Title: Auditory Perception of Complex Sounds

Author: Ira J. Hirsh

The studies summarized in this report concern auditory perceptual processes that underlie several aspects of complex pattern recognition -- whether of speech, of music, or of environmental sounds. These patterns differ from each other according to the characteristics of individual sound events and also characteristics of the pattern sequences themselves.

Among the sound characteristics, we have focused on pitch, quality and duration. We find that spectral properties of complex tones can be changed to yield changes in both apparent pitch and quality, that individuals differ with respect to relative performance on those dimensions, and that both pitch and quality or timbre can play similar grouping roles in auditory streams.

Most of the experimental work has concerned timing of successive sounds in sequences. We find that at slow rates, listeners detect equally well small temporal offsets or jitters at different positions in the sequence. Increasing the frequency of one of the tones, or increasing the duration of one or two of the successive intervals produces changes in performance at or near the changes. Some of these timing effects are also manifest in the
AUDITORY PERCEPTION OF COMPLEX SOUNDS

ADMINISTRATIVE CHRONOLOGY

AFOSR Grant 84-0335 was originally awarded to Central Institute for the Deaf for work to begin 1 September 1984, Principal Investigators J.L. Lauter and I.J. Hirsh. A Technical Report of September 1985 was submitted to the AFOSR Program Manager on 15 October 1985.

A second award was made for a two-year period beginning 1 September 1985 with I.J. Hirsh as Principal Investigator, after J.L. Lauter left Central Institute to pursue her research at the University of Arizona. A Technical Report for the period 1 September 1985 to 31 August 1986, dated 4 February 1987, was submitted to the Program Manager on 20 February 1987.

Continuing support for the third year began on 1 September 1986. The Program Manager has received, in lieu of the Annual Technical Report, the Progress Report contained in the January 1987 Research Proposal for renewal of AFOSR support.

The following Final Technical Report covers the three-year period from 1 September 1984 to 31 August 1987.

1. SUMMARY

The studies summarized in this report concern auditory perceptual processes that underlie several aspects of complex pattern recognition — whether of speech, of music, or of environmental sounds. These patterns differ from each other according to the characteristics of individual sound events and also characteristics of the pattern sequences themselves.

Among the sound characteristics, we have focussed on pitch, quality and duration. We find that spectral properties of complex tones can be changed to yield changes in both apparent pitch and quality, that individuals differ with respect to relative performance on those dimensions, and that both pitch and quality or timbre can play similar grouping roles in auditory streams.

Most of the experimental work has concerned timing of successive sounds in sequences. We find that at slow rates, listeners detect equally well small temporal offsets or jitters at different positions in the sequence. Increasing the frequency of one of the tones, or increasing the duration of one or two of the successive intervals produces changes in performance at or near the changes. Some of these timing effects are also manifest in the rhythmic aspects of spoken sentences.

2. RESEARCH OBJECTIVES

Our overall aim is to increase our understanding of how human listeners discriminate, recognize and identify complex sequential patterns of sound. We know that such complex sequences as melodies, phrases or sentences in speech, or
environmental sound patterns are distinguished according to two general categories of cues: (1) characteristics of the individual sound events, their loudness, pitch, quality, duration; and (2) characteristics of the sequences, their timing, the order of events within the sequence, and other temporal form properties yet to be specified. In addition, there is likely a third category namely (3) the interaction between the sound-event characteristics, and those of the sequence.

Our approach to these problems of auditory pattern recognition has followed this rationale (Lauter and Hirsh, 1985). In the first of these three years, Lauter (1985) and Singh (1984, 1987) were concerned with listeners' discrimination of pitch and sound quality, or timbre. At the same time, Hirsh, Grant and Singh (in preparation) began examining the sensitivity of listeners to small changes in the timing of simple sound patterns.

The second year saw the addition of studies of the interaction between pitch and event timing. Further complexities in the stimulus patterns required a whole new system for creating sounds and controlling experiments. Development of this new system, with considerable support from AFOSR, occupied several months, but enabled us to study further more complicated timing patterns (Hirsh and Monahan, Submitted) and further interactions between sound duration and timing (Ibid), and between pitch or timbre and sequential properties (Singh, 1987), in the third year.

2.1 Status of research

2.1.1 Pitch and Timbre. In a pitch-discrimination task, we asked listeners to judge if the second tone in a three-tone sequence was the same, or different, from the first and third tones, which were identical. The intention was to see how different listeners performed as a function of the "pitch strength" of the stimuli. The "pitch strength" of a sound is known to vary as a function of its spectral density (Fastl and Stoll (1979), Hartmann et al. (1985)). We thus intended to study discrimination performance for stimuli with clear, strong pitch, such as pure tones, in comparison with complex-tones and stimuli with more ambiguous, weak pitch, such as narrow-band, filtered noises. We began by comparing discrimination performance for 1000 Hz. In one case the stimuli were pure tones of 1000 Hz. The second tone in the three-tone sequence was made different by varying amounts of frequency change. A second group of stimuli comprised complex tones, in which 1000 Hz was a harmonic and the "different" tone was the same complex but with the 1000 Hz component now shifted by different amounts. The analysis of results showed great differences between individuals with respect to their discrimination performance for these two kinds of stimuli. In general, the 1000 Hz frequency component had to be changed by greater amounts as a member of a complex, than as the sole source of pitch in a pure tone to elicit equivalent performance (Lauter, 1985).

The task defined for both sets of stimuli was one of pitch discrimination. Different listeners however, appear to have adopted different perceptual criteria for making their judgements. Listeners who performed well with complex tones, may have detected a change in the frequency of a component on the basis of a perceived change in the overall "quality" of the sound. For other listeners operating on a pitch-based criterion, the same frequency change may not have been detected because it did not lead to a salient enough pitch change.
for the tone as a whole. Both these are examples of frequency discrimination, with timbre, or quality being the perceptual correlate in the first case, and pitch in the second case. Frequency discrimination may thus be manifested in several different ways. One may not be able to predict who will be a sensitive pitch discriminator based on timbre discrimination performance, and vice versa.

Another study on the relation between pitch and timbre tested the relative importance of pitch patterns and timbre patterns in determining the perceptual organization of complex-tone sequences. A previous, long literature on auditory-stream segregation concerned the apparent separation of a string of pitches into streams or sub-melodies, when the pitches spanned markedly different frequency regions. We have shown (Singh, 1984; 1987) that such segregation may also be mediated by differences in timbre, with similar timbres forming a perceptual group. Further, such grouping may be used to demonstrate a type of trade-off between pitch and timbre.

From the results obtained in the experiments on frequency discrimination and perceptual grouping of complex-tones, it appears that the perceptual correlates of spectral changes in the design of complex tones may differ for different listeners, different tasks, and different sequential contexts. The fact that pitch and timbre appeared to trade in a continuous way in the grouping experiments further seems to imply that these perceptual attributes may be derived via a common spectrally-based process that is sensitive to timing differences between events in a sequence. Before further experimentation on the influence of time relations on the relative salience of pitch and timbre, we propose to expand on the idea that both pitch and timbre discrimination may be subserved by frequency discrimination. Changes in the frequency of components of a complex tone may influence both its pitch and its timbre.

A new set of experiments on spectral discrimination addresses this issue of confounded percepts. In a paradigm similar to the one described above for pitch discrimination, listeners will be required to make two judgements. One, a generic frequency discrimination of the middle tone in an AAA or ABA sequence, and another about the nature of the difference — a pitch change, or some other qualitative (timbre) change. By providing more options for listener-response, we hope to determine the relation between frequency changes and accompanying perceptual changes.

2.1.2 Duration. In some preliminary experiments, in which we emphasized the differences between the duration of individual sound events and the interval that separates the onsets of two or more of such events, we sought to establish whether or not the classical data on sound-duration discrimination (filled intervals) were applicable to durations in sequences. We find that such discrimination is similar in sequences and in separate pairs, and shows a constant Weber ratio of between .05 and .10 for base durations down to about 50 msec, below which the ratio goes up. These data will be incorporated in a report on interval timing (Hirsh, Grant and Singh, In preparation).

2.1.3 Interval timing. The first sequence property that we studied was the timing of successive sound onsets. In the simplest case we present a series of brief sounds, equally timed or isochronous. After some preliminary observations at rates or tempos, we designed an experiment in which sequences of 3, 6, or 10 tones were presented at "standard" isochronous rates corresponding to inter-tone
intervals of 50, 200, or 800 msec. One tone in the sequence was delayed by a variable amount, depending on the performance of the listener in discriminating this "different" sequence from the standard. The delay could be introduced at any tone in the pattern; we chose to delay tone 2 in a sequence of 3, tone 2, 3, and 5 in a sequence of 6, tone 2, 3, 5, or 9, in a sequence of 10 tones.

The results (See Figure 1) show that discrimination of a small change in the timing of an isochronous sequence of brief tones holds at about 5-7% for the two slower tempos and appears to be the same for all of the different temporal positions sampled. At the fastest tempo, with intervals of 50 msec, corresponding to rates of 20 tones/sec, the performance is worse with all $\Delta t/t > 10\%$. Further, the discrimination is better for delay positions late in the sequence than for earlier ones (Hirsh, Grant and Singh, In preparation).

2.1.4 Interval timing with a pitch change. Not only were the brief tones in the previous study isochronous, they were also monotones, that is all at the same pitch. Our findings for the monotone sequences show that offset discrimination is not dependent on position of offset for standard intervals of 200 and 800 msec. We now introduce a pitch change in one of the tones of a 6-tone sequence with a standard timing interval of 200 msec. The pitch change can be introduced at positions 2, 3, or 5 in the sequence. Figures 2 and 3 show the interactions between the position of the pitch change and position of the temporal offset, for downward and upward frequency changes respectively. No matter where in the sequence the pitch change occurs, temporal discrimination for the tone whose pitch is different is impaired.

2.1.5 Rhythmic patterns. The introduction of a pitch change in otherwise monotone and isochronous sequences resulted in increased absolute and relative difference limens (DLs) for delay of tones that were changed in pitch, for both directions and all positions of the pitch change. Such a pitch change might represent a demand for a shift in attention, or it may have been a source of accent (Thomassen, 1982). In the present two experiments we use the organization of time intervals as a source of accenting and grouping. Tones that initiate relatively longer intervals (measured from a tone's onset to the onset of the next tone) generally sound accented and also as if they end a rhythmic grouping. Additionally, Povel & Okkerman (1981) have reported that tones that end longer intervals also sound accented if they then begin a rhythmic group of 3 or more tones. Povel & Essens (1985) have demonstrated that temporal patterns that have equally-spaced temporal accents take less time to code and are tapped out with less variability than those having unequally-spaced accents. In the present experiments we have included 6-tone series that have intervals of two sizes, "long" (2) which is twice the length of "short" (1). In such series we can exemplify rhythms that have both equally and unequally spaced accents. Specifically, we were interested in whether the timing DLs for such patterns were affected by pattern tempo, interval sizes within the patterns, and also of what might be called temporal accent regularity or simplicity (Povel & Essens, 1985).

Listeners discriminated between 6-tone rhythmic patterns that differed only in the temporal position of one of the tones. In an adaptive, cued-2AFC procedure each trial comprised three patterns: a standard followed by 2 comparisons. At random, one comparison was always identical to the standard
while the other differed only by a delay in the onset of one tone relative to its position in the standard. On each trial the subject indicated which pattern contained the delayed tone and feedback was given. The subject's performance determined the amount of delay on the succeeding trial. The 6 tones of the patterns marked off 5 intervals. In the first study, 5 patterns comprised 4 "shorts", and 1 "long" interval: thus, 21111, 12111, 11211, 11121, and 11112, where the "long" (2) was twice the length of a "short" (1). In the second study, 8 patterns comprised 3 "shorts" and 2 "longs": thus, 22111, 12211, 11221, 21211, 12211, 11121, 21211, and 12112. Additionally, the pattern 11211 was replicated from the first study. Each pattern was tested 5 times, with a delay of the tone ending each of its 5 intervals (which concomitantly shortened the following interval). Patterns were tested at three tempos, where "short" was 50, 100, or 200 ms and "long" was 100, 200, and 400 ms, respectively.

Absolute discrimination (Δt in ms) was poorer, the slower the tempo; Figures 4 and 5 show this result for the first and second studies, respectively. On the other hand, relative discrimination (Δt/Δt "short") was better, the slower the tempo; Figures 6 and 7 show this result for the first and second studies, respectively. Generally, in the first study, and especially at the slowest tempo in the second, discrimination of delay was poorest when the delay occurred between two "longs", was best between two "shorts", and of intermediate value between "short-long" or "long-short". These results are in line with Weber's Law and agree with previous findings of Bharucha & Pryor (1986) who tested for DLs in temporal sequences employing an entirely different method. Three models were generated to predict the discriminability of delay: the first predicted discriminability of delay as an inverse function of the sum of the lengths of adjacent intervals that are changed by the delay; the second and third, predicted that discriminability was related to perceived accent which is usually reported at the beginning and in some cases, at the ends of relatively lengthened intervals (Povel and Essens, 1985). The predictions made by the models are similar, but those of the first were correlated most with performance, accounting for more than half the variance in some conditions.

2.1.6 Timing in speech. The rhythmic structure of speech is given through accented and nonaccented syllables. It is then reasonable to hypothesize that variations in the metrical structure of the sentence will, everything else being equal, have effects on the timing of these smaller units. The purpose of our production experiment was to determine the nature of the effects and the level to which these effects extend. We examined the timing of sentences made up of 3 metrical feet, where each foot was either an iamb or an anapest. The shortest sentence contained 3 iambs (6 syllables) and the longest contained 3 anapests (9 syllables) as shown:

```
((Your) machines){will (soon) [C][V][C]}{(at) my desk).
```

Foot 1 Foot 2 (Target) Foot 3

The stressed syllable of the second foot was a target word of C1VC2 structure, where C1 varied over [p, t, k, b, d, g], V was either [a] or [i], and C2 was either [s] or [z]. There were thus 192 sentences: first-foot type (2) x second-foot type (2) x third-foot type (2) x 6 stops x 2 vowels x 2 fricatives.
Recordings were made of 3 male and 3 female native speakers of American English reading randomized lists of the 192 sentences. The recordings were then spectrographically analyzed and the following intervals were segmented and durations measured: the syllables comprising each foot, the closure for the initial stop of the target word, its VOT (if any), the durations of the vowel and the final fricative.

Preliminary results indicate a significant effect of the type of metrical foot on the duration of a preceding foot, i.e., second-foot affects first, and third-foot affects second. The first-foot shortens when it is followed by an anapest as opposed to an iamb, and the second-foot exhibits the same shortening pattern. We examined each subject's data separately to find if production behavior was consistent. We ran analyses of variance for each subject using the 24 sentences with different target words as replications of each sentence pattern. We found that all 6 subjects shortened the first-foot when the second was an anapest, but the effect was significant (p < .01) for only 3 subjects. On the other hand, all 6 subjects showed significant shortening of the second-foot when the third was an anapest.

Since each foot was made up of either 2 or 3 syllables, we decided to examine subject consistency for each syllable of each foot. Only the stressed syllables of the first- and second-foot were shortened consistently when the following foot was an anapest. The effect of shortening the stressed syllable was significant for 3 subjects for the first-foot and for all 6 subjects for the second-foot.

The different behaviors may involve different strategies of sentence production planning. The 3 totally consistent subjects may plan the timing of the sentence 2 feet at a time. The other 3 start the first-foot in the same way regardless of what may follow, but by the time they are producing the second-foot they are already looking forward to the following foot and the end of the sentence. That different subjects may use different strategies in the production and timing of the phonetic elements of their language is well documented. This research adds foot structure and, consequently, sentence length to the variables involved in planning and producing an utterance.

2.2 Interpretations and Conclusions

We have shown that the temporal form of a sequence is influenced by characteristics of the component sounds, as well as by features of the temporal or metrical form itself. We have not carried out a sufficient variety of conditions to know whether the important features or aspects can be regarded as belonging to the stimulus pattern itself, or whether such features get constructed by perceptual or cognitive processing.

3. PUBLICATIONS AND MANUSCRIPTS


4. OTHER REFERENCES CITED IN TEXT


5. PARTICIPATING PROFESSIONALS

Ira J. Hirsh

Northwestern University, Evanston, IL  A.M.  1943  Speech
Harvard University, Cambridge, MA  M.A.  1947  Psychology
Harvard University, Cambridge, MA  Ph.D.  1948  Psychology

Dissertation title: "Interaural Summation and Inhibition."

Caroline Monahan

Boston University, Boston, MA  M.A.  1966  Social Psychology
University of California, Los Angeles, CA  Ph.D.  1984  Psychology


Punita Singh

Washington University, St. Louis, MO  M.A.  1984  Psychology

Thesis title: "Dimensional Tradeoffs in the Perception of Complex Tone Sequences."

6. PRESENTATIONS

Ira J. Hirsh

"Perception of rhythmic patterns." Association for Research in Otolaryngology, Clearwater Beach, FL, 5 February 1985

"Temporal aspects of auditory perception." Tokyo Medical and Dental University, Japan, 4 September; East China Normal University, Shanghai, 8 September; Institute of Psychology, Chinese Academy of Sciences, Beijing, 12 September 1987.

Caroline B. Monahan

"How does the processing of contour, interval and hierarchical information differ for pitch and time patterns?" Invited paper given at the Symposium on Music Perception, Department of Psychology, University College, London, June 1986.

7. CONSULTING AND ADVISING

Ira J. Hirsh
Review manuscripts for Perception & Psychophysics, and for the Journal of the Acoustical Society of America.

Chair, National Research Council's Commissions on Behavioral and Social Sciences, 1982-1987.

Caroline B. Monahan
Review manuscripts for Perception & Psychophysics.

Punita Singh
Figure 1  Discrimination of a temporal offset in one of a series of tones as a function of the position of the offset.
Figure 2 Discrimination of a temporal offset in a sequence of 6 tones, one of which, in positions 2, 3, or 5, has a downward shift in frequency. Abscissa shows position of offset.
Figure 3 Discrimination of a temporal offset in a sequence of 6 tones, one of which, in positions 2, 3, or 5, has an upward shift in frequency. Abscissa shows position of offset.
Figure 4 Absolute DLs for delay (Δt in ms) for five temporal patterns comprising six tones and five intervals (21111, 12111, 11211, 11121 and 11112). In each pattern, tones 2, 3, 4, 5, and 6 were delayed. Δt was measured at each of three tempos, shown as the parameter. Upper 5 panels show Δt for these conditions when 20 ms tones initiated each interval. Lower 5 panels show Δt for the same patterns and serial positions of delayed tone for two slower tempos when 50 ms tones initiated each interval. Each point is the average of 3 subjects' average DL across 4 replications of conditions. Standard errors are based on an n of 3; standard errors less than 1 ms are not shown. Absolute DLs are an inverse function of tempo: the faster the tempo, the
Figure 5 Absolute DLs for delay (Δt in ms) for 9 temporal patterns comprising 6 tones and 5 intervals (top row, 22111, 12211, 11221; middle row, 21211, 12121, 11212; bottom row, 21121, 12112, 11211). In each pattern, tones 2, 3, 4, 5, and 6 were delayed. Δt was measured at each of three tempos, shown as the parameter. 20 ms tones initiated each interval. Each point is the average of 4 subjects' average DL across 3 replications of conditions. Standard errors are based on an n of 4; standard errors less than 1 ms are not shown. Absolute DLs are an inverse function of tempo: the faster the tempo, the lower the absolute DL.
Figure 6. Relative DLs for delay (Δt/t) where t is the shorter interval in the pattern tested for the same conditions described in Figure 1. Standard errors are not shown, since analysis was on absolute DLs. Relative DLs are a direct function of tempo: the slower the tempo, the smaller the relative DL, contrary to Weber's Law.
Figure 7 Relative DLs for delay (Δt/t) where t is the shorter interval in the pattern tested for the same conditions described in Figure 4. Standard errors are not shown since analysis was on absolute DLs. Relative DLs are a direct function of tempo: the slower the tempo, the smaller the relative DL, contrary to Weber's Law.
End Date Filmed 4-88 DTIC