the higher alternative was chosen, the within-octave mask would assume a frequency of 595 Hz, the lower octave mask a value of 297 Hz, and the higher octave mask a value of 1,190 Hz, though only one of these alternatives would be employed, in accord with the level of the masking variable chosen.

All tones were adjusted to an equal subjective loudness in the following manner. All test and masking tones between 400 and 600 Hz were presented at 80 dB SPL. Tones between 350 and 400 Hz were presented at 82.5 dB SPL and tones between 600 and 1,750 Hz were presented at 78 dB SPL. Both the test tones and the masking tones had a 100-msec duration.

The experimental task was exactly the same as in Experiment I. The only difference was that the subjects were instructed to call the highest tone H, the middle tone M, and the lowest tone L, on any given three-tone trial. Six sequences directly comparable to those employed in Experiments I and II were thus possible: LMH (ABC), LHM (ACB), MLH (BAC), MHL (BCA), HLM (CAB), and HML (CBA). Subjects responded by pressing one of six buttons labeled with the sequence name.

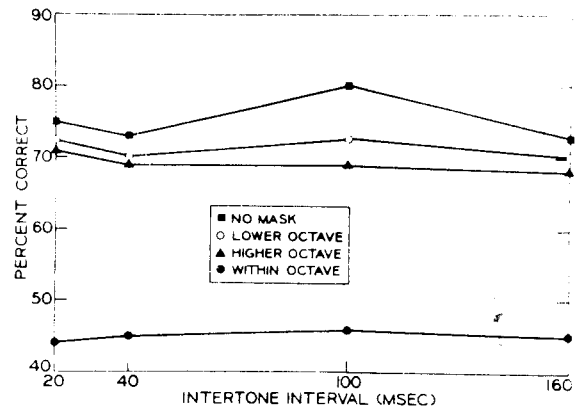


Figure 10. Percentage of correct identifications of the test sequence as a function of the masking condition and the duration of the silent intertone interval between the test and masking tones, in Experiment IV. Changes in the no-mask condition across the intertone interval is a dummy variable since no masking tones were presented.

Results

Data were analyzed for Days 2-5. As can be seen in Figure 10, a substantial effect of the masking condition was again obtained. The within-octave mask produced much greater interference than did the outside of the octave masks. As in Experiments I and II, the higher and lower octave masks yielded essentially equal performance. Though the outside of the octave masks did produce some interference relative to the no-mask condition, the difference was less than 5% overall. Figure 11 gives the results for four individual subjects. For 11 of the 13 subjects, there was no consistent difference between the no-mask and the outside-of-the-octave masking conditions. For two subjects, however, the no-mask condition was easier than either the higher or lower octave conditions. The difference between these conditions in the group data would appear, then, to be largely the contribution of two subjects. These findings were supported by analysis of variance. The analysis revealed a significant main effect for the masking condition, $F(3,36) = 43.74$, $p < .001$. Specific comparisons among levels of the masking variable revealed that the within-octave mask was significantly different from the other three conditions, $F(1,36) = 22.48$, $p < .001$. In contrast, the no-mask condition did not differ significantly from the outside-of-the-octave masking conditions ($F < 1$).

Figure 10 also reveals that there were no differential effects of the intertone interval in any of the masking conditions. Performance was essentially constant over the four intervals. As in the earlier studies, the no-mask condition is simply a dummy variable. The absence of masking in any condition was supported by the results of the analysis. While a significant main effect was found for the intertone

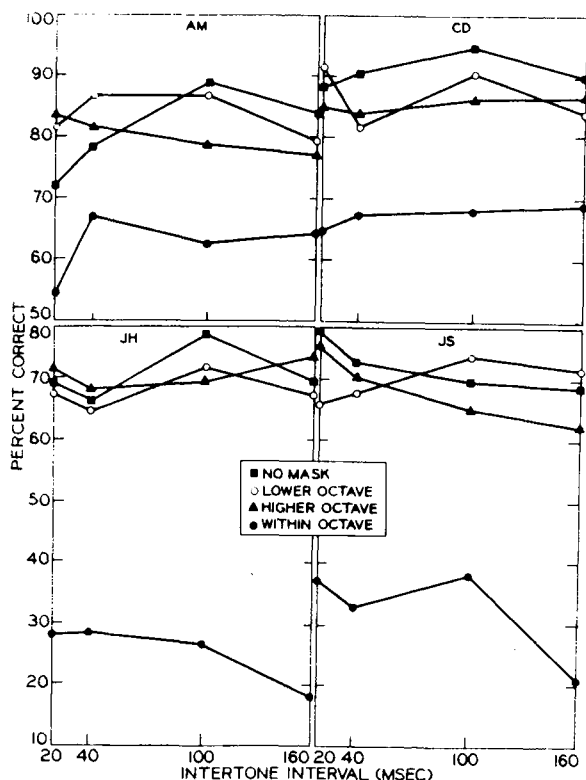


Figure 11. Percentage of correct identifications of the test sequence as a function of the masking condition and the duration of the silent intertone interval, for subjects C.D., A.M., J.S., and J.H., in Experiment IV.

interval, $F(3,36) = 3.28$, $p < .05$, the Masking Condition by Intertone Interval interaction was not significant.

The individual sequences also produced different results. HML and LMH—in which there is a consistent rise or fall in pitch—were identified substantially better than the other four sequences in the outside-of-the-octave and the no-mask conditions. HML was identified approximately 20% more accurately than HLM, MHL, MLH, or LHM, while LMH was identified approximately 30% more accurately than these four sequences. As can be seen in Figure 12, HML had no advantage over the other four conditions in the within-octave masking condition. However, LMH was identified approximately 20% more accurately under this condition. These conclusions were supported by the results of the analysis for sequence effects. Significant effects were found for the type of sequence, $F(5,60) = 23.09$, $p < .001$, and the Sequence by Masking Condition interaction, $F(15,180) = 4.49$, $p < .001$. The Sequence by Intertone Interval interaction was not significant.

Discussion

Experiment IV replicated the basic results of Experiments I and II. The within-octave masks again

produced a substantial decrement in performance, while the higher and lower octave masks did not differ significantly from the no-mask condition. This provides further support for the conclusion that only those masking tones within the octave containing the test tones will interfere with perception of the test tones. Masks drawn from a different octave will be independent of the test tones and will have no influence upon their perception. In contrast to Experiment I, a masking function was not obtained in the within-octave condition. Performance varied by less than 2% across the intertone intervals. Essentially, the same results for individual sequences were obtained as had been found in the earlier studies. HML (CBA) and LHM (ABC) were identified more accurately than the other four sequences in the no-mask and outside-of-the-octave masking conditions. In the within-octave condition, HML had no advantage over the other four sequences, though LMH was identified more accurately.

These results of Experiment IV have a number of implications for the phenomenon under investigation. The replication of the effects of the different masking conditions, when the sequences could not be identified on the basis of absolute pitch, argues that relational information—in the form of intervals or contours—is not only necessary to perform the task, but is, in fact, sufficient. Performance in Experiment IV was similar to that found in Experiments I and II—where the sequence could have been identified on the basis of absolute pitch—quantitatively and qualitatively. That is, in addition to replicating the main effects, the overall level of performance was similar in these studies. Consequently, it might be concluded that absolute pitch information was unnecessary for sequence identification, as subjects perceived the sequences equally accurately with and without this information. These results lend even further support to the hypothesis that relational information, not absolute pitch, is critical in processing a sequence of pure tones.

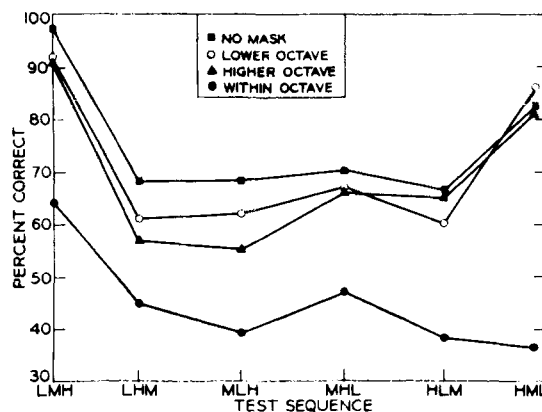


Figure 12. Percentage of correct identifications of each test sequence as a function of the masking condition, in Experiment IV.

The absence of masking in the within-octave condition would also be expected on the basis of a relational judgment theory. It was argued earlier that masking occurred in Experiment I—and for most subjects in Experiment II—since subjects switched to an analysis of the tones on the basis of absolute pitch when relational information proved to be inadequate in the within-octave condition. In this case, subjects would presumably hear the first tone and its following mask and then switch to an analysis of absolute pitch, missing the first tone. This strategy would be relatively successful at long intertone intervals—as seen in Experiment I—since resolution of the pitch of the second and third test tones would allow the first tone to be identified retrospectively. In Experiment IV, however, such a strategy would be ineffective. If the subject missed the first test tone, resolving the relationship between the second and third tones would not greatly reduce ambiguity as to the identity of the sequence. For example, if the subject determined that Tone 2 was higher than Tone 3, this would not allow him to determine whether the sequence was HML, MHL, or LHM. Consequently, the absence of masking in the within-octave condition is completely consonant with the relational judgment hypothesis, though it is at variance with any theory based upon absolute identification of the test tones within the sequence.

The individual sequence effects are also in line with the relational judgment hypothesis. The logic here is identical to that for Experiments I and II. If contour and interval information is critical for sequence identification, then HML (CBA) and LMH (ABC)—which involve a consistent rise or fall in pitch—should be identified better than the other sequences whenever this information is left intact. The superior performance on sequences HML (CBA) and LMH (ABC) in the outside-of-the-octave and the no-mask conditions supports such an interpretation. In the within-octave condition, however, when relational information was presumably disrupted, HML (CBA) and LMH (ABC) should have no advantage over the other sequences. This prediction was supported for HML (CBA), but not for LMH (ABC) in which performance was more accurate than for any of the other sequences. No immediate explanation for this discrepancy suggests itself.

GENERAL DISCUSSION

The results of the current research have provided evidence concerning both the stream segregation effect and the auditory recognition process. Experiments I, II, and IV provide an additional demonstration of the power of the effect. Moreover, they reveal that the segregation of tones into separate streams occurs at the earliest stage of perceptual analysis.

Masking tones drawn from different octaves than the test tones failed to interfere with the perception of these tones, while masks drawn from the same octave caused substantial interference. A comparison of Experiment III to Experiments I, II, and IV indicates that the presence of an auditory structure is critical for eliciting the effect. When a single tone and mask are presented, interference occurs from tones drawn from any octave. Finally, the results of Experiment IV revealed that the critical information for processing a melody is relational information in the form of contours and/or intervals.

A relational judgment theory was offered as a possible explanation for the results and, by implication, for the stream segregation effect. The key assumption underlying such an explanation is that the information used in processing single tones and musical sequences differs. In perceiving single tones, the processing system must rely exclusively on absolute pitch information. In perceiving an auditory sequence, an additional type of information is provided by the relationship between successive test tones in the form of either contours or intervals—which enables the recognition process to analyze the sequence without reference to absolute pitch information. A variety of results (Deutsch, 1969, 1972a, b, 1973a, b; Dowling, 1973; Dowling & Fujitani, 1971) can be interpreted in terms of relational judgments.

The applicability of the relation judgment theory to both the Deutsch (1972b) and Dowling (1973a) papers has been demonstrated above and is unsurprising in that the relational judgment model involves both a synthesis and an extension of this work. However, it can also be applied to related work which less clearly shares the same theoretical orientation. A direct application can be made to the work of Bregman and his associates (Bregman & Campbell, 1971; Bregman & Dannenbring, 1973; Bregman, Note 3) involving report of temporal order. Rather than tracing out the manner in which the relational hypothesis could handle each of these studies, its application to the well-known Bregman and Campbell (1971, Experiment I) study will be delineated, since all other experiments can be handled analogously. In the Bregman and Campbell study, a series of six tones—alternating between two frequency ranges—was presented repeatedly to subjects for a written report of order. The central finding was that subjects could accurately order tones within a range, but could not order tones across ranges. If subjects were treating these sequences as melodies, it seems tenable that as relational information could not be abstracted over the distance of several octaves which separated tones in this experiment, relationships would have been determined between successive tones within the same frequency range. In this case, the three tones within a given frequency range would be processed

independently of the tones in the other range. The subjects then would have no way of reporting relationships across frequency ranges since they did not perceive the tones in this alternating order. This interpretation cannot only handle the Bregman and Campbell results, but it jibes well with their report that their subjects thought the tones had been presented in blocks, rather than simply being unable to remember the presentation order.

The relational judgment model is also highly consonant with a recent study concerning the perception of auditory sequences, to which a different interpretation was given by the authors. Nickerson and Freeman (1974) presented their subjects with sequences of four tones, which repeated for a total of 4 sec. Six sequences were employed: ABCD, ABDC, ACBD, DCBA, CDBA, and DBCA, where A represents the lowest pitch and D the highest pitch. Three different extents of frequency separation were used to construct the sequences. In the most narrow set, all tones were contained within one-third of an octave (586, 631, 680, and 732 Hz); in the medium set, successive tones were separated by about one-third of an octave (469, 586, 782, and 916 Hz); and the widest set involved over an octave's separation between the two low tones (300 and 375 Hz) and the two high tones (1,144 and 1,431 Hz). The subjects' task was to determine which of the six sequences was presented by reporting a number between 1 and 6 associated with the sequence. Nickerson and Freeman found better performance with an increasing frequency separation between test tones. Sequences were identified most accurately in the wide set and least accurately in the narrow set. In addition, performance was critically dependent on individual sequences. In line with the results from the current research, sequences ABCD and DCBA—in which there was a consistent rise or fall in pitch—were identified more accurately than were the other four sequences. Of particular interest, sequences ACBD and DBCA were confused with each other much more than with any of the other four sequences and more often than any other pair of sequences was confused.

Although Nickerson and Freeman (1974) advanced a rather complex explanation to account for their data, it appears somewhat idiosyncratic and is not capable of handling much of the other work on stream segregation easily. In contrast, the relational judgment hypothesis can account for the Nickerson and Freeman results in a manner consistent with that proposed for the earlier studies. The six sequences employed in the Nickerson and Freeman (1974) study could be readily distinguished simply on the basis of contour information, without the necessity of extracting absolute intervals. That is, the sequences ABCD, ABDC, ACBD, DCBA, CDBA, and DBCA correspond to UUU, UUD, UDU, DDD, UDD, and DUD, respectively, where U and D correspond to

up-and-down changes in pitch direction. If directional judgments were being employed, then the larger the frequency difference between successive tones, the easier the determination of pitch direction. As absolute intervals need not be extracted, there seems no reason that the recognition process would be constrained in its judgments over large frequency differences. Were this the case, performance should have been better in the wide set—when directional judgments would be easier—and worse in the narrow set—when directional differences would be more difficult to discriminate. In addition to explaining the main effect of the study, the relational judgment hypothesis is in accord with the effects for individual sequences. Sequences ABCD and DCBA would be more easily recognized due to a consistent pitch direction for successive tones, analogous to the explanation for sequences ABC (LMH) and CBA (HML) in the current studies. Sequences ACBD and DBCA provide even more interesting evidence. Of the six sequences used in the study, only these two involved successive reversals in pitch direction (up, down, up) or (down, up, down). This characteristic would make them easily discriminable from the other sequences. However, if the subject missed the very first direction change, he would have difficulty distinguishing which of the two alternating sequences was occurring during the 4-sec repetition of the sequence. As a result, ACBD and DBCA would be virtually indistinguishable from each other on the basis of direction of pitch changes. In this case, the great confusion between these sequences that was obtained would be highly predictable. Consequently, the Nickerson and Freeman (1974) results are completely predictable from the relational judgment hypothesis.

To summarize briefly, it has been demonstrated that structural information plays a critical role in the analysis of an incoming stimulus even at the earliest stages of processing. The presence of an auditory structure alters the manner in which the individual tones are perceived within the melody. Rather than utilizing absolute pitch information, the processing system can rely upon relational information in identifying the sequence, thereby decreasing both the amount of information which must be abstracted from a stimulus and the time needed to process that stimulus. More specifically, it was suggested that, in processing short musical sequences, either exact intervals or melodic contour are critical. Within this context, the stream segregation phenomenon can be seen as one instance of the operation of a general mode of processing. The effect is striking in that it represents an auditory illusion in which a usually effective processing strategy has resulted in nonveridical perception of order information.

These findings would appear to have more general implications. They suggest that a distinction should

perhaps be made between the information which is potentially available in a complex stimulus and the information which is used in processing that stimulus. If such a distinction is valid, and the nature of the stimulus does in part determine the manner in which it is processed, then no unambiguous conclusions can be drawn concerning process variables, unless the relevant stimulus parameters are first delineated. Consequently, more attention needs to be directed to uncovering the relevant information in a stimulus and possible attendant variations in modes of processing.

REFERENCE NOTES

1. Van Noorden, L. P. A. S. *Rhythmic fission as a function of tone rate*. IPO Annual Progress Report, 1971, 6, 9-12.
2. Massaro, D. W., & Idson, W. L. *Auditory channels*. Paper presented at Psychonomic Society, Boston, November 1974.
3. Bregman, A. S. *Effects of stream segregation on the perception of order*. Paper presented at American Psychological Association, Honolulu, September 1972.

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.
- BREGMAN, A. S., & CAMPBELL, J. Primary auditory stream segregation, perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 1971, 89, 244-249.
- BREGMAN, A. S., & DANNENBRING, G. L. The effects of continuity on auditory stream segregation. *Perception & Psychophysics*, 1973, 13, 308-312.
- BROADBENT, D. E. *Perception and communication*. New York: Pergamon Press, 1958.
- COLE, R. A., & SCOTT, B. Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology*, 1973, 27, 441-449.
- DEUTSCH, D. Music recognition. *Psychological Review*, 1969, 76, 300-307.
- DEUTSCH, D. Effect of repetition of standard and comparison tones on recognition memory for pitch. *Journal of Experimental Psychology*, 1972, 93, 156-165. (a)
- DEUTSCH, D. Octave generalization and tune recognition. *Perception & Psychophysics*, 1972, 11, 411-412. (b)
- DEUTSCH, D. Octave generalization of specific interference effects in memory for tonal pitch. *Perception & Psychophysics*, 1973, 13, 271-275. (a)
- DEUTSCH, D. Interference in memory between tones adjacent in the musical scale. *Journal of Experimental Psychology*, 1973, 100, 228-231. (b)
- DEUTSCH, D. Generality of interference by tonal stimuli in recognition memory for pitch. *Quarterly Journal of Experimental Psychology*, 1974, 26, 229-234.
- DIVENYI, P. L., & HIRSCH, I. J. Identification of temporal order in three-tone sequences. *Journal of the Acoustical Society of America*, 1974, 56, 144-151.
- DIVENYI, P. L., & HIRSCH, I. J. The effect of blanking on the identification of temporal order on three-tone sequences. *Perception & Psychophysics*, 1975, 17, 246-252.
- DORMAN, M. F., CUTTING, J. E., & RAPHAEL, L. J. Identification of vowel order: Concatenated versus formant connected sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, 1, 121-129.
- DOWLING, W. J. Rhythmic fission and perceptual organization. *Journal of the Acoustical Society of America*, 1968, 44, 364. (Abstract)
- DOWLING, W. J. The perception of interleaved melodies. *Cognitive Psychology*, 1973, 5, 322-337.
- DOWLING, W. J., & FUJITANI, D. S. Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, 1970, 40, 524-531.
- FITZGIBBONS, P. J., POLLATSEK, A., & THOMAS, I. B. Detection of temporal gaps within and between tonal groupings. *Perception & Psychophysics*, 1974, 16, 522-528.
- GARNER, W. R. The stimulus in information processing. *American Psychologist*, 1970, 25, 350-358.
- GARNER, W. R. *The processing of information and structure*. Potomac, Md: Erlbaum Associates, 1974.
- GIBSON, J. J. *The perception of the visual world*. Boston: Houghton Mifflin, 1950.
- GIBSON, J. J. *The senses considered as perceptual systems*. Boston: Houghton Mifflin, 1966.
- HEISE, G. A., & MILLER, G. A. An experimental study of auditory patterns. *American Journal of Psychology*, 1951, 64, 68-77.
- MASSARO, D. W. Preperceptual auditory images. *Journal of Experimental Psychology*, 1970, 85, 411-417.
- MASSARO, D. W. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 1972, 79, 124-145. (a)
- MASSARO, D. W. Stimulus information versus processing time in auditory recognition. *Perception & Psychophysics*, 1972, 12, 50-56. (b)
- MASSARO, D. W. *Experimental psychology and information processing*. Chicago: Rand McNally, 1975.
- MASSARO, D. W., & COHEN, M. C. Preperceptual storage in speech recognition. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception*. Heidelberg: Springer-Verlag, in press.
- MASSARO, D. W., & KAHN, B. J. Effects of central processing on auditory recognition. *Journal of Experimental Psychology*, 1973, 97, 51-58.
- MILLER, G. A., & HEISE, G. A. The trill threshold. *Journal of the Acoustical Society of America*, 1950, 64, 637-638.
- NEISSER, V. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.
- NICKERSON, R. S., & FREEMAN, B. Discrimination of the order of the components of repeating tone sequences: Effects of frequency separation and extensive practice. *Perception & Psychophysics*, 1974, 16, 471-477.
- NORMAN, D. A. Temporal confusions and limited capacity processors. *Acta Psychologica*, 1967, 27, 293-297.
- ORTMANN, O. On the melodic relativity of tones. *Psychological Monographs*, 1926, 35(1, Whole No. 162).
- SCHARF, B. Critical bands. In J. V. Tobias (Ed.), *Foundation of modern auditory theory* (Vol. I). New York: Academic Press, 1971.
- TREISMAN, A. M., Shifting attention between the ears. *Quarterly Journal of Experimental Psychology*, 1971, 23, 157-167.
- WARREN, R. Auditory temporal discrimination by trained listeners. *Cognitive Psychology*, 1974, 6, 237-256.
- WARREN, R., & OBUSEK, C. Identification of temporal order within auditory sequences. *Perception & Psychophysics*, 1972, 12, 86-90.
- WARREN, R., OBUSEK, C., FARMER, R., & WARREN, R. Auditory sequence: Confusion of patterns other than speech and music. *Science*, 1969, 164, 586-587.
- WATSON, C. S., WROTON, H. W., KELLY, W. J., & BENBASSAT, C. A. Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals. *Journal of the Acoustical Society of America*, 1975, 57, 1175-1185.