

in K. Boff, L. Kaufman and J. Thomas (Eds.)
Handbook of Perception and Human Performance, 1986, Wiley

CHAPTER 32

AUDITORY PATTERN RECOGNITION

DIANA DEUTSCH

University of California, San Diego, La Jolla, California

CONTENTS

- | | | | |
|---|-------------|---|--------------|
| 1. Auditory Grouping Phenomena | 32-2 | 1.4. Grouping and Selective Attention, 32-20 | |
| 1.1. Parsing of Sounds of Complex Spectral Composition, 32-3 | | 1.4.1. Voluntary and Involuntary Grouping, 32-20 | |
| 1.1.1. Harmonicity of Spectral Components, 32-3 | | 1.4.2. Consequences of Attention Focusing, 32-20 | |
| 1.1.2. Time-Variant Relationships, 32-4 | | 2. Shape Analysis for Pitch Structures | 32-21 |
| 1.1.3. Familiarity, 32-5 | | 2.1. Auditory Shape Analysis as a Multileveled Process, 32-21 | |
| 1.2. Grouping of Sound Sequences in Space, 32-5 | | 2.2. Passive Versus Active Processing, 32-21 | |
| 1.2.1. Auditory Illusory Conjunctions, 32-5 | | 2.3. Feature Abstraction, 32-21 | |
| 1.2.2. The Scale Illusion, 32-6 | | 2.3.1. Octave Equivalence, 32-21 | |
| 1.2.3. Grouping of Nonsimultaneous Sound Sequences, 32-7 | | 2.3.2. Interval and Chord Equivalence, 32-21 | |
| 1.2.4. The Hypothesis of a Slow Switching Mechanism, 32-8 | | 2.3.3. Categorical Perception of Musical Intervals, 32-22 | |
| 1.2.5. The Octave Illusion, 32-9 | | 2.3.4. Global Cues, 32-22 | |
| 1.2.6. Grouping of Phase-Shifted Tones, 32-14 | | 2.3.5. Interval Class, 32-22 | |
| 1.3. Grouping of Rapid Sound Sequences, 32-16 | | 2.4. Higher-Order Abstractions, 32-25 | |
| 1.3.1. Grouping by Frequency, 32-16 | | 2.5. Hierarchical Encoding of Pitch Sequences, 32-27 | |
| 1.3.2. Grouping by Frequency Proximity, 32-16 | | 2.6. The Influence of Short-Term Memory on Perception of Pitch Patterns, 32-29 | |
| 1.3.3. Temporal Coherence as a Function of Frequency Proximity and Tempo, 32-16 | | 2.6.1. Interference Effects in Short-Term Memory for Pitch, 32-29 | |
| 1.3.4. Grouping by Frequency Proximity in Relation to Repetition, 32-16 | | 2.6.2. Facilitation Through Repetition in Short-Term Memory for Pitch, 32-31 | |
| 1.3.5. Frequency Proximity and the Perception of Temporal Relationships, 32-17 | | 2.6.3. The Influence of Relational Context on Pitch Comparison Judgments, 32-31 | |
| 1.3.6. Grouping by Good Continuation, 32-18 | | 2.7. Contour as a Cue in Recognition of Pitch Patterns, 32-32 | |
| 1.3.7. Grouping by Sound Quality, 32-19 | | 2.8. Scale and Key Structure in Recognition of Pitch Patterns, 32-32 | |
| 1.3.8. Grouping by Amplitude, 32-19 | | 2.9. Memory for Hierarchically Organized Pitch Patterns, 32-33 | |
| 1.3.9. Grouping by Temporal Position, 32-19 | | 3. Analysis of Timbre | 32-34 |
| 1.3.10. Grouping by Spatial Location, 32-19 | | 3.1. Timbre and Fourier Analysis, 32-34 | |
| 1.3.11. Closure: The Auditory Continuity Effect, 32-19 | | | |

3.2. Investigation of Timbre by Analysis and Synthesis, 32-35	
3.3. Multidimensional Models of Timbre, 32-35	
3.4. Role of Context in Timbre Perception, 32-36	
4. Perception of Temporal Relationships	32-37
4.1. Perception of Temporal Order, 32-37	
4.1.1. Modes of Order Perception, 32-37	
4.1.2. Perception of the Order of Two Events, 32-37	
4.1.3. Perception of the Order of Three or More Events, 32-37	
4.1.4. Order Perception in Continuously Cycling Sound Patterns, 32-37	
4.1.5. Theories of Order Perception, 32-38	
4.2. Perception of Rhythm, 32-38	
4.2.1. Subjective Rhythmic Grouping, 32-38	
4.2.2. Grouping by Temporal Proximity, 32-38	
4.2.3. Grouping by Accent, 32-39	
4.2.4. Grouping by Other Principles, 32-39	
4.2.5. The Run Principle and the Gap Principle, 32-39	
4.2.6. Rhythmic Hierarchies, 32-40	
5. Summary	32-43
Notes	32-43
References	32-44

Research on hearing has traditionally been concerned with simple detection, discrimination, and scaling tasks. However, the last decade has seen a flowering of interest in higher-level mechanisms concerned with auditory grouping, shape perception, memory, and so on. This new development has been due largely to technological advances that have enabled researchers to generate complex auditory stimuli with precision and flexibility. Those entering the field have been rewarded by the discovery of an elaborately structured and highly differentiated system that possesses some remarkable properties.

Two major influences on research into auditory pattern recognition may be identified. The first stems from related work in perceptual and cognitive psychology. For example, the multi-leveled approach to auditory shape perception has been strongly motivated by theoretical and experimental work on the perception of visual shape. As another example, research into memory for sound structures has been influenced by findings on memory for verbal materials.

A second major influence derives from music theory. Fundamental concepts such as octave equivalence and interval equivalence have been in the mainstream of traditional music theory since the time of Pythagoras. Several developments in contemporary music theory have also provided input. For example, the theory of 12-tone composition, developed by Schoenberg, is based on an implied theory of shape analysis for pitch structures. Another example is the hierarchical theory of tonal music, developed early in this century by Schenker, which has points of similarity with the theory of transformational grammar developed later by Chomsky. In addition, composers of electronic and computer music have provided the major impetus to recent experimental work on the perception of sound quality or timbre, an area of research with broad implications for auditory perception in general.

This chapter is divided into four main sections. In the first, auditory grouping phenomena are investigated. This section deals with questions concerning the perceptual fusion and separation of components of a complex sound spectrum, the grouping of sound elements emanating from different spatial locations, and the grouping of sounds that occur in rapid succession. The second section is concerned with the perception and recognition of patterns formed of pitch combinations. The third section deals with the perception of timbre or sound quality. The fourth section is concerned with the perception of temporal order and of rhythm. The final section summarizes the findings in these different subfields.

1. AUDITORY GROUPING PHENOMENA

We may distinguish two basic but interrelated questions in considering how the auditory system groups stimuli into perceptual configurations. The first involves the stimulus dimensions along which grouping principles operate. When presented with a complex signal, the auditory system may group elements according to some rule based on frequency, on amplitude, on temporal or spatial position, or on some multidimensional attribute such as timbre. As will be shown, any of these attributes may serve as a basis for grouping, and further, there are complex and rigid rules determining which attribute will be used. Such rules can often be well interpreted in terms of strategies that are most likely to lead to the correct conclusions in interpreting our auditory environment. Second, we may enquire into the principles that govern grouping along any given dimension. The Gestalt psychologists proposed that we form groupings on the basis of certain simple principles, such as proximity, good continuation, similarity, and common fate (Wertheimer, 1923). As described elsewhere in this *Handbook*, these have been shown to be important descriptive principles for grouping in vision. We shall show here that this is true for hearing also. It may plausibly be argued that grouping in conformity with such principles enables us to interpret our environment most effectively (Bregman, 1978; D. Deutsch, 1975c; Gregory, 1970; Hochberg, 1974; Sutherland, 1973). Sounds that are similar are likely to be coming from the same source, and sounds that are dissimilar are likely to be coming from different sources. A sequence of sounds is more likely to be coming from a single source if it contains frequency transitions that are gradual rather than abrupt. Components of a sound spectrum that modulate in synchrony are more likely to be coming from a single source than those that modulate out of synchrony.

The view of auditory grouping as a process of unconscious inference may be traced to Helmholtz (1859/1954) (see Note 1). He speculated how, given the complex, time-variant spectrum produced by several musical instruments playing simultaneously, the listener reconstructs the auditory environment so that some components of the spectrum fuse perceptually to produce the impression of a single sound, while others are heard as separate melodic lines sounding in parallel. He wrote:

Now there are many circumstances which assist us first in separating the musical tones arising from different sources, and secondly, in keeping together the partial tones of each separate source. Thus when one musical tone is heard for some time before being joined by the second, and then the second continues after the first has ceased, the separation in sound is facilitated by the succession of time. We have already heard the first musical tone by itself and hence know immediately what we have to deduct

from the compound effect for the effect of this first tone. Even when several parts proceed in the same rhythm in polyphonic music, the mode in which the tones of different instruments and voices commence, the nature of their increase in force, the certainty with which they are held and the manner in which they die off, are generally slightly different for each . . . but besides all this, in good part music, especial care is taken to facilitate the separation of the parts by the ear. In polyphonic music proper, where each part has its own distinct melody, a principal means of clearly separating the progression of each part has always consisted in making them proceed in different rhythms and on different divisions of the bars. . . . All these helps fail in the resolution of musical tones into their constituent partials. When a compound tone commences to sound, all its partial tones commence with the same comparative strength; when it swells, all of them generally swell uniformly; when it ceases, all cease simultaneously. Hence no opportunity is generally given for hearing them separately and independently. (pp. 59–60)

1.1. Parsing of Sounds of Complex Spectral Composition

A basic task for auditory theory is to determine the relationships between elements of an ongoing sound spectrum that give rise to the perception of a single sound and those that give rise to the perception of several simultaneous sounds. Without these processes of fusion and separation, intelligible listening would not be possible. Presumably mechanisms have evolved that cause us to fuse together those elements of the sound spectrum that are likely to be coming from the same source and to separate out those elements that are likely to be coming from different sources. Three factors will be considered here. The first is harmonicity of spectral components; the second is synchronicity; the third is familiarity with certain sound complexes.

1.1.1. Harmonicity of Spectral Components. It has been argued from various lines of evidence that harmonic sounds are more likely to be perceived as fused than are nonharmonic sounds (see Note 2). Stringed and blown instruments have partials that are harmonic or nearly harmonic, and such partials unite to produce the impression of a single tone. In contrast, bells and gongs have partials that are nonharmonic, and these produce more diffuse sound impressions (Mathews & Pierce, 1980). De Boer (1976) has shown that harmonic complexes tend to produce unitary and unequivocal pitch sensations, whereas certain types of nonharmonic complex do not merge, but instead produce multiple pitch sensations. Since most forced vibration systems such as the voice have partials whose frequencies are harmonic or close to harmonic, such findings are as expected on the hypothesis that our auditory system has evolved to interpret sound patterns in terms of the sources from which they emanate.

We may next enquire whether the phase relations between the partials of a tone affect the fusion of its image. This question was investigated by Kubovy (Note 3). He created a set of harmonically related sinusoids, all of equal amplitude, and all beginning with a positive zero-crossing and therefore having a common zero-crossing at the frequency of the fundamental. One of these sinusoids was then moved out of phase for a few hundred milliseconds. It was then moved back into phase, while another was moved out, and so on. A perceptual segregation was produced by these means, so that a melody was heard that corresponded to the out-of-phase sinusoids.

Later, Kubovy and Jordan (1979) constructed stimuli consisting of the third to fourteenth harmonics of a 200-Hz fun-

damental, which were played in the sine phase. At intervals of roughly 300 msec, the phases of all components but one were reset to 0°, and the phase of the remaining component was set to a different phase angle. The out-of-phase components formed a scale that either ascended or descended, and subjects judged the direction of this scale. The results are shown in Figure 32.1. It can be seen that for phase shifts greater than 40° subjects showed near-perfect identification of scale direction. These experiments therefore demonstrate the perceptual effect of phase relationships on the fusion of single tones composed of harmonically related complexes: Phase shifting a component of the complex results in its perceptual segregation.

Tones whose fundamental frequencies are related by simple ratios fuse more readily than tones that are not so related. In a demonstration of this phenomenon, Rasch (1978) presented two chords in succession. The lower tones of each chord were identical, and the higher tones formed a sequence that either ascended or descended. The subjects' task was to judge whether the higher tones formed a "low-high" sequence or a "high-low" sequence. Detection thresholds were taken as the measure of the extent to which the subjects could separate out the component tones of each chord. The lower tones all had a fundamental frequency of 250 Hz. The higher tones had fundamental frequencies that either were 500 and 750 Hz or deviated slightly from these values.

These results of the experiment are shown in Figure 32.2. It can be seen that, as the relationships formed by the fundamental frequencies of the higher and lower tones deviated from simple ratios, detection performance gradually improved, indicating a decreased tendency to fuse together the higher and lower components of the chords.

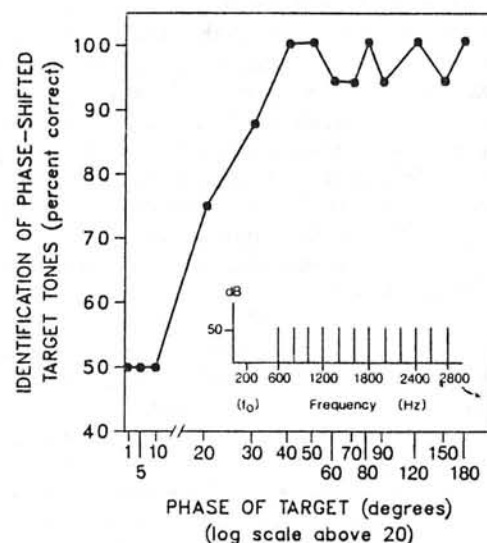


Figure 32.1. Percentage of correct identification of phase-shifted target tones as a function of phase shift in degrees. The stimuli consisted of the third to fourteenth harmonics of a 200-Hz fundamental, which were played in the sine phase. At intervals of around 300 msec, the phases of all components except one were reset to 0°, and the phase of the remaining component was set to a different phase angle. The out-of-phase components formed a scale that either ascended or descended, and subjects identified the direction of the scale. Near-perfect identification was shown for phase shifts greater than 40 deg. (From M. Kubovy & R. Jordan, Tone-segregation by phase: On the phase sensitivity of the single ear, *Journal of the Acoustical Society of America*, 1979, 66. Reprinted by permission.)

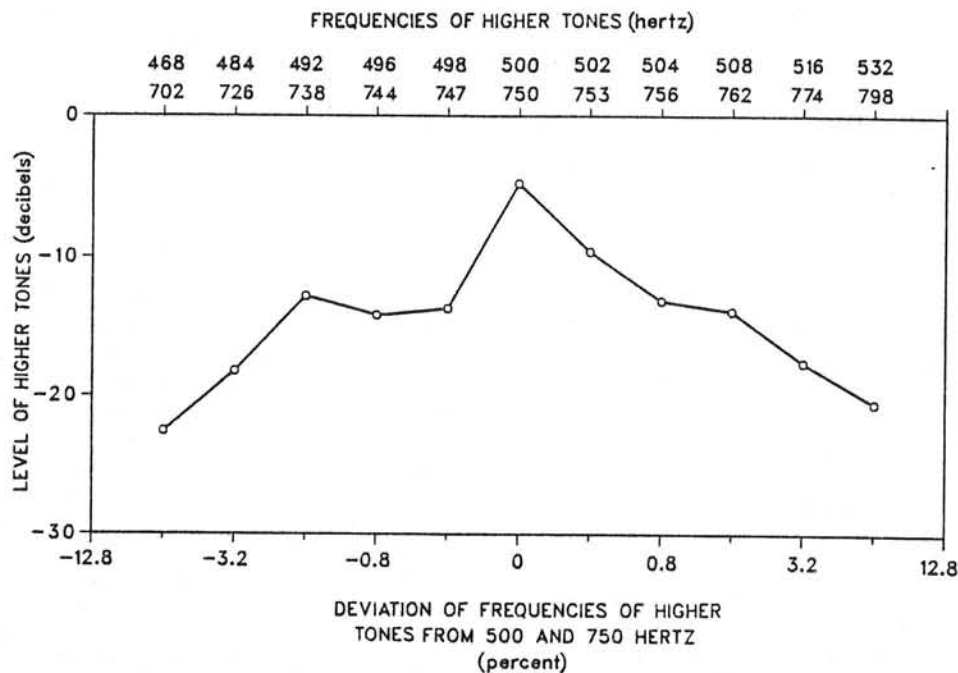


Figure 32.2. Detection thresholds for higher tones in the presence of lower tones. Two chords were presented in succession. The lower tones of the chords were both at 250 Hz, and the higher tones formed either a "low-high" sequence or a "high-low" sequence. Either higher tones were at 500 and 750 Hz, or they deviated slightly from these values. Subjects judged whether a "low-high" sequence or a "high-low" sequence had been presented. Detection thresholds fell gradually with increasing deviation from the 500-Hz and 750-Hz values, in roughly symmetrical fashion.

1.1.2. Time-Variant Relationships. One factor that may be hypothesized to contribute to the impression of a single fused sound is coordinated modulation in the steady state. In forced-vibration systems, any perturbation of the driving force will result in perturbations of components of the spectrum that are proportional to their frequencies. Thus a complex of sinusoids that is modulated in correlation is likely to be emanating from a single source. McNabb and Chowning (quoted by McAdams, 1982) have demonstrated informally that a harmonic tone complex with a spectral power distribution conforming to that of a vowel produces only a weak vocal sensation, and only weak perceptual fusion. However, if a small amount of frequency modulation is superimposed on all the spectral components simultaneously, they sound strongly fused. Similar observations have been reported informally by McAdams (1982).

By the same token, if we hear a complex of sinusoids with uncorrelated modulation functions, the likelihood is that the components of the complex are emanating from different sources. McAdams (1982) reports an informal experiment employing a complex stimulus in which a transition was made from perfectly correlated to two uncorrelated frequency modulation functions. For harmonic tone complexes, the listener's percept shifted from a single fused image to two distinct images. The effect was uncertain for inharmonic tone complexes.

A related finding was obtained by Rasch (1978), using the sequence detection task described in Section 1.1.1. He showed that, when the higher tones of the chords were frequency modulated while the lower tones remained unmodulated, detection of whether the chords formed a "low-high" sequence or a "high-low" sequence was enhanced, so that uncorrelated modulation

resulted in decreased fusion of the simultaneously presented tones.

How does onset asynchrony of two simultaneous tones affect perceptual fusion? Rasch (1978) used the same detection task to study the effect of delaying the lower tones of the chords relative to the higher tones. As shown in Figure 32.3, detection performance was strongly influenced by this manipulation. Each 10 msec of delay was associated with roughly a 10-dB downward shift of threshold. For a delay of 30 msec, threshold for perception of the high tone was close to that for the high tone presented alone.

Rasch further noted that the phenomenological effect of asynchrony was very strong. Whereas in the synchronous conditions a single "sound object" was perceived, in the asynchronous conditions the two tones stood apart very clearly. However, the onsets of the two tones were not separately audible, so that they were perceived as two separate but simultaneous sounds. This is an example of the continuity effect. (See Section 1.3.11.)

A related finding was obtained by Bregman and Pinker (1978). These authors presented a two-tone complex in alternation with a third tone and introduced various conditions of onset-offset asynchrony between the simultaneous tones in the complex. As the degree of asynchrony increased, the likelihood also increased that one of the simultaneous tones would form a melodic stream with the third tone. Bregman and Pinker argued that the asynchrony of the simultaneous tones resulted in a decreased tendency for these tones to be treated as coming from the same source and so facilitated a sequential organization by frequency proximity between one of these simultaneous tones and the alternating tone.

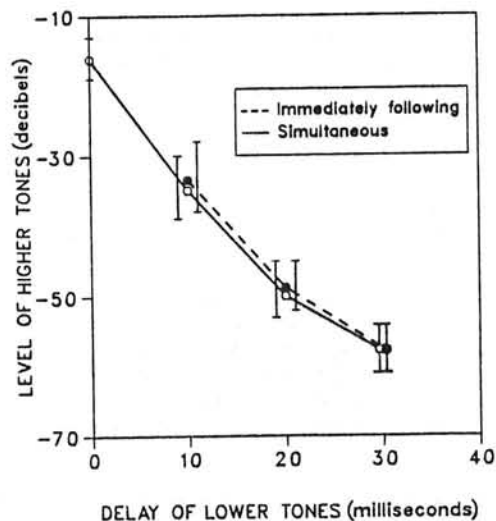


Figure 32.3. Detection thresholds for higher tones in the presence of lower tones. The paradigm used was as described in Figure 32.2. The lower tones were at 250 Hz, and the higher tones were at 500 Hz and 750 Hz. Either the higher tones ended simultaneously with the lower tones (solid line), or they ended immediately following onset of the lower tones (dashed line). Thresholds were virtually unaffected by amount of overlap but were strongly affected by delay of the lower tones. Each 10 msec of delay produced roughly a 10-dB downward shift in threshold. (From R. A. Rasch, *The perception of simultaneous notes such as in polyphonic music*, *Acustica*, 1978, 40. Reprinted with permission.)

Dannenbring and Bregman (1978) investigated the effects of several variables on the tendency of one component of a complex tone either to fuse with the other components or alternatively to be pulled out into a different melodic stream. The stimuli consisted of a complex of three pure tones (at 500, 1000, and 2000 Hz) that alternated repeatedly with a single "captor" pure tone (at 500, 1000, or 2000 Hz). The amplitudes of the components of the complex tone either were equal or increased or decreased with frequency. The amplitude of the "captor" tone was always equal to that of the "target" component of the frequency with which it alternated. The relative onsets and offsets of the components of the complex tone were also varied. Subjects judged the repetition rate of the captor tone. If this rate was judged to be slow, the components of the complex tone were considered to be fused into a single unit. However, if this rate was judged to be fast, the target component of the complex tone was considered to have been pulled into the same stream as the captor.

Various findings emerged from this study. First, the tendency for the formation of melodic streams was found to be greater when the repeating tone was at 500 Hz than when the tone was at one of the other two frequencies. Second, the tendency to fusion was greatest for tones in which the relative amplitudes of the components decreased with frequency, a situation most like that commonly encountered in the natural environment. Third, when the target components led the other components of the complex tone at onset, there was an increased tendency to produce melodic streams. This was also true when the target component lagged the other components at offset. However, when the target component lagged the others at onset or led them at offset, no such effect occurred.

The effects of fusion and separation of two gliding tones were studied by Steiger and Bregman (1981). Here the tones

glided in parallel on a log frequency scale, and the glides were repeatedly presented in alternation with a pure tone "captor" glide. Subjects judged whether the stimulus was "fused" (i.e., whether the sequence appeared as an isochronous alternation of a pure tone with a rich tone) or "decomposed" (i.e., whether the sequence appeared to contain three tones in each cycle). The tendency for the stimulus to be judged as decomposed was enhanced when the captor and target glides were in the same frequency range, and also when the captor and target glides had the same orientation.

A sudden change in the amplitude of a component of a tone complex can cause this component to stand out perceptually. This was demonstrated by Kubovy (Note 4). He presented subjects with an eight-tone chord whose components were successively turned off abruptly for 80 msec and then restored to full amplitude. This manipulation occurred at a rate of three per second. The subjects perceived a melody that corresponded to the order in which the tones were subjected to this momentary amplitude disparity. For this pitch segregation effect to occur, it was necessary that the frequency spacing between successive tones be greater than the critical band.

1.1.3. Familiarity. Sounds with familiar spectral shapes, such as human voices and musical instrument tones, appear to fuse more readily than sounds with unfamiliar spectral shapes. Informal observations show that the percept of a particular vowel is lost when its spectral envelope is shifted slightly in frequency, even though the relative amplitudes are preserved. Other factors such as the relative growth and decay of individual partials also appear to contribute to familiarity. Unfortunately no quantitative data on the issue are available at present.

1.2. Grouping of Sound Sequences in Space

A useful technique for studying grouping phenomena in hearing is to present two different pitch sequences in parallel, one to the left of the listener, and the other to the right. In most experiments, stimuli have been presented dichotically via headphones; however, in some experiments stimuli have been presented via spatially separated loudspeakers. This technique enables different stimulus dimensions to be set in opposition to each other as bases for grouping. Thus, for example, grouping by frequency or by amplitude may be opposed to grouping by spatial location. At the same time, different principles governing grouping along any given dimension may be set in opposition to each other. For example, grouping by proximity may be opposed to grouping by good continuation. This section describes findings obtained with this technique and discusses their theoretical implications.

1.2.1. Auditory Illusory Conjunctions. When two sequences of tones emanate simultaneously from different regions of space, and the onsets and offsets of these tones are synchronous, striking perceptual illusions are generally produced. We may characterize a tonal stimulus as a bundle of attribute values, that is, as having a pitch, a location, a loudness, and a timbre. In the situation just outlined, these bundles of attribute values fragment and recombine, so that illusory conjunctions result. (See also Treisman, Chapter 35.) This anomalous recombination suggests that all auditory stimuli are at some stage in the processing system fragmented into their separate attributes, and that this process of fragmentation is followed by a process of perceptual synthesis in which the different attribute values

are recombined. Under most circumstances the stimuli are reconstructed correctly; however, we should not assume that this necessarily occurs.

Striking individual differences are manifest in the types of illusion that are produced in this situation. Further, these differences correlate strongly with handedness and may be related to patterns of cerebral dominance. This implies that they have an innate basis.

1.2.2. The Scale Illusion. One example of the creation of strong illusory conjunctions is provided in the scale illusion (D. Deutsch, 1975c, 1975e). The configuration that produced the illusion is illustrated in Figure 32.4(a). It can be seen that this consisted of a major scale (see Note 5), which was presented simultaneously in both ascending and descending form. When a tone from the ascending scale was delivered to one ear, a tone from the descending scale was simultaneously delivered to the other ear, and successive tones in each scale alternated from ear to ear. This pattern was repeatedly presented ten times without pause. All tones were sine waves of equal amplitude and 250 msec in duration.

When presented with this configuration, no subject perceived the sequence of tones that was delivered to one ear or to the other, and none perceived a full ascending or descending scale. Instead, the successive tones were always grouped together on the basis of frequency range. All subjects perceived a sequence of four tones that repeatedly descended and then ascended. Beyond this, percepts were divisible into two categories. Most subjects also perceived a second stream of lower tones that repeatedly ascended and then descended. The second stream moved in contrary motion to the first [Figure 32.4(b)]. This percept therefore included all the pitches in the configuration; however, these were separated into two streams on the basis of frequency range. A minority of subjects perceived instead

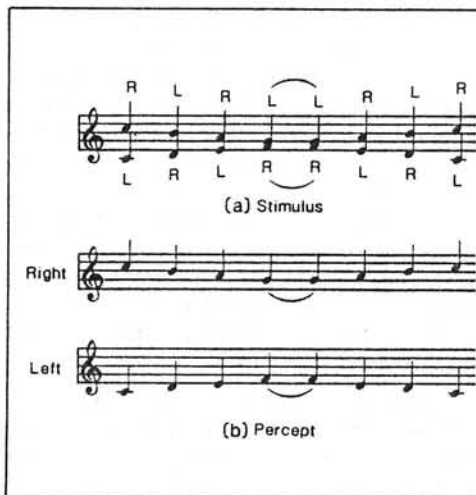


Figure 32.4. (a) Stimulus configuration that produced the scale illusion. This consisted of a major scale, presented simultaneously in both ascending and descending form. When a tone from the ascending scale was delivered to one ear, a tone from the descending scale was simultaneously delivered to the other ear, and successive tones in each scale alternated from ear to ear. All tones were of equal amplitude and 250 msec in duration. There were no pauses between tones. (b) Percept most commonly obtained. This consisted of two melodic lines, a higher one and a lower one, that moved in contrary motion. The higher tones all appeared to be emanating from one earphone, and the lower tones from the other earphone. (From D. Deutsch, Two-channel listening to musical scales, *Journal of the Acoustical Society of America*, 1975, 57. Reprinted with permission.)

Table 32.1. Numbers of Right-Handers and Left-Handers Perceiving Both the Higher and the Lower Pitch Sequences in the Scale Illusion ("Both"), and Those Perceiving Only the Higher Pitches ("Single")

Handedness	Streams	
	Both	Single
Right	34	7
Left	15	14

The right-handers tended significantly to hear both streams; however, the left-handers did not show such a tendency (From D. Deutsch, Two-channel listening to musical scales, Journal of the Acoustical Society of America, 1975, 57. Reprinted with permission.)

only one stream of four tones that repeatedly descended and then ascended. This corresponded to the higher sequence of tones; little or nothing of the lower sequence was perceived.

Table 32.1 shows the numbers of right-handed and left-handed subjects who obtained these two categories of percept. As can be seen, the two handedness groups differed significantly on this measure. Further, in considering those subjects who perceived both streams, significant differences between the two handedness groups also emerged. Most right-handers obtained an illusion whereby the higher tones all appeared localized in one ear and the lower tones in the other ear. As shown in Table 32.2, there was a highly significant tendency to perceive the higher tones in the right ear and the lower tones in the left ear, and also to maintain a given localization pattern when the earphone positions were reversed. The remaining right-handers obtained a variety of idiosyncratic localization percepts, as did those who perceived only one stream. Most left-handed subjects who perceived both streams also localized all the higher tones in one ear and all the lower tones in the other ear. However, as shown in Table 32.2, these subjects did not display the same localization tendencies as did the right-handers. The remaining left-handed subjects reported a variety of idiosyncratic localization percepts.

Table 32.2. Localization Patterns in the Scale Illusion, Displayed for Those Subjects who Perceived All the Higher Tones in One Ear and All the Lower Tones in the Other Ear

Handedness	Localization		
	RR	LL	Both
Right	21	1	8
Left	2	5	4

Numbers of right-handed and left-handed subjects obtaining a given localization pattern are displayed. Subjects listened to the sequence with earphones placed first one way and then the other. RR: Higher tones localized in the right ear and lower tones in the left ear on both presentations. LL: Higher tones localized in the left ear and lower tones in the right ear on both presentations. Both: Higher tones localized in the right ear and lower tones in the left ear on one presentation, with opposite localization pattern on the other. Right-handers tended strongly to hear the higher tones on the right and the lower tones on the left; however, left-handers did not show this tendency.

AUDITORY PATTERN RECOGNITION

To summarize these findings, in considering what *attribute* was used as a basis for grouping, organization by spatial location never occurred; rather organization was always on the basis of frequency (see also Kubovy, 1981). Second, in considering which *principle* was used, organization was always on the basis of frequency proximity. Either listeners heard two melodic lines, one corresponding to the higher tones and the other to the lower tones, or they heard the higher tones alone. Third, there were substantial individual differences in the way that this configuration was perceived, both in terms of *what* was perceived and in terms of *where* the sounds appeared to be coming from. These individual differences correlated strongly with handedness.

Auditory illusory conjunctions have been shown to occur under broader circumstances also. Butler (1979) presented the scale illusion pattern either through headphones, or through spatially separated loudspeakers in a free sound field environment. The subjects, who were musically trained, notated separately the sequence coming from the speaker on their right and the sequence coming from the speaker on their left. Almost all notations reflected grouping by frequency range. Thus a higher melodic line appeared to be coming from one speaker, and a lower melodic line appeared to be coming from the other. Even when the tones were generated on a piano, and when in addition differences in timbre and loudness were introduced between the tones coming from the two speakers, essentially the same results were obtained. Similar results were also obtained when different contrapuntal patterns (see Note 6) were employed. Further, when differences in timbre were introduced, no subject was able to identify these differences correctly. Instead, all subjects perceived a change in tone quality that appeared to be emanating from both headphones or both loudspeakers. Butler concluded that under these conditions channeling by pitch range is so strong as to persist through a wide range of timbral changes, changes in envelope characteristics of the tones, imperfectly timed attacks, inconsistencies in duration, and varying loudness.

This powerful illusion appears as a good example of unconscious inference in perception. Our auditory environment is very complex, and the assignment of sounds to their sources is rendered difficult by the presence of echoes and reverberation. So when a sound mixture is presented such that both ears are stimulated simultaneously, we cannot judge from first-order localization cues (see Note 7) alone which components of the total spectrum should be assigned to which source. We therefore need to utilize other cues in making such judgments. One such cue is similarity of frequency spectrum. Similar sounds are likely to be coming from the same source, and different sounds from different sources. It is therefore reasonable for the listener to conclude that tones in one frequency range are coming from one source, and that tones from a different frequency range are coming from another source. The tones are therefore perceptually reorganized in space in accordance with this interpretation (D. Deutsch, 1975c).

1.2.3. Grouping of Nonsimultaneous Sound Sequences. If the above line of reasoning is correct, we should expect that perceptual grouping of parallel pitch sequences would be strongly influenced by the salience of the first-order localization cues. If, in contrast to the conditions just described, such cues were strong and unambiguous, channeling by spatial location would be expected to take precedence over channeling by frequency range. One can produce such a situation by employing sequences in which the tones at the two ears are clearly separated in time.

To examine this hypothesis, perceptual grouping was examined as a function of the temporal relationships between the signals arriving at the two ears (D. Deutsch, 1979a). Subjects were asked to identify rapid melodic patterns whose component tones switched from ear to ear. In one set of conditions, input was to one ear at a time; in another set, input was to both ears simultaneously. It was predicted that when input was to one ear at a time identification of the melody should be difficult, reflecting perceptual grouping by spatial location. However, when both ears receive input simultaneously, identification of the melody should be much easier.

Subjects were presented with sequences of pure tones. Each sequence consisted of ten repetitions of a basic eight-tone melody. All tones were of equal amplitude and 30 msec in duration, with tones within a melody separated by 100-msec pauses. Two such melodies were employed, and the subjects identified on each trial which of these had been presented.

The experiment employed four conditions, which are illustrated in Figure 32.5. In Condition A, all tones of the melody were presented simultaneously to both ears. In Condition B, the component tones of the melody were distributed in random fashion between the ears. Condition C was identical to Condition B except that the melody was accompanied by a drone. Whenever a tone from the melody was presented to the right ear, the drone was simultaneously presented to the left ear, and vice versa. Condition D was identical to Condition C except that the drone was always presented to the same ear as the tone from the melody.

The percentages of correct identifications of the melodies in the different conditions of the experiment are shown on Figure 32.5. It can be seen that excellent performance was obtained in Condition A, in which the melodies were presented binaurally. In contrast, performance in Condition B, in which the tones from the melodies were distributed between the ears, was very poor. The procedure of switching the tones from ear to ear thus produced a considerable decrement in identification performance. However, in Condition C, in which a contralateral drone was presented so that input was to both ears simultaneously, the performance level was again very high. This finding cannot be attributed to processing the harmonic relationships between the drone and the melody because in Condition D, in which the drone was presented to the same ear as the melody component, performance was below chance. In this last condition, input was no longer to the two ears simultaneously.

This experiment demonstrates that temporal relationships between tones emanating from different spatial locations are important factors in determining how the tones are perceptually grouped. When signals are emanating from two locations simultaneously, as in Condition A and C, it is easy to integrate the information arriving at the two ears into a single perceptual stream. However, when the signals coming from the two locations are clearly separated in time, as in Conditions B and D, grouping by spatial location is so powerful as to prevent the listener from combining the tones to produce an integrated percept.

We may next enquire what happens in the intermediate case, where inputs to the two ears overlap but are not strictly synchronous. This condition brings us closer to normal listening, and also to the case where streams of speech are presented in parallel to both ears. A second experiment investigated the effects of onset-offset asynchrony between the components of the melody and the contralateral drone. In the asynchronous conditions, all tones were again 30 msec in duration, and the drone either led or lagged the melody components by 15 msec,

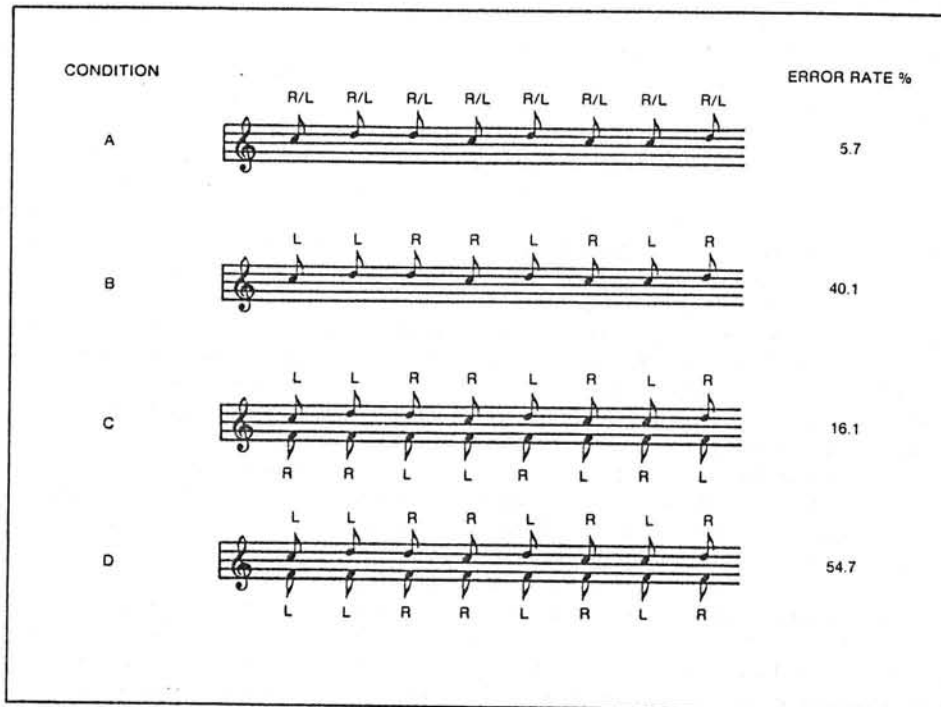


Figure 32.5. Percentage of errors in identification of melodic patterns when the component tones of the patterns switched between ears. On each trial, ten repetitions of a basic eight-tone pattern were presented. All tones were 30 msec in duration, and tones within a pattern were separated by 100-msec pauses. Two such melodies were employed, and subjects identified on each trial which of these had been presented. In Condition A (melody presented binaurally) excellent performance was obtained. In Condition B (melody distributed between ears) performance was very poor. In Condition C (contralateral drone accompanying melody) performance levels were again high. In Condition D (ipsilateral drone accompanying melody) performance was below chance. (From D. Deutsch, *Binaural integration of melodic patterns*, *Perception and Psychophysics*, 1979, 25. Reprinted with permission.)

or the right ear tones led or lagged the left ear tones by 15 msec. Performance levels in these conditions were significantly lower than when the melody components and the drone were strictly synchronous, and they were also significantly higher than when the melody components switched between ears without an accompanying drone. This is as expected on the present line of reasoning.

A similar experiment was performed by Judd (1979). Two repeating stimulus patterns were constructed, from four square-wave tones, each 100 msec in duration. The two patterns were as shown on Figure 32.6. It can be seen that, taking each channel separately and treating the patterns as cyclically repeating, the tones in the two patterns were identically ordered. However, when the channels were combined, two different melodic patterns emerged instead. Subjects were presented with pairs of these patterns and were required to judge whether the members of each pair were the same or different. On half of the trials, the silent gaps between the tones were replaced by noise. It was found that performance was better in the noise-filler condition than in the silent gap condition. Judd interpreted this finding as due to the noise degrading the localization information, which encouraged grouping of successive tones on the basis of frequency range rather than spatial location.

Schubert and Parker (1956) performed an experiment that may be interpreted similarly. These authors measured the amount of interference in speech perception that was produced by switching the signal from ear to ear. They found that adding

noise to the contralateral ear reduced this interference effect (Figure 32.7). It may plausibly be argued that the ongoing speech-noise signal was interpreted by the listener in terms of two sources, one emitting noise and the other emitting speech, whereas the ongoing speech-silence signal was interpreted by the listener in terms of two independent speech sources.

1.2.4. The Hypothesis of a Slow Switching Mechanism. The problem of degradation of processing when information is switched from ear to ear has been addressed in other contexts. For instance, Cherry and Taylor (1954) studied the intelligibility of speech that switched back and forth from ear to ear. They found that intelligibility dropped substantially at alternation rates of around 3 Hz and interpreted these findings in terms of a limitation in the rate at which we are able to switch attention between ears. However, Huggins (1964) found that the maximum dip in intelligibility shifted in parallel with a shift in the rate of the presented speech. He argued from this result that the performance decrement was due to interference in processing the basic units of speech, and not to a limitation in attention switching time.

A related paradigm involves recall of lists of digits that are dichotically presented. When two such dichotic lists were delivered at fast rates, recall was found to be better by ear than by temporal order, the latter task requiring switching between ears (Broadbent, 1954, 1958). Further, subjects showed poorer recall of successive lists of digits when these were presented

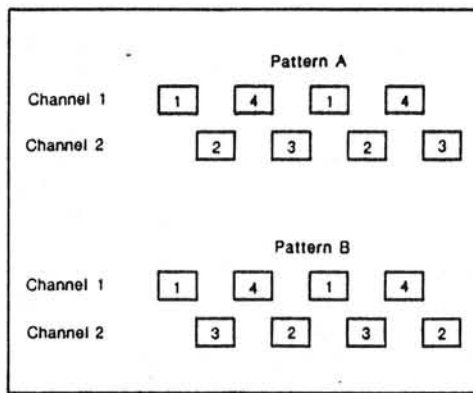


Figure 32.6. Stimulus configurations employed to investigate the effect of contralateral noise on the ability to discriminate melodic patterns whose component tones alternated between ears. Tones were 100 msec in duration, with fundamental frequencies of (1) 912 Hz, (2) 1024 Hz, (3) 1150 Hz, and (4) 1290 Hz. Discrimination performance was enhanced when the gaps between the tones were replaced by noise. (From T. Judd, *Comments on Deutsch's musical scale illusion, Perception and Psychophysics*, 1979, 26. Reprinted with permission.)

alternately to the two ears than when they were presented binaurally (A. Treisman, 1971). This finding cannot be ascribed to perceptual interference with the basic units of speech, since there was no disruption of the verbal items in these experiments. Some difficulty in the ability to switch attention between the ears was therefore hypothesized.

In contrast to the above arguments for a switching limitation, powerful general arguments may be made against the idea that information from the two ears cannot be dealt with in rapid succession. In everyday listening, the information ar-

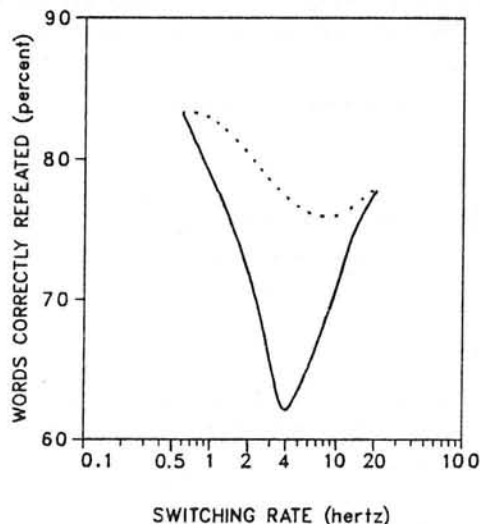


Figure 32.7. Percentages of words correctly repeated as a function of rate at which the speech signal was switched from ear to ear. The lower curve shows the results for trials with silence in the contralateral ear. The upper (dotted) curve shows the results for trials in which noise was delivered to the contralateral ear. The contralateral noise resulted in enhanced speech intelligibility, especially at switching rates of around 4 Hz, where intelligibility was otherwise substantially reduced. (From E. D. Schubert & C. D. Parker, Addition to Cherry's findings on switching speech between two ears, *Journal of the Acoustical Society of America*, 1956, 27. Reprinted with permission.)

riving at the two ears is never identical, and the running cross-correlations performed on this information are very important for several functions. One such function is localization, and the other is the suppression of echoes and reverberation (Haas, 1951; Tobias & Schubert, 1959; Wallach, Newman, & Rosenzweig, 1949). The auditory elements that are compared for such functions may be separated by only a few microseconds. Such an ability to utilize information entering the two ears in rapid succession is not consistent with the notion of a slow switching mechanism.

Two conflicting sets of phenomena have therefore been reported, one arguing for a decrement in processing information where rapid switching between ears is involved, and the other arguing against such a decrement. We may resolve this conflict on the following line of reasoning. An important function of our auditory system is to separate out the signals emanating from different sources. If such perceptual separations were not accomplished we would not know which elements of the acoustic spectrum to link with, so as to form high-order abstractions. It is necessary, therefore, that there exist mechanisms that inhibit the formation of higher-order linkages between acoustic elements that are likely to be emanating from different sources. Since our acoustic environment is very complex, such mechanisms must be flexible and employ multiple criteria. Thus certain configurations involving input to the two ears would be interpreted as coming from the same source, so that integration of this information should be easy. Yet other configurations would best be interpreted as emanating from different sources, so that integration should be difficult. According to this hypothesis, when a decrement in integrating information arriving at the ears occurs, this is due not to capacity limitation, but rather to a mechanism that we have evolved to prevent confusion in monitoring our auditory environment (see Bregman, 1978, 1981, for an analogous argument based on findings involving various monaural tasks).

1.2.5. The Octave Illusion. In the experiments described in Section 1.2.2, when tones were presented to both ears simultaneously with synchronous onsets and offsets, sequential grouping by frequency proximity was the rule. Grouping by ear of input occurred only when there were temporal separations between the stimuli presented to the two ears. We now turn to an examination of certain situations in which grouping by ear of input occurs even though such input is strictly simultaneous. It will be seen that this happens only under special conditions of frequency relationship between the tones presented in sequence at the two ears.

One such situation is illustrated in Figure 32.8(a). This shows the stimulus pattern that gives rise to the octave illusion (D. Deutsch, 1974, 1975c). It can be seen that two tones that were spaced an octave apart (400 and 800 Hz) were repeatedly presented in alternation. The identical sequence was delivered to the two ears simultaneously; however, when the right ear received the high tone the left ear received the low tone and vice versa. So in fact the listener was presented with a single, continuous, two-tone chord, but the ear of input for each component switched repeatedly.

This configuration produced a number of illusory percepts, the most common of which is illustrated in Figure 32.8(b). It can be seen that this consisted of a single tone that alternated from ear to ear, and whose pitch simultaneously alternated from one octave to another in synchrony with the localization shift. When the earphones were placed in reverse position, most

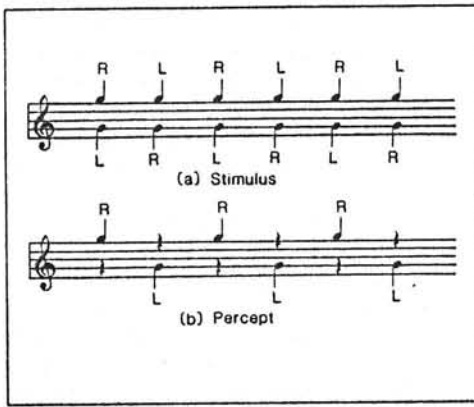


Figure 32.8. (a) Stimulus pattern giving rise to the octave illusion. Musical notation is approximate. The lower tones were at 400 Hz and the higher tones were at 800 Hz. All tones were of equal amplitude and 250 msec in duration. There were no gaps between tones. The sequence was continuously presented for 20 sec. (b) Percept most commonly obtained. This consisted of a single tone that alternated from ear to ear and whose pitch simultaneously alternated from one octave to the other in synchrony with the localization shift. (From D. Deutsch, *An auditory illusion*, *Nature*, 252. Copyright 1974 by Macmillan Journals Ltd. Reprinted with permission.)

listeners found that the apparent locations of the high and low tones remained fixed. Thus it seemed to these listeners that the earphone that had been producing the high tones was now producing the low tones, and that the earphone that had been producing the low tones was now producing the high tones.

If we assume that there are two separate brain mechanisms, one for determining what pitch we hear and the other for determining where the sound is located, we are in a position to advance an explanation for this illusion. The model is diagrammed in Figure 32.9. To determine the perceived pitch, the information arriving at one ear is followed, and the information

arriving at the other ear is suppressed. However, each tone is localized in the ear receiving the higher-frequency signal, regardless of which frequency is in fact perceived (D. Deutsch, 1975c). The combined output of these two mechanisms, for the case of the listener whose pitch percept corresponds to the frequencies presented to the right ear, should result in the percept of a high tone to the right alternating with a low tone to the left. For the case of the listener whose pitch percept corresponds to the frequencies presented to the left ear instead, the resultant percept should be that of a high tone to the left alternating with a low tone to the right.

This model received confirmation in a further experiment (D. Deutsch & Roll, 1976). Subjects were presented with the basic pattern shown in Figure 32.10(a). This again employed tones standing in octave relation. It can be seen that one ear received three high tones followed by two low tones, while simultaneously the other ear received three low tones followed by two high tones. This basic pattern was repeatedly presented ten times without pause.

As expected from the model, most subjects perceived a pattern of pitches that corresponded to the frequencies presented either to the right ear or to the left ear. In other words, they heard a repeating sequence consisting either of three high tones followed by two low tones, or of three low tones followed by two high tones. However, each tone was localized in the ear that received the higher frequency. This is illustrated in Figure 32.10(b). When Channel A was presented to the right ear and Channel B to the left, the listener heard a repeating sequence of three high tones to the right followed by two low tones to the left. When, however, Channel A was presented to the left ear and Channel B to the right, the listener now heard a repeating sequence of two high tones to the right followed by three low tones to the left.

Most subjects in the D. Deutsch (1974) experiment perceived a single high tone in one ear alternating with a single low tone in the other ear. However, some subjects instead perceived a

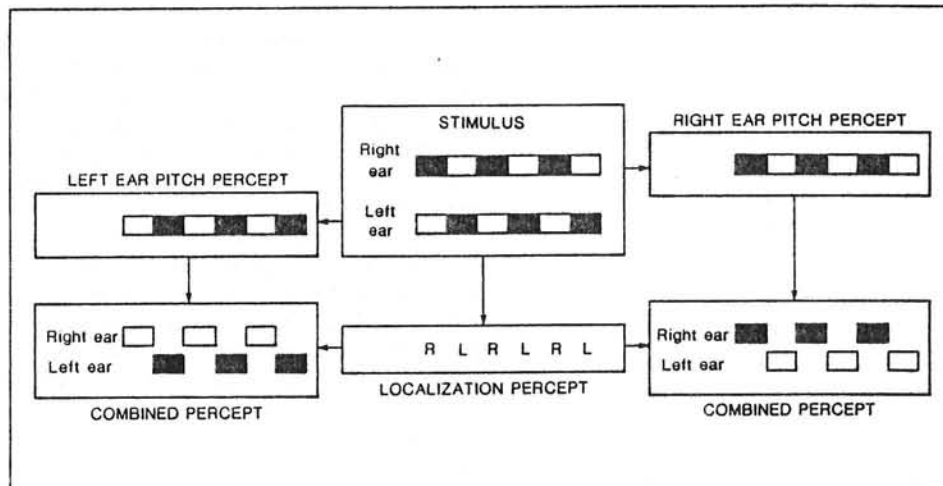


Figure 32.9. Diagram showing how the outputs of the pitch and localization mechanisms combine to produce the octave illusion. Filled boxes indicate high tones (800 Hz) and unfilled boxes indicate low tones (400 Hz). The pitch mechanism follows the sequence of frequencies presented to one (dominant) ear rather than to the other. However, the localization mechanism follows the higher-frequency signal, regardless of whether the higher or the lower frequency is perceived. The outputs of these two mechanisms combine to produce the percept of a high tone in one ear alternating with a low tone in the other ear. (From D. Deutsch, *The octave illusion and auditory perceptual integration*, in J. V. Tobias & E. D. Schubert (Eds.), *Hearing research and theory*, Vol. 1, Academic Press, Inc., 1981. Reprinted with permission.)

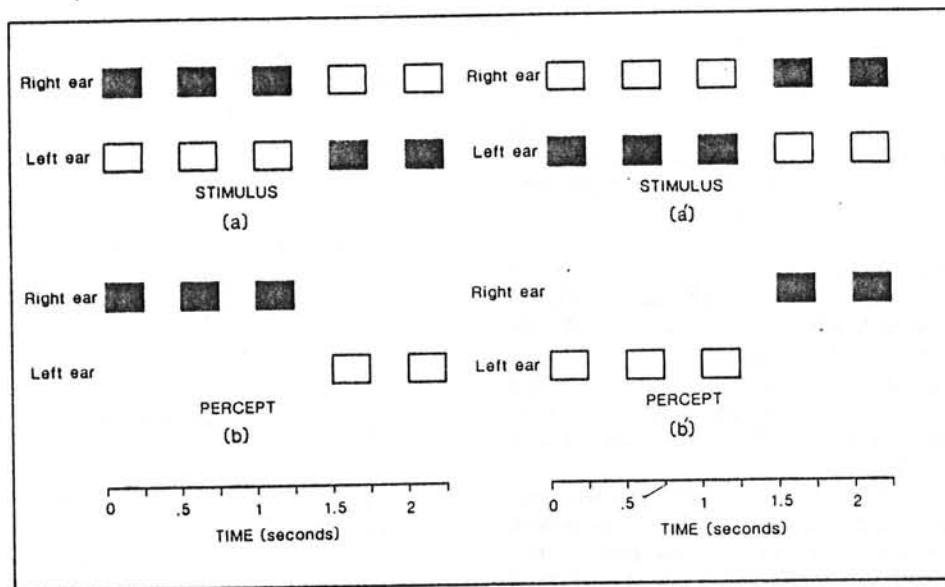


Figure 32.10. Stimulus patterns and percepts in experiment to test hypothesized basis for the octave illusion. Filled boxes represent tones of 800 Hz and unfilled boxes represent tones of 400 Hz. The basic patterns shown were presented ten times without pause. In accordance with the hypothesis, most subjects reported the pattern of pitches that was presented to the right ear; yet all subjects localized each tone to the ear receiving the higher-frequency signal. (From D. Deutsch & P. L. Roll, Separate 'what' and 'where' decision mechanisms in processing a dichotic tonal sequence, *Journal of Experimental Psychology: Human Perception and Performance*, 2. Copyright 1976 by American Psychological Association. Reprinted with permission.)

single tone that alternated from ear to ear, whose pitch either did not change or changed only slightly with a shift in its apparent location. Other subjects heard more complex patterns, such as two low tones that alternated from ear to ear with an intermittent high tone in one ear. Such patterns were usually unstable, exhibiting frequent changes with continued listening.

The individual differences in perception of this illusion were found to correlate with handedness. As shown in Table 32.3, the proportion of subjects reporting complex percepts was substantially higher in the left-handed than in the right-handed population (see also Craig, 1979). A second handedness correlate concerned the localization patterns for the high and low tones. As shown in Table 32.4, most right-handers heard the high tone on the right and the low tone on the left, regardless of the positions of the earphones (see also Geffen & Reynolds, 1982; McClurkin & Hall, 1981). In contrast, the left-handers did not show a significant tendency to localize the high and low tones

Table 32.3. Distributions of Right-Handers and Left-Handers Obtaining Different Categories of Percept in the Octave Illusion

Handedness	Octave	Single Pitch	Complex
Right (n = 53)	58%	25%	17%
Left (n = 33)	52%	9%	39%

Percentages of right-handers and left-handers are displayed. "Octave" indicates the percept of a single tone that alternates from ear to ear, whose pitch simultaneously alternates from one octave to the other. "Single Pitch" indicates the percept of a single tone that alternates from ear to ear, whose pitch either does not change or shifts slightly with a change in localization. "Complex" comprises a number of different complex percepts. The proportion of subjects obtaining complex percepts was considerably higher among left-handers than among right-handers. (From D. Deutsch, *An auditory illusion*, *Nature*, 151. Copyright 1974 by Macmillan Journals Ltd. Reprinted with permission.)

Table 32.4. Localization Patterns for the Two Pitches at the Two Ears in Subjects Obtaining the "Octave" Percept in the Octave Illusion

Handedness	RR	LL	Both
Right	25	5	1
Left	6	7	4

Each subject was given two presentations of the sequence, for 20 sec each time, with earphones placed first one way and then the other. The numbers of right-handers and left-handers obtaining a given localization pattern are displayed. RR: High tone localized in the right ear and low tone in the left on both presentations. LL: High tone localized in the left ear and low tone in the right on both presentations. Both: High tone localized in the right ear and low tone in the left on one presentation; and high tone localized in the left ear and low tone in the right on the other. Right-handers tended strongly to hear the high tone in the right and the low tone in the left; however, left-handers did not display this tendency.

either way, and showed a greater tendency to change their localization patterns.

Given the strong correlates with handedness in perception of the octave illusion, it is interesting to consider the neurological differences on which such correlates might be based. The overwhelming majority of right-handers are left-hemisphere dominant, but this is true of only about two-thirds of left-handers. Further, the majority of right-handers have a clear dominance of the left hemisphere; however, a substantial proportion of left-handers have some bilateral representation (Goodglass & Quadfasel, 1954; Hécaen & de Ajureaguerra, 1964; Hécaen & Piercy, 1956; Milner, Branch, & Rasmussen, 1966; Subirana, 1969; and Zangwill, 1960). It appears reasonable to assume

that these patterns of dominance are reflected in percepts of the octave illusion in two ways. First, the localization of the high tone on the right and the low tone on the left reflects left-hemisphere dominance, with the localization of the high tone on the left and the low tone on the right reflecting right-hemisphere dominance. Second, unambiguous localization patterns reflect clear dominance, with complex percepts reflecting more cerebral equipotentiality.

Localization patterns have been shown to correlate not only with handedness, but also with familial handedness background. In a study by D. Deutsch (1983b), subjects with left- or mixed-handed parents or siblings were found less likely to localize the high tone on the right and the low tone on the left than were subjects without left- or mixed-handed parents or siblings. This was found true for right-handed, mixed-handed, and left-handed populations.

A further question of interest is whether the interactions underlying the localization and pitch effects in the octave illusion occur between pathways conveying information from the two ears, or whether instead pathways conveying information from different regions of auditory space are involved. To investigate this question, the stimuli were presented through spatially separated loudspeakers rather than earphones (D. Deutsch, 1974, 1975c). An analogous illusion was obtained under these conditions: The subjects perceived a high tone that appeared to be coming from one speaker, which alternated with a low tone that appeared to be coming from the other speaker. This effect was obtained even with the two speakers placed side by side, facing the listener, which shows that highly specific regions of auditory space were involved here.

We shall now consider only what sequence of pitches is perceived in the octave illusion and leave aside the issue of where the tones appear to be located. In the octave illusion, channeling of pitch sequences was always on the basis of spatial location. However, in the scale illusion, channeling was always on the basis of frequency proximity instead. Yet the stimuli producing these two illusions were in several ways very similar. In both cases, repeating sequences of sine-wave tones at equal amplitudes and durations were presented, with synchronous onsets and offsets. Also in both cases, the frequencies presented to one ear always differed from the frequencies simultaneously presented to the other ear. Nevertheless, radically different channeling strategies arose in response to these two stimulus patterns. It is particularly noteworthy that, when two tones standing in octave relation were simultaneously presented in the scale illusion, both these tones were generally perceived. But when two tones standing in octave relation were simultaneously presented in the octave illusion, only one of these tones was generally perceived. Such differences in channeling strategy must therefore arise from differences in the patterns of frequency relationship between successive tones.

Another characteristic of the stimulus producing the octave illusion was that the frequency emanating from one side of space was always the same as the frequency that had just emanated from the opposite side. It therefore seemed plausible to hypothesize that this sequential relationship was responsible for producing channeling by spatial location. A further set of experiments was performed to test this hypothesis (D. Deutsch, 1980a, 1981).

In the first experiment, listeners were presented with sequences consisting of 20 dichotic chords. Two conditions were compared, using the basic patterns illustrated in Figure 32.11(a). The pattern in Condition 1 consisted of the repetitive presen-

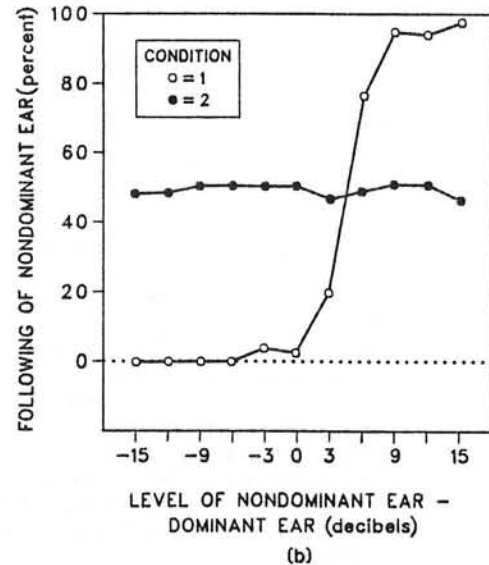
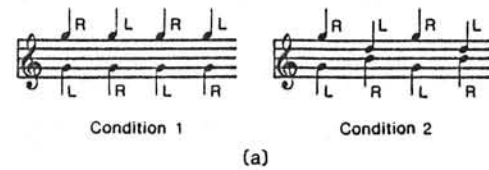


Figure 32.11. (a) Configurations used in first experiment examining effects of sequential interactions on ear dominance. Each sequence consisted of 20 dichotic chords. In Condition 1, the two ears received the same frequencies in succession; however, this was not true in Condition 2. (b) Percentage of following of nondominant ear in these two conditions, as a function of amplitude differences between the tones at two ears. In Condition 1, the dominant ear was followed until a critical level of amplitude relationship was reached, and the nondominant ear was followed beyond this level. However, there was no following on the basis of ear of input in Condition 2. (From D. Deutsch, *Ear dominance and sequential interactions*, *Journal of the Acoustical Society of America*, 1980, 67. Reprinted with permission.)

tation of a single chord. The tones comprising this chord stood in octave relation and alternated from ear to ear in such a way that when the high tone was in the right ear the low tone was in the left ear and vice versa. Here the two ears received the same frequencies in succession. The sequence presented to the right ear began with the high tone and ended with the low tone on half of the trials, while this order was reversed on the other half. The subjects were asked to judge whether the sequence began with the high tone and ended with the low tone or whether it began with the low tone and ended with the high tone. It was thus possible to infer which ear was being followed for pitch.

In Condition 2, the basic pattern consisted of the repetitive presentation of two dichotic chords in alternation. The tones comprising the first chord formed an octave and the second a minor third; thus the entire four-tone combination constituted a major triad. Note that here the two ears did not receive the same frequencies in succession. The right ear received the higher tone of the first chord and the lower tone of the last chord on half of the trials. The order was reversed on the other half of the trials.

The relationship between the amplitudes of the tones presented to the two ears was varied systematically across trials, and plots were made of the extent to which each ear was followed

as a function of these amplitude relationships. The results are displayed in Figure 32.11(b). It is evident that in Condition 1 the frequencies presented to one ear were followed until a critical level of amplitude relationship was reached, and the frequencies presented to the other ear were followed beyond this level. However, there was no following on the basis of ear of input in Condition 2, even when the signals presented to the two ears differed substantially in amplitude. Subjects instead followed on the basis of frequency proximity: Three of the subjects consistently followed the low tones, and one subject consistently followed the high tones. This result is in accordance with the assumption that channeling by spatial location here occurs when the same frequencies emanate in succession from different regions of auditory space.

In a second experiment only two dichotic chords per trial were presented. The comparison was again between two conditions. These employed the basic patterns shown in Figure 32.12(a). In Condition 1, the basic pattern consisted of two presentations of the identical chord. The component tones of this chord formed an octave, in such a way that one ear received first the high tone and then the low tone, while simultaneously the other ear received first the low tone and then the high tone.

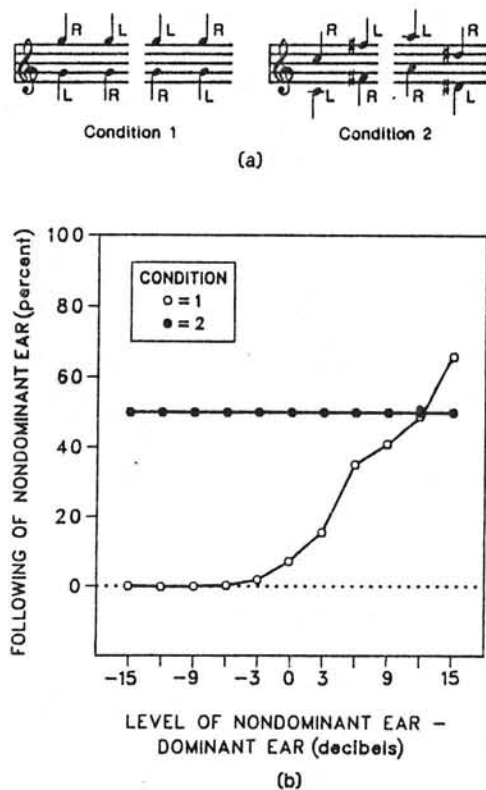


Figure 32.12. (a) Configurations used in second experiment examining effects of sequential interactions on ear dominance. Only two dichotic chords were presented on each trial. In Condition 1, the two ears received the same frequencies in succession, but this was not true in Condition 2. (b) Percentage of following of nondominant ear in these two conditions, as a function of amplitude differences between the tones at the two ears. In Condition 1, the dominant ear was followed until a critical level of amplitude relationship was reached, and the nondominant ear was followed beyond this level. However, there was no following on the basis of ear of input in Condition 2. (From D. Deutsch, Ear dominance and sequential interactions, *Journal of the Acoustical Society of America*, 1980, 67. Reprinted with permission.)

Throughout this condition the identical frequencies were employed. The basic pattern in Condition 2 consisted of two dichotic chords. In each case the component tones of the chord formed an octave, but the tones in the two chords were of different frequencies. Two pairs of chords were employed, and trials employing these different chord pairs occurred in strict alternation. In this way, any given chord was repeated only after a substantial time period during which several other chords had been interpolated.

The results are displayed in Figure 32.12(b). This again shows the extent to which each ear was followed as a function of the amplitude relationships between the tones at the two ears. In Condition 1, following was clearly on the basis of ear of input. But such following did not occur in Condition 2, even when there were substantial amplitude differences between the tones at the two ears. Instead, these sequences were consistently followed on the basis of their overall contour: The subjects' patterns of response indicated an ascending sequence when the second chord was higher than the first, and a descending sequence when the second chord was lower than the first. Such a result held even when the tones at the two ears differed substantially in amplitude.

Thus in both experiments when the same frequencies emanated successively from different spatial locations, channeling by spatial location always occurred. Otherwise, channeling was on the basis of frequency range. It is noteworthy that relative amplitude turned out not to be an important factor in either experiment. Following by frequency proximity or contour occurred despite large amplitude differences between the signals arriving at the two ears. When following was by ear of input, a shift from following one ear to the other occurred not at the point where the amplitude balance shifted from one ear to the other, but at some other different level of amplitude relationship that varied from subject to subject (see Note 8). This finding lends support to Kubovy's (1981) "Theory of Indispensable Attributes," in which it is argued that the auditory system will organize stimuli on the basis of frequency, as opposed to other attributes such as location or amplitude.

A further question is whether the absence of following by ear of input in the second condition of these two experiments was due to the delay between successive presentations of the same frequencies to the two ears or to the interpolation of different frequencies. A further experiment was performed to study the effect of interpolated frequencies. The patterns employed are shown in Figure 32.13(a). These two patterns were identical except that in Condition 2 a single tone was interpolated between the two presentations of the dichotic chord. Listeners were asked to ignore this tone. As can be seen from Figure 32.13(b), following of the preferred ear was less pronounced in the condition where the extra tone was interpolated than in the condition where there was no interpolated tone.

To investigate the effect of temporal delay, the time interval between onsets of the successive tones was varied. Two methods of varying this temporal parameter were used. Either the durations of the tones were altered, or gaps were interpolated between them [Figure 32.14(a)]. The results, shown in Figure 32.14(b), demonstrated that the degree of following of the preferred ear lessened with increasing time between onsets of the identical frequencies at the two ears. Whether such a time increase was produced by lengthening the durations of the tones or by interpolating silent gaps between them did not matter. Thus channeling by preferred spatial location was shown to be reduced both by interpolated information and by temporal delay.

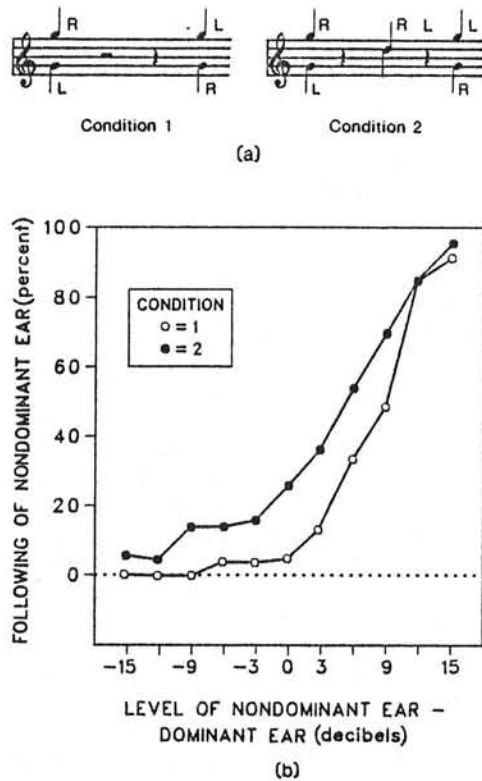


Figure 32.13. (a) Configurations used in third experiment examining the effects of sequential interactions on ear dominance. Conditions 1 and 2 were identical except that in Condition 2 a single binaural tone was interpolated between the two dichotic chords and subjects were asked to ignore this tone. (b) Percentage of following of nondominant ear in these two conditions, as a function of amplitude differences between the tones at the two ears. The interpolation of a single tone in Condition 2 significantly reduced the size of the ear dominance effect. (From D. Deutsch, Ear dominance and sequential interactions, *Journal of the Acoustical Society of America*, 1980, 67. Reprinted with permission.)

We may ask how a system producing such a set of perceptual phenomena could be useful to us. It may be that these phenomena are of value in permitting us to follow new, ongoing auditory information with a minimum of interference from echoes and reverberation. Under natural conditions, when we hear the same frequency emanating in close temporal succession from two regions of auditory space, the second occurrence is in all probability an echo. This explanation becomes less probable as the delay between two such occurrences is lengthened. Further, if different frequencies are interpolated between two occurrences of the same frequency, this interpretation also becomes less probable. It seems, therefore, that the effects we have found are based on a mechanism that serves to counteract misleading effects in our auditory environment (see Note 9). Another such mechanism is the precedence effect, as described by Wallach, Newman, and Rosenzweig (1949) and by Haas (1951). Here a single auditory image may be obtained when the same frequency emanates from two different spatial locations, with onset disparities of less than around 70 msec.

1.2.6. Grouping of Phase-Shifted Tones. Another approach to the issue of grouping by frequency and by spatial location was developed by Kubovy and his co-workers. Kubovy, Cutting, and McGuire (1974) presented a set of eight simultaneous and continuous sine-wave tones to both ears. All tones were at equal

amplitude, and their frequencies were such that they comprised a major scale (see Note 5). The tones were interaurally phase shifted in sequence, with the result that a melody was heard that corresponded to the phase-shifted tones. However, the melody was not detected when the stimulus was presented to either ear alone. At the phenomenological level, the melody was heard as inside the head but displaced to one side of the midline, while a background noise was heard as displaced to the other side, so it appeared to the listener as though a source in one spatial location was generating the melody and another source in a different spatial location was generating the noise. A diagrammatic illustration of this experimental situation is shown in Figure 32.15, taken from Kubovy (1981). This effect is analogous to the Julesz stereogram (Julesz, 1971).

As pointed out by Kubovy (1981), there are two potential interpretations of this effect. On one hand, the segregation of the melody from the noise could be based on *concurrent-difference cues*; that is, the target tone may appear segregated because its interaural disparity differs from that of the background tones. Alternatively the segregation effect could be based on *successive-difference cues*; that is, the target tone may appear segregated because it has changed its interaural disparity whereas the others had not. Two further configurations were therefore devised

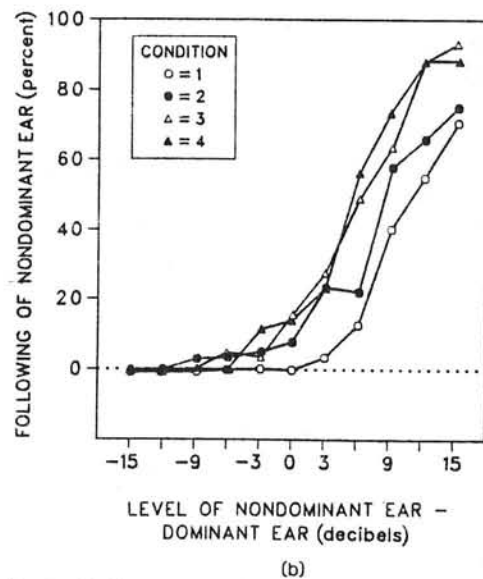
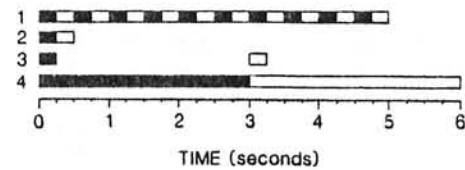


Figure 32.14. (a) Configurations used in experiment to investigate the effects of temporal delay on ear dominance. (b) Percentage of following of non-dominant ear in the different conditions of this experiment. The strength of ear dominance was reduced with increasing time between onsets of the same frequencies at the two ears. There was no effect depending on whether the differences between onsets were produced by lengthening the durations of the tones or by interpolating silent gaps between them. (From D. Deutsch, The octave illusion and auditory perceptual integration, in J. V. Tobias & E. D. Schubert (Eds.), *Hearing research and theory*, Vol. 1, Academic Press, Inc., 1981. Reprinted with permission.)

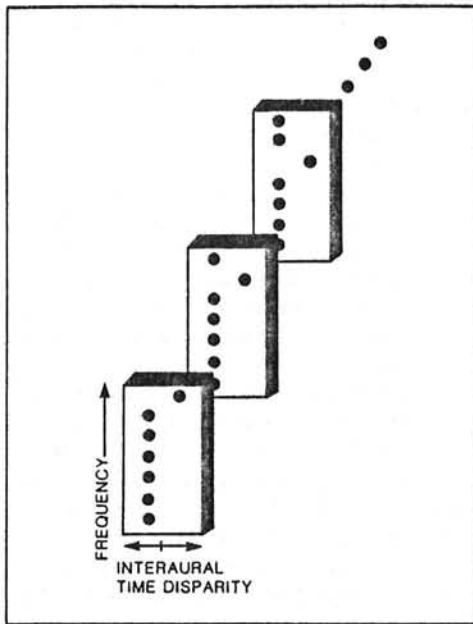


Figure 32.15. Stimulus configuration such as employed by Kubovy, Cutting, and McGuire (1974) to demonstrate grouping of phase-shifted tones. Each slab represents a segment of sound about 300 msec in duration. The abscissa represents interaural time disparity, which produces a shift in the localization of the phase-shifted tone. The ordinate represents frequency on a logarithmic scale. When presented with such a configuration, the listener perceives a descending scale. (From M. Kubovy, Concurrent pitch-segregation and the theory of indispensable attributes, in M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*, Lawrence Erlbaum Associates, 1981. Reprinted with permission.)

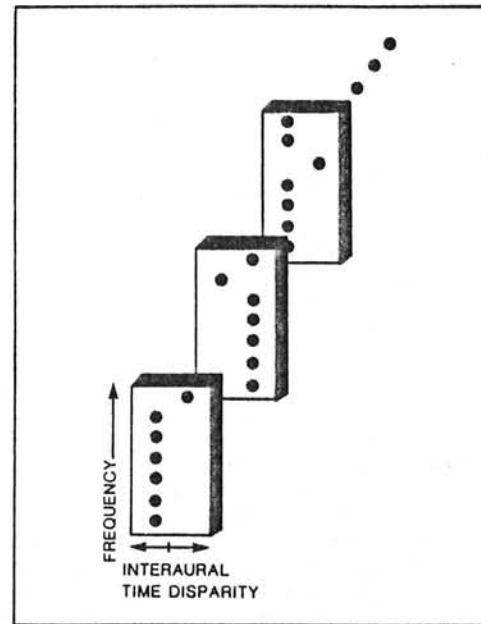


Figure 32.16. Stimulus configuration that has only a concurrent-difference cue. All changes occur in the background tones; these have their phases interaurally shifted. The target tone itself does not change. The target tone appears segregated because its interaural disparity differs from that of the background tones. (From M. Kubovy, Concurrent pitch-segregation and the theory of indispensable attributes, in M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*, Lawrence Erlbaum Associates, 1981. Reprinted with permission.)

to determine which of these two factors was responsible. The first is illustrated in Figure 32.16, which displays a sequence of stimuli in which the target is distinguished from the background only by concurrent-difference cues. Essentially, all changes occur in the background tones; the target tone itself does not change. The second configuration is illustrated in Figure 32.17, which displays a sequence of stimuli that contain only successive-difference cues. The first stimulus consists of four tones to the right and three to the left. The second stimulus is identical to the first except that one of the tones has shifted from right to left. The third stimulus is identical to the second except that one of the tones has shifted from left to right.

Subjects were presented with these three types of stimulus configuration, consisting of confounded, successive-, and concurrent-difference cues. For each type, either an ascending or a descending scale was presented, and subjects identified on forced choice the direction of the scale. The pure successive-difference cues were found to be as effective as the confounded cues. The concurrent-difference cues were less effective, though with these stimuli, performance levels were still above chance. (The poorer performance was here attributed to the fact that, in order to generate concurrent-difference cue stimuli, successive-difference cues were necessarily applied to the background, thus providing contradictory information.)

In another experiment, Kubovy and Howard (1976) studied the effect of interpolating silent intervals between temporally adjacent chords in which successive-difference cues had been introduced. The purpose of the experiment was to measure the amount of time for which the auditory information in the first chord persisted, so that it could be compared with that in the

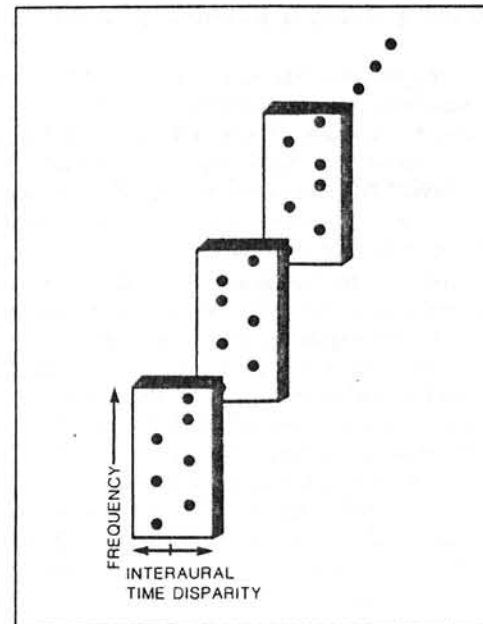


Figure 32.17. Stimulus configuration that has only a successive-difference cue. The first stimulus consists of four tones to the right and three to the left. The second stimulus is identical to the first, except that one of the tones has been interaurally phase shifted from left to right. The target tone appears segregated because it has changed its interaural disparity whereas the others have not. (From M. Kubovy, Concurrent pitch-segregation and the theory of indispensable attributes, in M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*, Lawrence Erlbaum Associates, 1981. Reprinted with permission.)

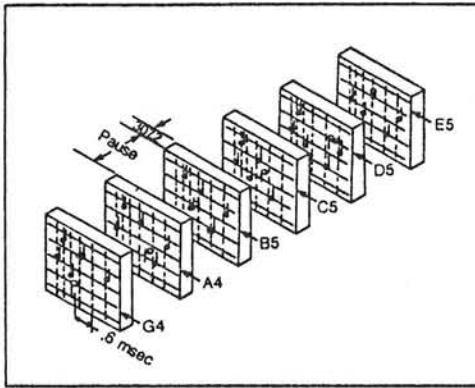


Figure 32.18. Stimulus configuration in experiment to study effect of interpolating silent intervals between temporally adjacent chords in which successive-difference cues had been introduced. Each tone had a different interaural time disparity, and a variable pause was introduced between successive tones. In this example, the listener perceives an ascending scale. An estimate of roughly 1 sec for the persistence of this type of memory was obtained, though there was considerable individual variation. (From M. Kubovy, *Concurrent pitch-segregation and the theory of indispensable attributes*, in M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*, Lawrence Erlbaum Associates, 1981. Reprinted with permission.)

next chord. All chords consisted of six simultaneous tones, around 300 msec in duration, and presented at equal amplitude. Each tone had a different interaural disparity, as shown in Figure 32.18. A variable pause was introduced between successive chords. Subjects judged whether an ascending or a descending scale had been presented. The experiment yielded an estimate of roughly 1 sec for the persistence of this type of memory, though considerable individual differences were observed.

1.3. Grouping of Rapid Sound Sequences

1.3.1. Grouping by Frequency. In the auditory mode, frequency appears to be the most sensitive dimension along which grouping principles operate. This is particularly well illustrated in experiments involving rapid sequences of tones. The next four sections investigate the consequences of grouping by proximity along the frequency dimension and then describe evidence for grouping by good continuation.

1.3.2. Grouping by Frequency Proximity. When a rapid sequence of tones is drawn from more than one frequency range, the sequence tends to split apart perceptually so that two or more melodic lines are perceived in parallel. This phenomenon is exploited in musical composition by the technique of pseudopolyphony, or compound melodic line. Here a single instrument plays a rapid sequence of tones that are drawn from different pitch ranges, with the result that more than one melodic stream is perceived in parallel. Figure 32.19(a) shows a segment of music that exploits this principle. In Figure 32.19(b) the same segment of music is depicted, with log frequency and time mapped into two dimensions of visual space. It is interesting to note that grouping by proximity clearly emerges in the visual analogue, just as it does in the perceived music.

G. A. Miller and Heise (1950) performed one of the first experimental demonstrations of this grouping effect. They presented subjects with a sequence that consisted of two tones alternating at a rate of 10 sec^{-1} . When the frequencies of these tones differed by less than 15%, the sequence was heard as a single coherent string. However, as the frequency disparity between the alternating tones increased, the sequence was heard

instead as two repeating and unrelated tones. This phenomenon has come to be termed *fission*. Heise and G. A. Miller (1951) examined the perception of rapid sequences of tones that were composed of several different frequencies. When the frequency of one of these tones differed sufficiently from the rest, it was heard as in isolation from them.

Later Dowling (1973a) presented two well-known melodies with their component tones alternating at a rate of 8 sec^{-1} . He found that recognition of these melodies was very difficult when their pitch ranges overlapped; however, recognition was easy when their pitch ranges differed. He explained this finding in terms of the operation of the principle of proximity. When the components of the two melodies were proximal in pitch, they were perceptually combined into a single stream, with the result that they were difficult to disentangle. However, this problem did not arise when the melodies were in different pitch ranges.

1.3.3. Temporal Coherence as a Function of Frequency Proximity and Tempo. The term *temporal coherence* is employed to describe the subjective impression that a tonal sequence forms a connected series. In an experiment to study the conditions giving rise to this effect, Schouten (1962) varied the frequency relationships between successive tones in a sequence and also varied their presentation rate. He found that with an increase in the frequency separation between successive tones a reduction in presentation rate was required to maintain the impression of a coherent stream.

Later Van Noorden (1975) presented subjects with sequences of alternating tones and instructed them either to attempt to hear temporal coherence or to attempt to hear fission. He determined two boundaries by this method. The *temporal coherence boundary* defined the threshold frequency separation as a function of presentation rate required for the subject to hear the sequence as coherent. The *fission boundary* defined the threshold frequency separation as a function of presentation rate required for the subject to hear two disconnected series. These two boundaries are shown in Figure 32.20. It can be seen that, when the subjects were attempting to hear coherence, decreasing the presentation rate substantially increased the range of frequency separation within which the sequence was heard as a single stream. However, when the subjects were attempting to hear fission, decreasing the presentation rate had little effect on threshold. In the region between these two boundaries, subjects could alter their listening strategies at will and so hear either fission or coherence.

A later experiment by Bregman and Bernstein (quoted in Bregman, 1978) confirmed the interaction between frequency separation and presentation rate for judgments of temporal coherence and found that this effect was maintained throughout a considerable frequency range.

1.3.4. Grouping by Frequency Proximity in Relation to Repetition. It has been demonstrated in several experiments that the splitting of tonal sequences into separate streams on the basis of frequency proximity develops with repetition. Van Noorden (1975) determined the temporal coherence boundaries for two-tone, three-tone, and long repetitive sequences. In the case of three-tone sequences, the frequency change was either unidirectional or bidirectional. The results are shown in Figure 32.21. For unidirectional three-tone sequences, temporal coherence occurred at rates that were faster than for two-tone sequences. However, for bidirectional three-tone sequences, the upper limit for temporal coherence was lower than for two-tone sequences. For long repetitive sequences, the upper limit was lower still.

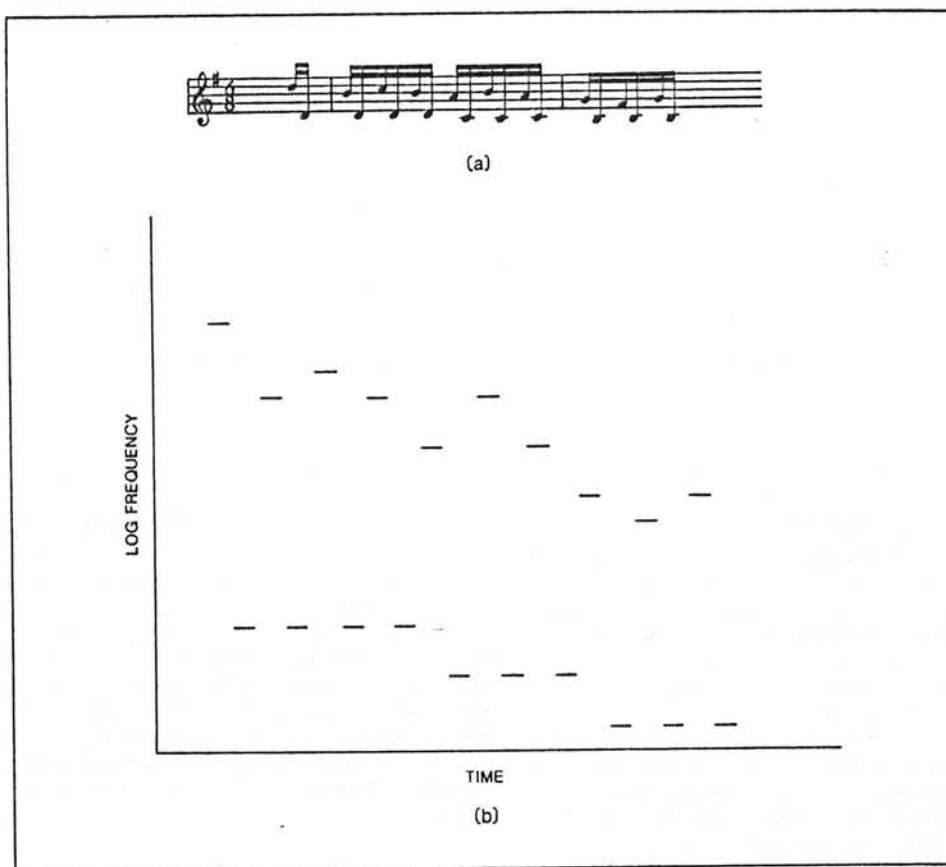


Figure 32.19. Grouping of melodic stimuli on the basis of frequency proximity. Two parallel melodic lines are perceived, each in a different frequency range (passage is from Beethoven's *Six Variations on the Duet "Nel cor piu non mi sento"* from Paisiello's *La Molinara*). (a) The passage in musical notation. (b) The passage with log frequency and time mapped into two dimensions of visual space. (From D. Deutsch, *The processing of pitch combinations*, in D. Deutsch (Ed.), *The psychology of music*, Academic Press, Inc., 1982. Reprinted with permission.)

Bregman (1978) presented subjects with sequences that consisted of two "high" tones and a single "low" tone. At rapid presentation rates this sequence split into two streams such that the upper stream appeared as an alternation of the two high tones and the lower stream as the repeating occurrence of the low tone. The number of tones packaged between 4-sec silent periods was varied, and subjects adjusted the presentation rate to correspond to the point at which splitting into two streams occurred. The results are shown in Figure 32.22. It can be seen that as the size of the package increased the threshold presentation rate for splitting the sequence into two streams decreased. This is in accordance with the results of Van Noorden (1975) described earlier. Bregman interpreted these findings as follows. Stream segregation (or fission) may be viewed as the result of a mechanism that "parses" the auditory environment; that is, it groups together components of the acoustic spectrum so as to reconstruct the original sources of the sounds. It is reasonable to expect that such a mechanism would accumulate evidence over time, and also with repeated presentation.

Further evidence for this view has been provided in an experiment by Bregman and Rudnick (1975). Here two test tones were embedded in a four-tone pattern and so were flanked by two "distractor" tones. Subjects were required to judge the order of the test tones, and it was found that the presence of the distractor tones made this task difficult. However, when another stream of tones, termed "captor" tones, was moved

close in frequency to the distractor tones, the distractors combined with the captors to form a single stream. The test tones were therefore left in a stream of their own. As a result, judgment of their order was facilitated. The authors argued that in this situation the subject is presented with two simultaneous streams of tones, and that the distractor can belong to either one of these, but not to both simultaneously. This is in accordance with the hypothesis of an auditory parsing mechanism: It is unlikely that any single tone would be emanating from more than one source simultaneously.

The cumulation of effect over time reported by Bregman (1978) is analogous to cumulation effects in the octave illusion. Here the strength of tendency to follow the frequencies presented to one side of space rather than to the other also cumulates with repeated presentation, and cumulates more rapidly as repetition rate increases. The strength of tendency to localize the perceived sound toward the source of the higher-frequency signal in this illusion also builds with repetitive presentation (D. Deutsch, 1976, 1978c). Such findings may also be well interpreted in terms of evidence accumulation.

1.3.5. Frequency Proximity and the Perception of Temporal Relationships. When a rapid sequence of tones splits into two separate streams, judgment of temporal relationships between elements of the different streams is impaired. Bregman and Campbell (1971) presented subjects with a repeating sequence

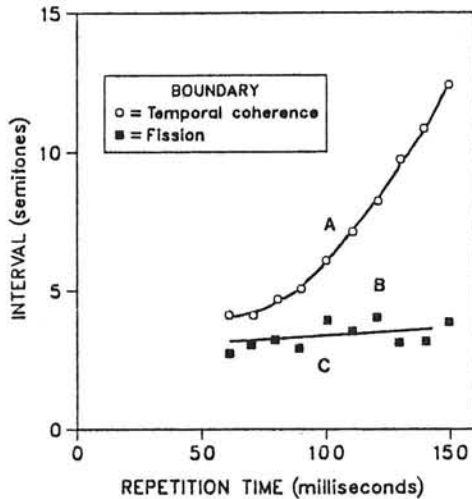


Figure 32.20. Temporal coherence boundary and fission boundary as a function of frequency relationship between alternating tones, and of presentation rate. When the subject was attempting to hear coherence, decreasing the presentation rate substantially increased the range of frequency separation within which the sequence was heard as a single stream. However, when the subject was attempting to hear fission, decreasing the presentation rate had little effect on threshold. In region A, the sequence could be heard only as two streams. In region C, it could be heard only as a single stream. In region B, the subject could choose to hear the sequence either way. (From L. P. A. S. Van Noorden, *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Technische Hogeschool, Eindhoven, the Netherlands, 1975. Reprinted with permission.)

of six tones, three taken from a high-frequency range and three from a low-frequency range. They found that when these tones occurred at a rate of 10 sec^{-1} subjects had difficulty in perceiving a pattern of high and low tones that was embedded in the sequence. Dannenbring and Bregman (1976) later reported that, when two tones alternated at high speeds so that they split into

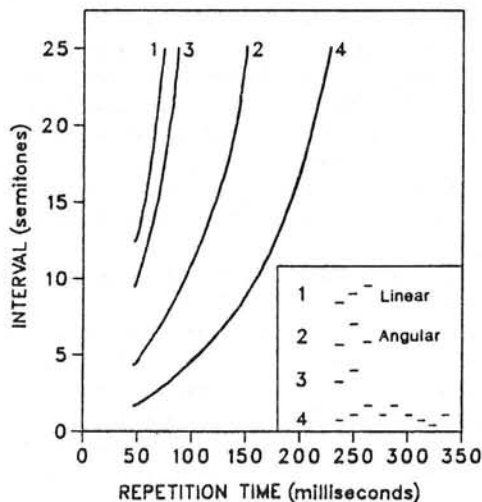


Figure 32.21. Temporal coherence boundary for two-tone, three-tone unidirectional, three-tone bidirectional, and continuous sequences. For unidirectional three-tone sequences, temporal coherence occurred at rates that were higher than for two-tone sequences. However, for bidirectional three-tone sequences, the upper limit for temporal coherence was lower than for two-tone sequences. For long repetitive sequences the upper limit was higher still. (From L. P. A. S. Van Noorden, *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Technische Hogeschool, Eindhoven, the Netherlands, 1975. Reprinted with permission.)

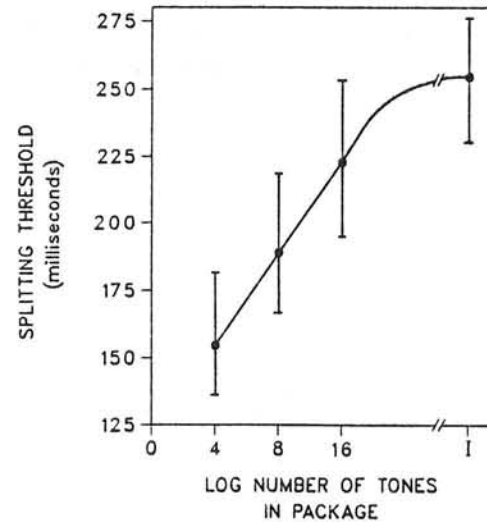


Figure 32.22. Threshold stream segregation as a function of number of tones in a temporal group or "package." Two "high" tones were presented in alternation with a single "low" tone, in temporal groups or packages. As the size of the package increased, the threshold rate for splitting the sequence into two streams decreased. Thus the mechanism that produces stream segregation accumulates evidence with repeated presentation. (From A. S. Bregman, *The formation of auditory streams*, in J. Requin (Ed.), *Attention and performance* (Vol. 7), Lawrence Erlbaum Associates, 1978. Copyright by International Association for the Study of Attention and Performance. Reprinted with permission.)

two perceptual streams, the tones appeared to be overlapping in time. In addition, Fitzgibbon, Pollatsek, and Thomas (1974) studied the perception of temporal gaps between rapidly presented tones. When these tones were in the same frequency range, the interpolation of a 20-msec gap was easily detected. However, when this same gap was interpolated between tones in different frequency ranges, its detection was considerably impaired.

Van Noorden (1975) examined the detection of temporal displacement of a tone that continuously alternated with another tone of different frequency. He found that as the presentation rate increased the threshold for detection of temporal displacement also increased. As shown in Figure 32.23, the greater the frequency separation between the alternating tones, the greater the increase in threshold.

This deterioration in temporal processing resulting from frequency disparity is not confined to continuous sequences but occurs with two-tone sequences also. Divenyi and Hirsh (1972) investigated the discrimination of size of a temporal gap between a pair of tones and found that performance deteriorated with increasing frequency separation between the tones. Williams and Perrott (1972) investigated the minimum temporal gap detectable between tone pairs. For tones of 100- and 30-msec duration, the detection threshold rose with increasing frequency separation between the tones. On the other hand, Van Noorden (1975) has demonstrated that such deterioration of temporal processing is considerably greater for continuous sequences than for two-tone sequences and may thus be considered to result from the formation of separate perceptual streams. This is illustrated in Figure 32.23.

1.3.6. Grouping by Good Continuation. The principle of good continuation has been shown to influence the grouping of tones that occur in rapid succession. Bregman and Dannenbring (1973) presented subjects with a repeating sequence consisting

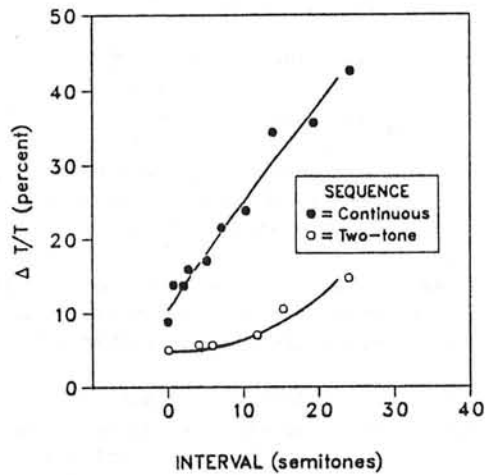


Figure 32.23. Open circles—represent the just noticeable temporal displacement $\Delta T/T$ of the second tone of a two-tone sequence as a function of frequency separation in semitones. Closed circles—represent the just noticeable temporal displacement $\Delta T/T$ of one tone in a continuous sequence of alternating tones as a function of frequency separation in semitones. The greater the frequency separation between successive tones, the higher the threshold for perception of temporal displacement. The effect was more pronounced with continuous sequences than with two-tone sequences. (From L. P. A. S. Van Noorden, *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Technische Hogeschool, Eindhoven, the Netherlands, 1975. Reprinted with permission.)

of a high tone alternating with a low tone. When the frequency disparity between these tones was such that they tended to segregate into two streams, segregation was reduced when the tones were connected by frequency glides. Also I. V. Nabelek, A. K. Nabelek, and Hirsh (1973) studied perception of complex tone bursts and found that when frequency glides were interpolated between the initial and final tones of the burst there was more pitch fusion than when such glides were not interpolated. Divenyi and Hirsh (1974) investigated identification of the order of three-tone sequences. For sequences in which the frequency changes were unidirectional, order perception was superior to that for sequences in which frequency changes were bidirectional. Similar findings were obtained by Nickerson and Freeman (1974), R. M. Warren and Byrnes (1975), and McNally and Handel (1977), using four-tone sequences.

1.3.7. Grouping by Sound Quality. The formation of perceptual groupings on the basis of sound quality is an example of the application of the principle of similarity. A striking demonstration of this phenomenon was created by R. M. Warren, Obusek, Farmer, and R. P. Warren (1969). These authors constructed sequences of four unrelated sounds that were repeatedly presented without pause. The sounds were a high tone (1000 Hz) a hiss (2000-Hz octave band noise) a low tone (796-Hz sine wave) and a buzz (400-Hz square wave). All sounds had a duration of 200 msec. Subjects were found unable to name the orders in which these sounds occurred; however, correct ordering was possible when the duration of each sound was increased beyond 500 msec. This effect is discussed in detail in Section 4.1.4.

Grouping by sound quality was also demonstrated informally in an experiment by Wessel (1979). A repeating three-tone ascending line was presented with two alternating timbres. When the timbral difference between the adjacent tones was

small, percepts were dominated by the ascending pitch lines. However, as the difference in spectral energy distribution between adjacent tones increased, percepts were transformed into two streams based on timbre. As a result, two interwoven descending lines were formed, each with its own timbral identity.

1.3.8. Grouping by Amplitude. Grouping by amplitude has been shown to occur in the perception of rapid sequences of tones. Dowling (1973a) found that, when melodies were interleaved in time, loudness differences between them enhanced the ability to hear the melodies as separate. Van Noorden (1975) investigated the perception of sequences of tones that were of identical frequency but whose amplitudes alternated between two values. When the amplitude difference between the alternating tones was less than 5 dB, a single coherent stream was perceived. However, as the amplitude difference increased, two parallel streams of different loudness were perceived instead. With substantial amplitude differences between the alternating tones, the auditory continuity effect was obtained (Section 1.3.11).

1.3.9. Grouping by Temporal Position. Sound sequences may be divided into subsequences on the basis of temporal position. Such grouping is most readily achieved by interpolating gaps between subsequences, and the evidence on this issue is described in Section 4.2.2.

1.3.10. Grouping by Spatial Location. Rapid sound sequences are under certain conditions grouped by spatial location. The evidence on this issue is discussed in detail in Section 1.2.3. Temporal relationships between the sounds in the different locations can be important in determining whether or not grouping by spatial location occurs, as can frequency relationships between tones that occur in sequence at the different locations.

1.3.11. Closure: The Auditory Continuity Effect. Several of the findings discussed above have demonstrated that the auditory system reorganizes sound sequences in accordance with expectations derived from our knowledge of the auditory environment. It has further been demonstrated that sounds that are not actually present in a stimulus configuration may be perceptually synthesized in accordance with such expectations.

When two sounds of differing amplitude are presented in alternation, the weaker sound may be perceived as continuing through the louder one (G. A. Miller & Licklider, 1950; Thurlow, 1957; Vicario, 1960). Furthermore, when a phoneme in a sentence is replaced by a noise of greater amplitude, the missing phoneme may be perceptually synthesized by the listener (R. M. Warren, 1970; R. M. Warren, Obusek, & Ackroff, 1972). Similar findings have been obtained with nonverbal sounds. This "auditory induction effect" occurs only under stimulus conditions in which one might reasonably conclude that the substituted sound had masked the missing one (Plomp, 1981; R. M. Warren, 1982).

Dannenbring (1976) presented a sine-wave tone whose frequency repeatedly glided up and down. He then substituted a loud noise burst for a portion of this ongoing tone and found that the tone still appeared to glide through the noise. Dannenbring and Bregman (1976) report further that, if the amplitude of the tone changed just before the noise burst, the tendency to perceive the tone as continuing through the noise was reduced. As the authors point out, the change in amplitude produced evidence that something had happened to the tone itself rather than its simply being masked, so as to make a masking hypothesis less likely (see, however, Steiger & Bregman, 1981).

1.4. Grouping and Selective Attention

1.4.1. Voluntary and Involuntary Grouping. In normal listening, we have the impression that we can direct our attention at will to any feature of the auditory environment. However, this impression may often be illusory, and the conditions under which attention is indeed under voluntary control require careful examination. Two issues need to be separated in this discussion. First, we may enquire into the role of voluntary attention in the initial division of an auditory configuration into groupings. Second, once such groupings have been established, we may enquire into the role of voluntary attention in determining which grouping is attended to.

Concerning the first issue, several configurations have been described in which a particular grouping principle is so powerful that the listener is generally unaware of alternative organizations. Thus most listeners when presented with the scale illusion (D. Deutsch, 1975e) form groupings so strongly on the basis of frequency proximity that they mislocalize the tones on this basis. As a result, when attending to the higher or the lower melodic line, they believe that they are attending to one spatial location rather than to another, although this is not in fact the case (D. Deutsch, 1975e). The same phenomenon exists with the contrapuntal patterns (see Note 6) devised by Butler (1979). Again, when presented with the octave illusion, most listeners believed that an intermittent high tone is delivered to one ear, and an intermittent low tone to the other ear. Yet in reality they are being presented with a single two-tone chord. Listeners never attend to the entire configuration, though they generally believe that they are doing so. Yet again, in the sequence devised by Kubovy et al. (1974) the listener hears a melody as though it were occurring in one spatial location and a background noise as though in the other. In reality a continuous two-tone chord is being presented, but it is not perceived as such, and so it is not attended to as such.

However, once such groupings have been established, we find that voluntary factors play a prominent role in determining which of two parallel groupings is attended to. Thus when listeners hear the scale illusion as two melodic lines in parallel (D. Deutsch, 1975e) they can direct their attention at will to either the higher or the lower one. The same holds for the contrapuntal patterns devised by Butler (1979). And again, on listening to the configuration of Kubovy et al. (1974), listeners can focus their attention at will on either the target melody or the background noise.

Strong involuntary factors are also involved in the formation of separate groupings from rapid sequences of tones. Thus, for example, voluntary attention focusing cannot readily overcome the difficulty in perceiving temporal relationships between elements that belong to different groupings, when these are configured as in Bregman and Campbell (1971), R. M. Warren et al. (1969), or D. Deutsch (1979a). In these experiments, the configurations were such as to induce powerful grouping on the basis of frequency proximity, sound type, or spatial location.

For configurations in which groupings are only weakly induced, voluntary attention focusing can exert an influence. For example, Van Noorden (1975) showed that, within a given range of frequency relationships between two alternating tones, and at certain presentation rates, the listener may direct his or her attention at will so as to hear either a single grouping or two separate groupings. This region of ambiguity is shown in Figure 32.20. The composer Robert Erickson addressed this issue with regard to grouping by pitch or by timbre in a composition entitled

LOOPS (Note 10). A repeating melodic configuration was performed by five instruments, with each instrument playing a different note in the manner of a hocket. Each pitch was therefore eventually played by every instrument. Although no formal data were collected, it is clear that on listening to this piece one can direct one's attention at will and so form configurations on the basis either of instrumental timbre or of pitch (see also Erickson, 1975).

As with sequences of simultaneous tones, once groupings are formed from rapid sequences, the listener may voluntarily switch attention from one grouping to another (e.g., see Van Noorden, 1975).

We may conclude that the initial division of a configuration into groupings is often outside the listeners' voluntary control, though ambiguous situations may be generated in which attention focusing can play a role. However, once such groupings have been established, voluntary attention is important in determining which of these is attended to. We may note that such a division of the attentional process into two stages corresponds in certain ways to the stages termed *preattentive* and *postattentive* by Neisser (1967) and by Kahneman (1973). These terms are generally taken to imply different depths of analysis at the two stages; however the question of depth of analysis is as yet unsettled (J. A. Deutsch & D. Deutsch, 1963; Keele & Neill, 1979).

1.4.2. Consequences of Attention Focusing. Finally, we consider the consequences of voluntary attention focusing on the processing of unattended material. Cherry (1953) and Cherry and Taylor (1954) presented subjects with two streams of speech, one to each ear, and asked them to shadow one of the streams. Subjects were able to report very little about the speech that had been presented to the nonattended ear (see also Kahneman, 1973). The present author has informally obtained an analogous result using melodic stimuli instead of speech. Two familiar melodies were generated on a piano, and these were simultaneously presented, one to each ear. Subjects were asked to shadow the melody presented to one ear by singing and later to report what had been presented to the other ear. It was found that the subjects were unable to name the second melody and could describe very little about it. Thus for nonverbal sound sequences also, attention focusing may have the effect of suppressing the unattended material from conscious perception.

It is particularly interesting that such a result should have been obtained for the case of music, since, in contrast to speech, music often consists of several streams of information in parallel. The important question therefore arises as to the extent to which the unattended signal is processed under these conditions. Broadbent (1958) originally proposed that in selective listening a filter selects out elements of a simultaneous configuration on the basis of gross physical characteristics, such as frequency range or spatial location. Stimuli that share a characteristic, so defining a relevant "channel," are then analyzed further, the other stimuli being filtered out. This theory ran into difficulties on experimental grounds. For example, semantic content may be a basis for channel selection (Gray & Wedderburn, 1960; A. Treisman, 1960). To handle such findings, A. Treisman (1960, 1964) suggested a modification of Broadbent's theory, in which the unattended message is not completely filtered out, but rather attenuated. J. A. Deutsch and D. Deutsch (1963) proposed alternatively that all input is perceptually analyzed by the nervous system, whether attended to or not. The analyzed information is weighted for pertinence, the weightings being determined both by long-term factors and by the current situation. On this

model, the information with the highest pertinence weighting controls awareness. Recent studies have provided evidence for the latter view (e.g., Corteen & Wood, 1972; Lewis, 1970; Shiffrin & Schneider, 1977); however, the issue remains controversial.

2. SHAPE ANALYSIS FOR PITCH STRUCTURES

2.1. Auditory Shape Analysis as a Multileveled Process

The analysis of auditory shape may be considered at several stages of abstraction. In the case of shapes built of pitch structures, we may first enquire into the types of abstraction that give rise to local features, such as intervals, chords, and tone chroma. Such features may be considered analogous to orientation and size of angle in vision. Other low-level abstractions give rise to global features such as contour, overall pitch range, general distribution of interval sizes, the proportion of ascending compared with descending intervals, and so on. Such low-level features are combined at a higher level to form more complex configurations, which are themselves abstracted so as to give rise to perceptual equivalences and similarities. At the highest levels of analysis, pitch structures are organized as hierarchies. Since sequential patterns of pitches are spread out in time, short-term memory mechanisms play an important role in determining how such patterns are perceived.

2.2. Passive Versus Active Processing

The multileveled approach to auditory shape analysis does not imply that such analysis proceeds serially from the lowest to the highest level; indeed, we shall see that this is very unlikely to be true. Investigations into mechanisms of visual shape analysis have led to a distinction between an early process, in which many low-level abstractions occur in parallel, and a later process, in which questions are asked of these low-level abstractions, based on hypotheses about the scene to be analyzed (Hanson & Riseman, 1978). Such a distinction between "bottom-up" and "top-down" processing is of importance to auditory shape analysis also; indeed perhaps of greater importance, since the auditory system is more prone to gross error than the visual system and therefore relies more heavily on extraneous cues.

2.3. Feature Abstraction

2.3.1. Octave Equivalence. Tones whose fundamental frequencies stand in the ratio of 2:1 are said to be in octave relation. Such tones possess a strong perceptual similarity, which is evidenced in various ways. In western musical notation, a tone is represented by a letter name, which specifies its position within the octave, together with a number, which specifies the octave in which it occurs. For example, the symbols C_2 , C_3 , and C_4 represent tones that stand in octave relation. In one version of Indian musical notation, a tone is also represented by a letter, which specifies its position within the octave, together with a dot or dots, which specify its octave placement. Thus the symbols \dot{m} , m , \ddot{m} , and $\overset{\cdot}{m}$ represent tones that are separated by octaves. Indeed, octave equivalence appears to be commonly assumed in most musical systems (Burns & Ward, 1982).

People with absolute pitch (i.e., those who can identify musical notes by letter name on hearing them) often place such

notes in the wrong octave. This provides further evidence for octave equivalence at the perceptual level (Bachem, 1954; Baird, 1917). Additional evidence comes from conditioning studies, in which generalization of response to tones separated by octaves has been observed both in humans (Humphreys, 1939) and in animals (Blackwell & Schlosberg, 1942). Yet further evidence for octave equivalence comes from the finding that certain interference effects that operate in pitch recognition (D. Deutsch 1972a, 1973a) also occur when the interference tones are displaced to different octaves (D. Deutsch, 1973b).

In similarity rating paradigms in which a large number of pitch values are employed, octave equivalence effects are not necessarily apparent (Allen, 1967; Kallman, 1982; Thurlow & Erchul, 1977). However, when explicit musical contexts are provided to the subjects, tone pairs that are separated by octaves are judged as closely similar (Krumhansl, 1979; Krumhansl & Shepard, 1979).

Because of the existence of octave equivalence effects, both psychologists and music theorists have argued that pitch should be analyzed in terms of at least two dimensions, the first representing overall pitch level, and the second defining the position of a tone within the octave. These two dimensions have been termed *tone height* and *tone chroma* by psychologists (Bachem, 1948; Burns & Ward, 1982; M. Meyer, 1904, 1914; Revesz, 1913; Ruckmick, 1929; Shepard, 1964), and *pitch* and *pitch class* by music theorists (Babbitt, 1960; Forte, 1973; Westergaard, 1975).

The subjective octave is slightly larger than the 2:1 ratio of the physical octave (Stumpf & M. Meyer, 1898). In an experiment by Ward (1954), subjects were repeatedly presented with two pure tones in succession and asked to adjust the frequency of one until it was exactly an octave above the other. The subjects' adjustments produced ratios that were reliably greater than 2:1. Further, the amount of deviation from the physical octave increased in the higher frequency ranges. Similar findings were obtained by Sundberg and Lindquist (1973) using complex tones. Burns (1974b) obtained analogous results with professional Indian musicians as subjects, showing that the phenomenon of octave stretch is not confined to our culture. A basis for this phenomenon was proposed by Terhardt (1971). He suggested that it is acquired early in life as a result of exposure to complex sounds, such as speech. Due to a mutual masking effect, the pitches of neighboring partials in such complex sounds move away from each other slightly, and Terhardt argued that we generalize from experience with such sounds in making octave judgments. Dowling (1973b) has suggested that octave stretch might alternatively simply reflect innate properties of the auditory system.

2.3.2. Interval and Chord Equivalence. A musical interval is perceived when two tones are presented either simultaneously or in succession. Further, intervals are perceived as the same in size when the fundamental frequencies of their component tones stand in the same ratio. The traditional western musical scale is based in part on this principle. The semitone is the smallest unit of this scale, and it corresponds to a frequency ratio of approximately 1:1.059. Intervals that comprise the same number of semitones are given the same name. Thus, an interval consisting of four semitones is called a major third; an interval consisting of seven semitones is called a perfect fifth; and so on (Figure 32.24). The perceptual equivalence of intervals composed of tones whose fundamental frequencies stand in the same ratio is also assumed by contemporary music theorists (Babbitt, 1960; Forte, 1973; Westergaard, 1975).

Approximate ratio	17:18	8:9	5:6	4:5	3:4	5:7	2:3	5:8	3:5	5:9	11:21	1:2
Number of semitones	1	2	3	4	5	6	7	8	9	10	11	12
Musical interval	Minor second	Major second	Minor third	Major third	Perfect fourth	Tritone	Perfect fifth	Minor sixth	Major sixth	Minor seventh	Major seventh	Octave

$$\log F_1 - \log F_2$$

Figure 32.24. The interval size continuum. This figure gives the number of semitones corresponding to each musical interval, together with the approximate frequency ratio to which it corresponds.

When three or more tones are presented simultaneously, there results the perception of a chord. One may characterize a chord in terms of its component intervals. For example, the major triad is composed of intervals corresponding approximately to the frequency ratios 2:3, 4:5, and 5:6, that is, to 7, 4, and 3 semitones, respectively. Such a characterization may however lead one to assume that two chords are perceptually equivalent when in fact they are not. Thus the minor triad is composed of the same set of intervals as the major triad (Figure 32.25), yet the major and minor triads sound quite different. It is therefore of perceptual importance that, in the major triad, the upper components form a ratio of 5:6 and the lower components a ratio of 4:5, while the reverse is true in the minor triad. It is interesting that certain contemporary music theorists do characterize chords that contain the same set of intervals as equivalent (Babbitt, 1960, 1965; Forte, 1973). This characterization has, however, been challenged on perceptual grounds by other music theorists (e.g., Browne, 1974).

2.3.3. Categorical Perception of Musical Intervals. Although musical intervals vary continuously in size, they are sometimes perceived categorically. Categorical perception is operationally defined according to three criteria. The first is the presence of distinct labeling categories separated by sharp boundaries. The second is the presence of peaks in discrimination performance near category boundaries, with poor discrimination performance within categories. The third is a close correspondence between the discrimination functions that are obtained experimentally and those that are predicted on the hypothesis that stimuli will be discriminated to the extent that they can be identified as different (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). Initially, categorical perception was thought to occur only in the case of consonants in speech; however, more recent exper-

iments have demonstrated its occurrence with nonspeech sounds also (Burns & Ward, 1974; Locke & Kellar, 1973; J. R. Miller, Wier, Pastore, Kelly, & Dooling, 1976; Zatorre & Halpern, 1979).

Category scaling identification functions have been obtained for melodic intervals over ranges of two to five semitones (Burns & Ward, 1974, 1978; Rakowski, 1976; J. A. Siegel & W. Siegel, 1977a, 1977b). Figure 32.26(a) displays the results from a typical subject with musical training. It can be seen that identification functions show sharp category boundaries such as are characteristic of category scaling data for speech materials. Figures 32.26(b), 32.26(c), and 32.26(d) show discrimination functions obtained from the same subject, together with those predicted from the identification functions assuming categorical perception. The agreement between the obtained and predicted discrimination functions is here comparable to that found for speech materials (Burns & Ward, 1978).

In contrast to findings from musically trained subjects, untrained subjects may show inconsistent identification functions. Further, a large effect of shifting the stimulus range has been obtained for subjects with musical training, while this effect was virtually nonexistent for those without musical training. These findings are evidence that categorical perception depends on training (Burns & Ward, 1978; J. A. Siegel & W. Siegel, 1977a).

Listeners appear unable to categorize stimuli reliably in less than semitonal increments (Burns, 1977; Rakowski, 1976). This difficulty has been found not to be confined to western listeners but to be true of Indian musicians also, despite the fact that Indian scales theoretically include microtones (Burns, 1974a, 1977). Such findings may reflect a fundamental limit to the number of discrete interval categories within an octave that listeners can handle.

In addition to melodic intervals, categorical perception has been found to occur for harmonic intervals and triads (Locke & Kellar, 1973; Zatorre & Halpern, 1979).

2.3.4. Global Cues. Global cues are employed in the recognition of pitch sequences. These include overall pitch range, the distribution of the sizes of simultaneous and successive intervals, the proportion of simultaneous compared with successive intervals, the proportion of ascending compared with descending intervals, and so on. Contour has been particularly well documented as a cue in the recognition of melodies (Dowling, 1978; Dowling & Fujitani, 1971; Kallman & Massaro, 1979; Werner, 1925; White, 1960). Such work is described in Section 2.7. It should here be noted that birds are able to discriminate rising from falling pitch patterns (Hulse, Humpal, & Cynx, 1984), showing that sensitivity to contour is not confined to the human case.

2.3.5. Interval Class. The principles of octave equivalence and interval equivalence have led certain music theorists to

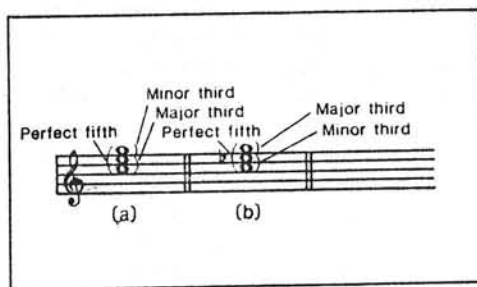


Figure 32.25. The C-major triad (a) and C-minor triad (b). The triads contain the identical set of intervals: the major third, the minor third, and the perfect fifth. However, in the major triad, the major third lies below the minor third, while the reverse is true of the minor triad. Since these triads are perceptually distinct, the ordering of their intervals is of perceptual importance. (From D. Deutsch, Musical recognition, *Psychological Review*, 76. Copyright 1969 by American Psychological Association. Reprinted with permission.)

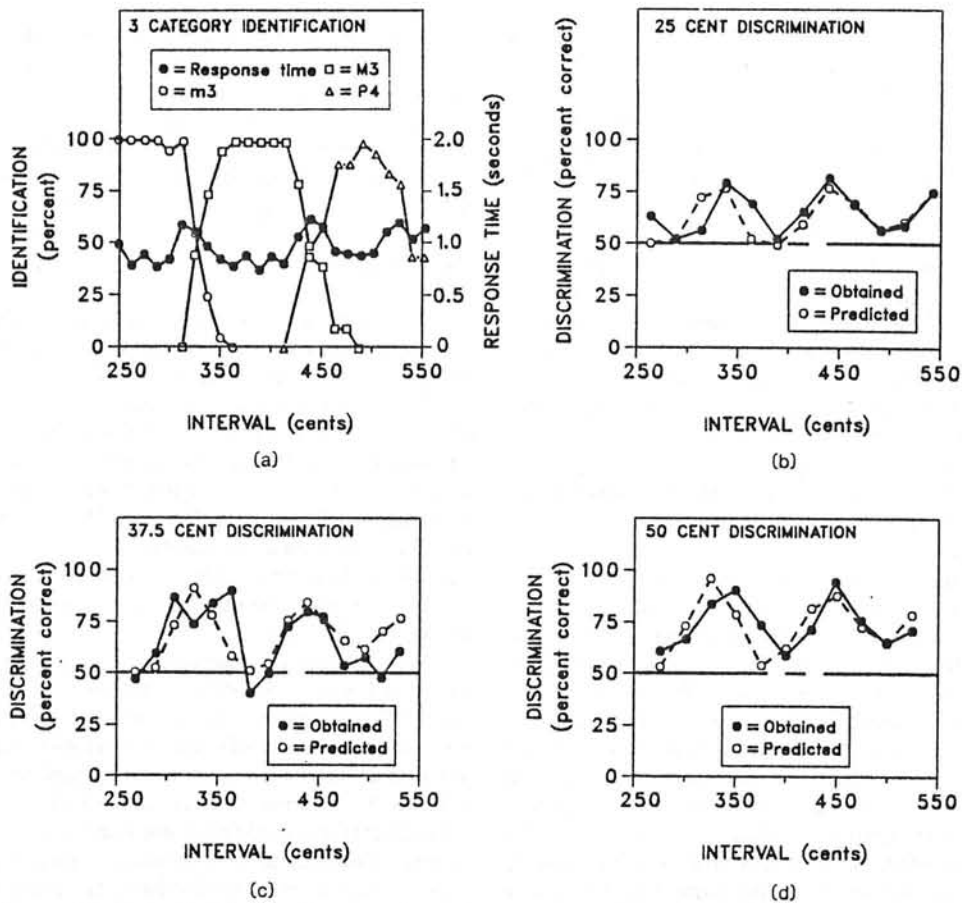


Figure 32.26. (a) Identification functions obtained from a musically trained subject for category scaling of isolated melodic intervals. Sharp category boundaries are apparent. (b), (c), and (d) Discrimination functions (solid lines) obtained from the same subject in a roving-level melodic interval discrimination experiment for interval separations of 25, 37.5, and 50 cents. (Percentage of correct discrimination is plotted at the mean value of the two intervals in a discrimination trial.) Also shown are the discrimination functions (dashed lines) predicted from the identification functions assuming categorical perception. There is good agreement between the obtained discrimination functions and those predicted from the identification functions assuming categorical perception. (From E. M. Burns & W. D. Ward, Categorical perception—Phenomenon or epiphenomenon: Evidence from experiments on the perception of melodic musical intervals, *Journal of the Acoustical Society of America*, 1978, 63. Reprinted with permission.)

assume that an equivalence exists between intervals composed of tones that are placed in different octaves while preserving pitch class (Babbitt, 1960, 1965; Forte, 1973). Such intervals are held to be in the same *interval class*. However, the perceptual validity of this characterization may be debated.

In traditional western music theory, harmonic intervals whose components have reversed position by being placed in different octaves are termed *inversions* (Piston, 1948). Thus a harmonic interval of n semitones is considered perceptually equivalent in certain respects to a harmonic interval of $12 - n$ semitones. Laboratory evidence for the perceptual similarity of inverted intervals has been obtained by Plomp, Wagenaar, and Mimpen (1973). Subjects were asked to identify intervals formed by simultaneous pairs of tones. Confusions were found to occur between intervals that were related by inversion. Further evidence for such equivalence has been provided by D. Deutsch and Roll (1974).

For the case of melodic intervals, the issue of perceptual equivalence based on interval class is complicated. D. Deutsch (1969) has proposed a neural network for the abstraction of octave information, and of interval and chord information, which

predicts that such equivalence should not be directly apprehended. In this network, information travels first to a unidimensional array of "tone height" and is then processed along two separate and parallel channels. Along the first channel there is convergence of input from neural units that underlie tones that are separated by octaves (see Note 11). The output of this channel results in octave equivalence effects for single tones and also in the harmonic equivalence of chords that are related by inversion. The patterns of input along the second channel are such as to mediate transposition of intervals and chords (see Note 12); however, there is no convergence of input based on the octave relation along this channel.

The two-channel model predicts that octave equivalence effects should occur for single tones, and also for simultaneously presented tones. Supporting behavioral evidence for this prediction has been described in Section 2.3.1, and earlier in the present section. However, the model also predicts that, where melodic intervals are concerned, octave equivalence effects should not directly operate. More specifically, it leads to the prediction that listeners should experience difficulty in recognizing well-known melodies in which interval class is preserved,

but in which the pitches of the tones are placed randomly in different octaves. (This prediction does not hold for listeners who know the identity of the presented melody, or who are given cues on which to base hypotheses. Such listeners should be able to perform the recognition judgment by confirming the individual pitch classes, so utilizing the first channel of the model.)

As a test of this prediction, D. Deutsch (1972c) presented subjects with the first half of the melody "Yankee Doodle," with the tones distributed randomly across three adjacent octaves, while preserving pitch class. The subjects were asked to identify the melody but were given no clues on which to base a hypothesis. Recognition was found no better than in a control condition in which the rhythm was retained but the pitch information removed entirely. However, when the subjects were later told the identity of the melody and heard it again, recognition was greatly facilitated. This result is in accordance with the two-channel model and shows that interval class cannot be considered a perceptual invariant.

Further supporting evidence for this view comes from an experiment by D. Deutsch (1979b) on consolidation of memory for melodic patterns. Subjects were presented with a standard melody that was followed by a comparison melody, and they judged whether the two were the same or different. The comparison was always transposed up from the standard. On half the trials this transposition was exact, and on the other half two of the tones were permuted. The experiment consisted of four conditions. In the first, the standard melody was presented once, followed by the comparison melody. In the second, the standard melody was repeated six times and then followed by the comparison melody. In the third, the standard melody was again repeated six times, but on half the repetitions the melody was transposed intact to the next-higher octave, and on the other half it was transposed intact to the next-lower octave. In the fourth condition, the standard melody was again repeated six times, but on each repetition the individual tones were displaced alternately to the next-higher and the next-lower octaves. So in this last condition, interval classes were preserved, though exact intervals were not preserved.

Exact repetition resulted in substantial improvement in recognition performance, and an improvement also occurred when the standard melody was repeated intact in the next-higher and the next-lower octaves. However, when the melody was repeated in such a way that its tones were placed alternately in the next-higher and the next-lower octaves, performance was significantly worse than when the melody was not repeated at all. This experiment again demonstrates that interval class cannot be treated as a perceptual invariant. Repetition of a set of intervals resulted in consolidation of memory for these intervals; however, repetition of a set of interval classes did not produce memory consolidation.

Idson and Massaro (1978) have proposed an alternative explanation for the "Yankee Doodle effect." They pointed out that the octave randomizing procedure results in an alteration in melodic contour and argued that this altered contour provides the listener with misleading information and so actively interferes with the recognition process. The authors found experimentally that, when the individual tones of a melody were displaced to different octaves but contour was preserved, recognition performance was enhanced relative to conditions in which contour was not preserved. A similar result was obtained by Dowling and Hollombe (1977).

The above line of reasoning is problematical, however, because contour alone can serve as a salient cue for melody recognition (Section 2.7). If, then, subjects are able to hypothesize the identity of a melody on the basis of contour alone, they can then confirm the hypothesized melody by reference to the individual pitch classes, and so without direct processing of interval class. So the finding that preservation of contour results in an improvement in recognition performance is in accordance with the two-channel model also (see, in addition, D. Deutsch, 1978d, 1982b).

Idson and Massaro (1978) proposed alternatively that melody recognition depends on two factors: first, recognition of individual pitch classes, and second, recognition of contour. If this hypothesis were correct, then there should be no difference in recognition performance for melodies that are presented without transformation, compared with those in which octave placement is varied but pitch class and contour are retained. However, Kallman and Massaro (1979) found that recognition performance was significantly better in the former case than the latter. This finding is in accordance with the two-channel model but cannot be explained on the hypothesis advanced by Idson and Massaro.

Additional evidence comes from comparing the findings of Idson and Massaro (1978) with those of Kallman and Massaro (1979). In the former study, subjects were furnished with the names of a small set of melodies and were tested for recognition of these melodies under various transformations for hundreds of trials. In contrast, subjects in the latter study were presented with each test melody only once and were not informed of their names. Recognition performance under octave displacement was considerably poorer in the latter study than in the former. This finding is in accordance with the two-channel model, but it cannot be accommodated on the hypothesis that recognition of an ordered set of pitch classes and contours is sufficient to identify a melody.

The possibility still remains, however, that alterations in contour could actively interfere with melody recognition, and thus could play some role in the "Yankee Doodle effect." The extent of such interference cannot be determined with the use of a recognition paradigm, because when contour is altered, this could lead to active interference, yet when contour is preserved, melodies could be recognized on this basis alone.

To circumvent this difficulty, an experiment was performed in which musically literate subjects listened to novel melodic patterns and recalled in musical notation what they had heard. Since no comparison was made with other melodic patterns, the issue of contour as a cue could not arise (D. Deutsch & Boulanger, 1984).

Examples of patterns employed in the different conditions of this experiment are shown in Figure 32.27. Each pattern consisted of a random ordering of the first six notes of the C-major scale (see Note 5). In the first, "higher octave," condition all tones were in the octave beginning on C₅. In the second, "lower octave," condition all tones were in the octave beginning on C₄. In the third, "across octaves," condition, the individual tones in the melody alternated between these two adjacent octaves. In this last condition, roughly two-thirds of the melodic intervals were larger than an octave; the remaining one-third spanned less than an octave.

Also shown in Figure 32.27 are the percentages of tones correctly notated in the correct serial positions in the different conditions of the experiment. Performance in the "across octaves"

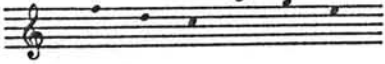


CONDITION		CORRECT NOTATIONS
Higher octave		62.7%
Lower octave		67.5%
Both octaves		31.8%

Figure 32.27. Examples of sequences employed in the different conditions of experiment on the effect of octave jumps on recall of melodic patterns. Also shown are the percentages of tones correctly recalled in the correct serial positions in these different conditions. Recall accuracy was substantially lower for melodic patterns that contained octave jumps than for those that did not. (From D. Deutsch & R. C. Boulanger, *Octave equivalence and the immediate recall of pitch sequences*, *Music perception*. Copyright 1984 by The Regents of the University of California. Reprinted with permission.)

condition was substantially poorer than that in the other two. Further analyses showed that errors in which the correct pitch class was noted but octave placement was incorrect were virtually absent in all conditions, so that the poorer performance in the "across octaves" condition could not be attributed to errors in octave placement. The decrement in recall for melodic patterns involving octave jumps is as predicted on the two-channel model but cannot be explained on the alternative hypothesis that melodic processing occurs through identification of an ordered set of pitch classes together with contour.

2.4. Higher-Order Abstractions

This section is concerned with the ways in which low-order features based on pitch are combined so as to give rise to perceptual equivalences and similarities. In the visual mode, shapes are recognized as equivalent when these differ in size or position in the visual field and (at least under some conditions) in orientation. This leads us to enquire whether auditory analogues of such visual shape abstractions also exist.

Early theorists have speculated concerning possible analogies between relationships in the pitch domain and relationships in visual space (Helmholtz, 1859/1954; Koffka, 1935; Mach, 1906/1959). Recent theoretical discussions have focused particularly on the mapping of one dimension of visual space into frequency and the other into time (Julesz & Hirsh, 1972; Kubovy, 1981). Various grouping phenomena in the perception of sound sequences have clear visuospatial analogues (Bregman, 1978; D. Deutsch, 1975c; Divenyi & Hirsh, 1978; Van Noorden, 1975), as exemplified by the musical passage in Figure 32.19.

In his seminal paper on shape perception, the Gestalt psychologist Von Ehrenfels (1890) pointed out that melodies retain their perceptual identities when they are transposed to different pitch ranges, provided that the relationships between the successive tones in the melodies are unchanged. He argued that in this respect melodies are akin to visual shapes, which preserve their identities when they are translated to different regions of the visual field. This leads us to enquire whether further equivalences may be demonstrated for auditory shapes which have counterparts in vision.

Most relevant to this question is the system of 12-tone composition, developed by Schoenberg, which is based on a theory

of shape analysis for pitch structures. This theory is in turn based on an intermodal analogy in which one dimension of visual space is mapped into pitch and another into time. Schoenberg argued that transformations that are analogous to rotation and reflection in vision give rise to perceptual equivalences for structures in pitch-time also. As he put it:

THE TWO-OR-MORE DIMENSIONAL SPACE IN WHICH MUSICAL IDEAS ARE PRESENTED IS A UNIT . . . The elements of a musical idea are partly incorporated in the horizontal plane as successive sounds, and partly in the vertical plane as simultaneous sounds. . . . *The unity of musical space demands an absolute and unitary perception.* In this space . . . there is no absolute down, no right or left, forward or backward. . . . To the imaginative and creative faculty, relations in the material sphere are as independent from directions or planes as material objects are, in their sphere, to our perceptive faculties. Just as our mind always recognizes, for instance, a knife, a bottle, or a watch, regardless of its position, and can reproduce it in the imagination in every possible position, even so a musical creator's mind can operate subconsciously with a row of tones, regardless of their direction, regardless of the way in which a mirror might show the mutual relations, which remain a given quantity. (Schoenberg, 1951, p. 229)

Schoenberg thus argued that a tone row, defined as a given ordering of the 12 tones of the chromatic scale, retains its perceptual identity under the following transformations: When it is transposed in pitch (transposition), when the directions of the successive intervals are reversed (inversion), when the tones are presented in reverse order (retrogression), and when the two latter transformations are both applied (retrograde-inversion). He further assumed that a perceptual equivalence exists for tones in the same pitch class and for intervals in the same interval class. These assumptions are illustrated in Figure 32.28.

Schoenberg's theory has served as the basis for much musical system building. The group-theoretic approach of Milton Babbitt and his followers has been particularly influential here. The elements of the group are 12-tone sets, which are represented as permutations of pitch or order numbers, and the operation is the multiplication of permutations (Babbitt, 1960, 1965; see also Perle, 1972, 1977).

This leads us to enquire whether the equivalence relations defined in 12-tone theory are perceived by the listener. The

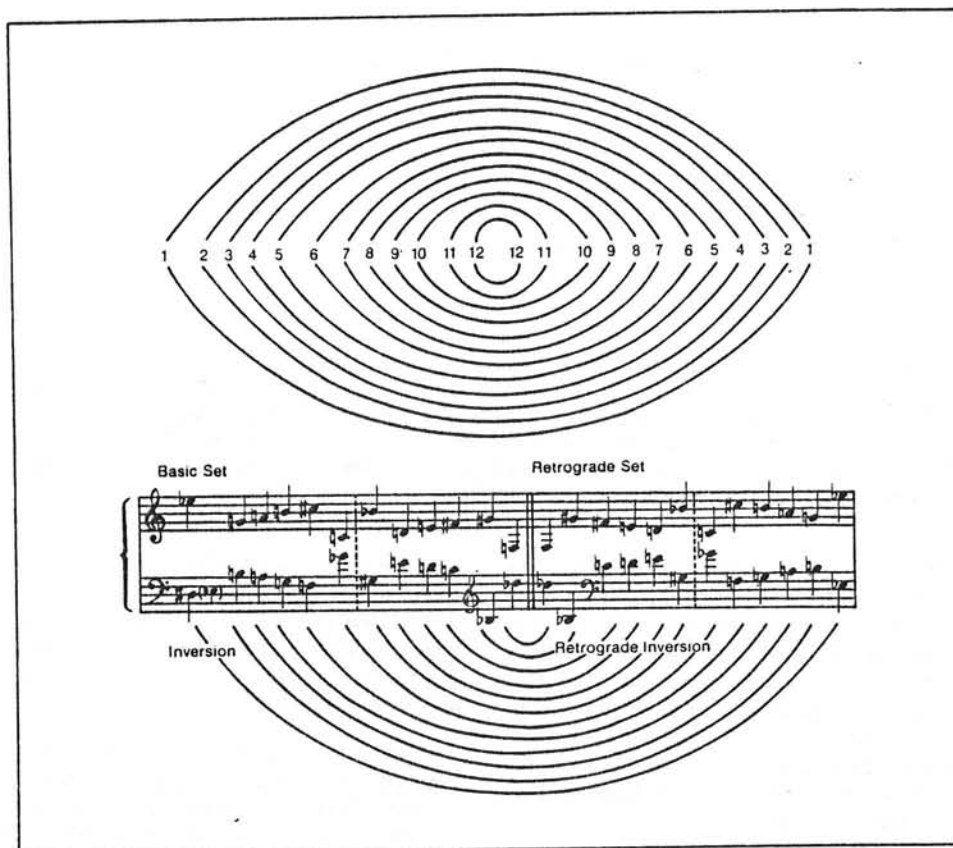


Figure 32.28. Schoenberg's concept of "musical space." Sequences of pitches are considered equivalent at an abstract level when they are transposed to a different pitch range (transposition), when all ascending intervals become descending intervals, and vice versa (inversion), when the tones are presented in reverse order (retrogression), when they are transformed by both these operations (retrogression-inversion), and when the component tones of the sequence are placed in different octaves. (From A. Schoenberg, *Style and idea*, Williams & Norgate, 1951. Reprinted with permission.)

issue of interval class has been discussed at length, and it has been shown that this cannot be taken to be a perceptual invariant. Concerning inversion and retrogression, we may note that there is a clear evolutionary advantage to mechanisms that enable us to recognize a visual object when it is presented in a different orientation. However, there is no analogous advantage to recognizing a sequence of sounds presented in reverse order, or whose pitch relationships are inverted.

The experimental evidence on the issue of equivalence under retrogression and inversion is equivocal. The ability of listeners to identify well-known melodies presented in retrogression was studied by White (1960). He found that identification performance was here no better than when the pitch information was removed entirely, with rhythm serving as the only cue. Dowling (1972) employed a short-term paradigm to study recognition of melodies under retrogression, inversion, and retrograde-inversion. Subjects were presented with a standard melody, followed by a comparison melody. In one set of conditions, the comparison was either unrelated to the standard or transformed by transposition, retrogression, inversion, or retrograde-inversion. In a second set of conditions, the comparison melody was further transformed so that its contour was preserved but the interval sizes were altered. Subjects were found to perform no better on exact transformations than on those that preserved contour alone. In a later study, Dowling (1978) provided evidence for

an interference effect on recognition of exact intervals, resulting from the listener's projecting the pitch information onto overlearned musical scales (see also D. Deutsch, 1977).

From analysis of tonal music, it would appear that retrogression and inversion are indeed perceived in short-term situations, provided that the memory load is not too heavy (e.g., see L. B. Meyer, 1973). However, inversion here takes place along highly overlearned pitch alphabets such as diatonic scales or triads. Rather than assuming a perceptual equivalence based simply on frequency ratios, it would appear that such operations are performed at a level of abstract encoding equivalent to the level that enables us to invert an overlearned alphanumeric sequence (D. Deutsch & Feroe, 1981; Simon & Sumner, 1968).

A considerable body of contemporary music theory is concerned with defining equivalence and similarity relations between sets of pitches. These theories assume equivalence under retrogression and inversion, as well as interval class identity (Chrisman, 1971; Forte, 1973; Howe, 1965; Lewin, 1960, 1962; Perle, 1972, 1977). A detailed examination of such theories is, however, beyond the scope of the present chapter.

A different approach to the structuring of pitch relationships stems from the classical theory of tonality (Helmholtz, 1859/1954; Rameau, 1722/1971) and treats as fundamental the intervals of the octave, the perfect fifth and the major third. Debates concerning tuning systems have utilized two-dimensional arrays

in which tones lying adjacent along one dimension are separated by major thirds, and tones lying adjacent along the other dimension are separated by perfect fifths. An evaluation of different schemes for tuning and temperament based on such arrays is provided by Hall (1974, 1980).

Longuet-Higgins (1962a, 1962b, 1978) has hypothesized that such two-dimensional arrays may form the basis of key attribution. As shown in Figure 32.29, the notes in a diatonic scale (see Note 5) form a compact group in this two-dimensional space, so that a key can be defined as a neighborhood in the space. Longuet-Higgins suggested that when a musical passage is presented listeners select a given region of this space, so attributing a key. If, however, their choices force them to make large jumps in this space, they select instead a different region where the tones are more compactly represented. A different key is thus attributed.

Considerations of octave equivalence have formed the basis of yet another approach to the description of pitch structures. As described by Drobisch (1846/1929), the perceptual similarity of tones standing in octave relation can be accommodated by deforming the unidimensional scale of pitch into a helix, with tones separated by octaves lying most proximal within each turn of the helix. Shepard (1964) has provided experimental evidence for such a helical representation in a harmonic setting. He generated a set of tones, each of which consisted of many sinusoidal components separated by octaves. The amplitudes of these components differed according to a fixed bell-shaped envelope. When such tones were presented in monotonically ascending steps, listeners perceived a sound that constantly ascended in pitch and never descended.

Shepard (1982) later proposed an elaboration of the helical model in which pitch is represented as a five-dimensional helical structure. Along one dimension, tones are ordered in accordance with pitch height. Two further dimensions accommodate the circular representation of tone chroma, and two more dimensions

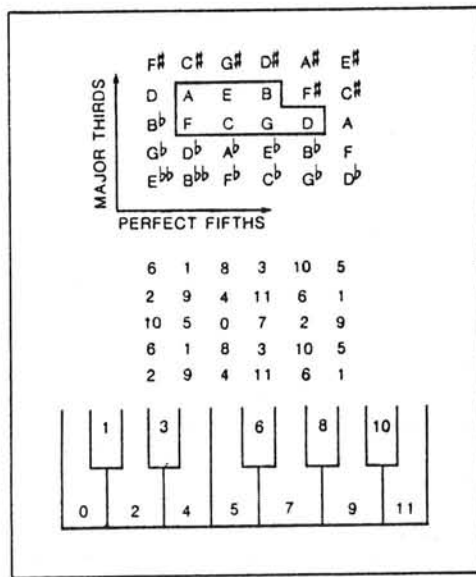


Figure 32.29. Two-dimensional array proposed for the representation of "tonal space." Tones lying adjacent along one dimension are separated by fifths; tones lying adjacent along the other dimension are separated by major thirds. (From H. C. Longuet-Higgins, *The perception of music, Interdisciplinary Science Reviews*, 1978, 3. Reprinted with permission.)

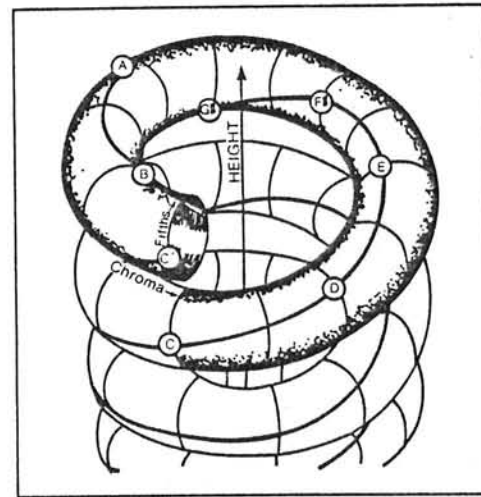


Figure 32.30. Representation of pitch as a double helix wrapped around a helical cylinder in five dimensions. (From R. N. Shepard, *Structural representations of musical pitch*, in D. Deutsch (Ed.), *The psychology of music*, Academic Press, Inc., 1982. Reprinted with permission.)

accommodate the circle of fifths. This model is illustrated in Figure 32.30. Shepard further demonstrated a simple affinity (in the mathematical sense) between this space and the space based on perfect fifths and major thirds.

One problem with specifying invariant geometric structures for pitch comes from evidence showing that a set of notes played in a particular musical key will induce an internal framework that is specific to that key. The internal representation of pitch relationships will thus be expected to differ depending on the key that is attributed. Risset (1978) has pointed out that the same melodic interval may be quite differently perceived when presented in different contexts. There is evidence that performing musicians will accordingly produce a given interval as larger or smaller in size depending on its tonal function (Schackford, 1961, 1962; Small, 1936). Krumhansl (1979) performed an experiment to determine the effects of tonal context on the perception of pitch relationships. Subjects were presented with a set of context tones, followed by two tones in succession. The context tones consisted of either the C-major triad or the C-major scale. On each trial, subjects judged the similarity of the first to the second tone in this context. Multidimensional scaling of the subjects' judgments yielded a three-dimensional conical structure around which tones were arranged according to pitch height. The tones of the C-major triad formed a closely related cluster near the vertex of the cone, and the remaining tones of the C-major scale formed a less closely related subset further from the vertex. Tones not in the C-major scale were still further dispersed (Figure 32.31). In addition, tones less central to the key were judged more similar to tones more central to the key than the reverse. These findings may be related to the tendency described by music theorists for less "stable" tones in a key to "resolve" to more "stable" tones (see also Krumhansl, 1983; Krumhansl, Bharucha, & Kessler, 1982; Krumhansl & Kessler, 1982).

2.5. Hierarchical Encoding of Pitch Sequences

In general, when observers are presented with artificial serial patterns that may be hierarchically encoded, they form encodings

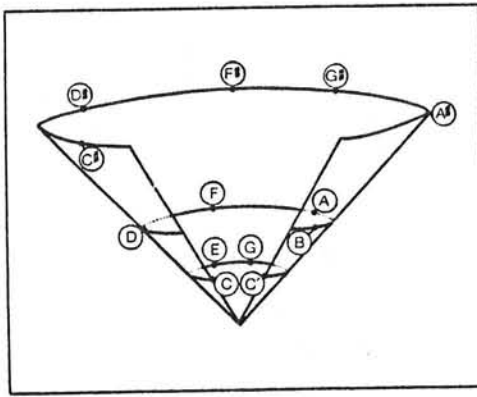


Figure 32.31. Three-dimensional representation of the interrelations between the tones of the chromatic scale spanning an octave, when presented in C-major context. (From C. L. Krumhansl, *The psychological representation of musical pitch in a tonal context*, *Cognitive Psychology*, 1979, 11. Reprinted with permission.)

that reflect the ways these patterns are structured (Bjork, 1968; Kotovsky & Simon, 1973; Restle, 1970; Restle & Brown, 1970; Simon & Kotovsky, 1963; Vitz & Todd, 1967, 1969). Such findings have given rise to models of serial pattern representation in terms of hierarchies of operators (D. Deutsch & Feroe, 1981; Greeno & Simon, 1974; Leeuwenberg, 1971; Restle, 1970; Simon, 1972; Simon & Kotovsky, 1963; Simon & Sumner, 1968; Vitz & Todd, 1967, 1969).

A good example of experimental evidence for such encoding has been provided by Restle (1970), and Restle and Brown (1970). Subjects were presented with arrays of lights that flashed on and off in repetitive sequence, and their task was to predict which light would flash on next. To illustrate the type of pattern employed, take the basic subsequence $X = (123)$. The operation X ("transposition +1 of X ") produces 1 2 3 2 3 4, the operation R ("repeat of X ") produces the sequence 1 2 3 1 2 3, and the operation M ("mirror image of X ") produces the sequence 1 2 3 6 5 4. Recursive application of such operations can generate long sequences that have compact structural descriptions. Thus the sequence 1 2 3 1 2 3 6 5 4 6 5 4 can be described as $M(R(X))$. This example corresponds to the structural tree shown in Figure 32.32.

Restle and Brown found with such sequences that the probability of error in prediction increased monotonically with the level of transformation along the structural tree. Thus the highest probability of error in the present example would occur at locations 1 and 7, and the next highest at locations 4 and 10. It was concluded from these and other findings that observers organize information in accordance with such structures.

The sequences in this study, however, were structured in such a way as to allow for only one parsimonious encoding. It is difficult, therefore, to generalize from findings on artificial serial patterns to the encoding of sequences that do not have such special characteristics. The same argument applies to the other experimental work cited above.

In considering how serial patterns may in general be hierarchically encoded, it is instructive to consider the organization of tonal music. Such music is strongly hierarchical in nature (Keiler, 1983; Lerdahl & Jackendoff, 1983a, 1983b; L. B. Meyer, 1973; Narmour, 1977, 1983; Salzer, 1962; Schenker, 1956, 1973), and it is reasonable to assume that its structure has evolved to make optimal use of our processing mechanisms. This

is particularly the case where the structure of pitch sequences is concerned.

D. Deutsch and Feroe (1981) have proposed a model that takes the structure of tonal music into account and also shows how its characteristics can be exploited so as to arrive at representations that are parsimonious and also capitalize on general tendencies of our processing systems. The model assumes that pitch sequences are represented as hierarchies, at each level of which elements are organized as structural units in accordance with laws of figural goodness (e.g., proximity; good continuation), and which tend to be of optimal chunk size. Elements at any given level are elaborated by further elements so as to form structural units at the next-lower level until the lowest level is attained.

A simplified set of rules for the system is as follows:

1. A structure A of length n is notated as $(A_0, A_1, \dots, A_{l-1}, *, A_{l+1}, \dots, A_{n-1})$, where A_j is one of the operators n, p, s, n^i or p^i . (A string of length k of an operator A is abbreviated kA .)
2. Each structure has associated with it an alphabet, α . The combination of a structure and an alphabet defines a sequence. This, in combination with the reference element r , produces a sequence of notes.
3. The effect of an operator is determined by that of the operator next to it, but on the same side as $*$. So, for instance, the operator n refers to moving one step up the alphabet associated with the structure. The operator p refers to moving one step down the alphabet. The operator s refers to remaining in the same position. The operators n^i and p^i refer to traversing up or down i steps in the alphabet respectively.

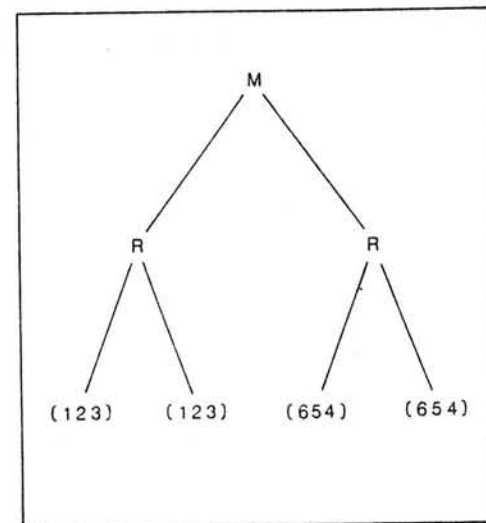


Figure 32.32. Structural tree corresponding to a particular sequence, according to Restle (1970). The sequence illustrated is of six events, which occur in the order 1 2 3 1 2 3 6 5 4 6 5 4. According to the theory, the basic subsequence is $X = (1 2 3)$. The operation R ("repeat of X ") produces 1 2 3 1 2 3. The operation M ("mirror image of X ") produces 1 2 3 6 5 4. The sequence in figure can be described as $M(R(X))$. When subjects are asked to predict the next event, they make the largest number of errors at locations 1 and 7, and the next largest at locations 4 and 10, as expected from the theory. (From D. Deutsch & P. L. Roll, *Separate 'what' and 'where' decision mechanisms in processing a dichotic tonal sequence*, *Journal of Experimental Psychology: Human Perception and Performance*, 2. Copyright 1976 by American Psychological Association. Reprinted with permission.)

4. The values of the sequence of notes ($A_0, A_1, \dots, *, \dots, A_{n-1}$), α, r , are obtained by taking the value of $*$ to be that of r .
5. Given two sequences $\mathcal{A} = (A_0, A_1, \dots, *, \dots, A_{n-1})$, α , and $\mathcal{B} = (B_0, B_1, \dots, *, \dots, B_{m-1})$, β , define the compound operator pr (prime). $\mathcal{A}[pr]\mathcal{B};r$ refers to assigning values to the notes produced from $(B_0, B_1, \dots, *, \dots, B_{m-1})$, such that the value of $*$ is identical to the value of A_0 , when the sequence \mathcal{A} is applied to r . Values are then assigned to the notes produced from $(B_0, B_1, \dots, *, \dots, B_{m-1})$, so that the value of $*$ is identical to the value of A_2 , and so on. This produces a sequence of length $n \times m$. Other compound operators are analogously defined, such as inv (inversion) and ret (retrograde).

The sequence shown in Figure 32.33(a) provides an example to illustrate the model. One could theoretically describe this sequence in terms of steps ascending the chromatic scale (see Note 13). One may state that a basic subsequence consisting of a step up this scale is presented four times in succession, the second presentation being four steps up from the first, the third being three steps up from the second, and the fourth being five steps up from the third. In terms of the present formalism, such a representation would take the cumbersome form of $\{(*, n, n^3, n, n^2, n, n^4, n); Cr\}$.

However, this description does not relate the key elements of the four subsequences in any meaningful fashion. Musical analysis would instead describe this sequence as represented on two hierarchical levels. On the higher level, shown in Figure 32.33(b), there is an arpeggiation of the C-major triad (the notes C-E-G-C). On the lower level, each note of the triad is preceded by a neighbor embellishment, so forming a two-note pattern. This representation is illustrated in the tree diagram on Figure 32.33(c). On the model, the sequence is represented as

$$\mathcal{A} = \{(*, 3n); 1\}$$

$$\mathcal{B} = \{(p, *); Cr\}$$



Figure 32.33. Pitch sequence (a), represented as on two hierarchical levels. The higher level (b) consists of an arpeggiation of the C-major triad. At the lower level, each note of the triad is preceded by a neighbor note, so that a two-note pattern is formed. This structure corresponds to the tree diagram in (c). (From D. Deutsch & J. Feroe, The internal representation of pitch sequences in tonal music, *Psychological Review*, 88. Copyright 1981 by American Psychological Association. Reprinted with permission.)

$$\mathcal{C} = \{A[pr]B; 1\}C$$

Such a representation has multiple cognitive advantages. First, since two alphabets are utilized, only single steps are employed at each level, in accordance with the principle of proximity. Second, this representation involves two structures, and taken as chunks of information, the first is a four-element chunk, and the second is a three-element chunk. The representation is therefore in terms of chunks of optimal size (Estes, 1972; Wickelgren, 1967). Third, since notes that are present at the higher level are also present at the lower level, the higher-level notes are given prominence, through redundancy of representation, and so serve to cement the lower-level notes together. (This contrasts with Restle's formalism, in which the reverse is the case.) These and other processing advantages are discussed at length in Deutsch and Feroe (1981).

Many useful insights into cognitive processing of pitch patterns may be gained from detailed examination of various music-theoretic accounts of hierarchical structure. (Since these accounts rely on familiarity with music theory, they will not be described here, but the interested reader is referred to the following sources: Keiler, 1983; Lerdahl & Jackendoff, 1983a, 1983b; L. B. Meyer, 1956, 1960, 1973; Narmour, 1977, 1983; Schenker, 1956, 1973).

2.6. The Influence of Short-Term Memory on Perception of Pitch Patterns

The accuracy with which a pattern of pitches is perceived depends on the accuracy with which the individual pitches in the pattern can be related to each other. The most simple relational judgment that can be made here is whether two tones occurring in the pattern are the same or different. As will be shown, such judgments are heavily dependent on a number of variables.

2.6.1. Interference Effects in Short-Term Memory for Pitch. When two tones that occur in succession are to be judged as the same or different in pitch, recognition accuracy declines very slowly over a silent retention interval (Bachem, 1954; Harris, 1952; Koester, 1945; Wickelgren, 1966, 1969a). In contrast, the interpolation of a sequence of extra tones during the retention interval results in a substantial decrement in performance. This is true even when the subjects are instructed to ignore the interpolated tones. The disruptive effect due to the interpolated tones is specific in nature and is not based on an overloading of some general limited-capacity storage system. The interpolation of a sequence of spoken digits instead does not cause a performance decrement, even when recall of these digits is required (D. Deutsch, 1970).

The interference effect of a tone that forms part of an interpolated sequence depends on the pitch relationship between this tone and the first test tone. D. Deutsch (1972b) demonstrated this phenomenon using the paradigm illustrated in Figure 32.34. Subjects were presented with a first test tone, which was followed by a sequence of interpolated tones, and then by a second test tone. Either the test tones were identical in pitch or they differed by a semitone. The subjects were asked to ignore the interpolated tones and to judge whether the test tones were the same or different. The relationship between the tone in the second serial position of the interpolated sequence (the "critical tone") varied in increments of one-sixth of a tone between identity and a whole tone separation.

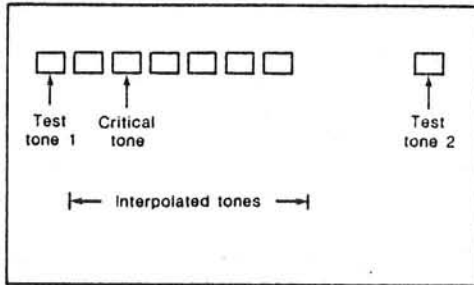


Figure 32.34. Representation of paradigm to examine the effect on pitch recognition accuracy of a critical tone that formed part of a sequence that was interpolated between two test tones. Either the test tones were identical in pitch or they differed by a semitone, and subjects judged whether they were the same or different. The relationship between the critical tone and the first test tone varied in increments of one-sixth of a tone between identity and a whole tone separation. (From D. Deutsch, *The organization of short-term memory for a single acoustic attribute*, in D. Deutsch & J. A. Deutsch (Eds.), *Short term memory*, Academic Press, Inc., 1975. Reprinted with permission.)

The results of the experiment are shown on Figure 32.35. When the first test tone and the critical tone were identical in pitch, memory was facilitated. As the pitch separation between these tones increased, the error rate also increased, then peaked at a separation of two-thirds of a tone and returned to baseline at a separation of roughly a whole tone.

This pattern of results may be explained by assuming that pitch memory is the function of an array whose elements are activated by tones of specific pitch. These elements are tonotopically organized on a log frequency continuum. Inhibitory interactions take place between elements along this array that are a function of the distance separating them. These interactions are assumed to be analogous to recurrent lateral inhibitory interactions in systems that handle sensory information at the incoming level (Ratcliff, 1965). It is assumed that when these

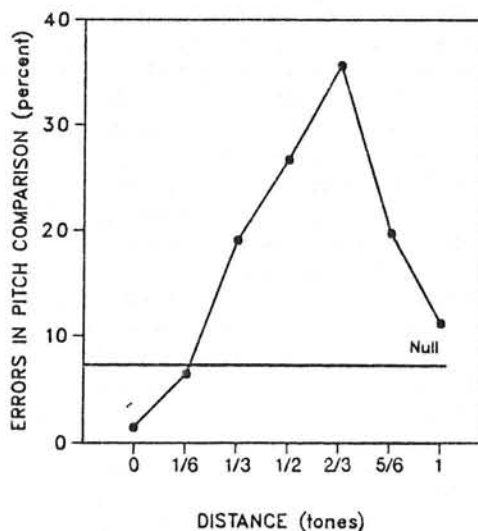


Figure 32.35. Percentage of errors in pitch comparisons, plotted as a function of the separation in pitch between a critical interpolated tone and the first test tone. The critical tone was always in the second serial position of a sequence of six interpolated tones. The error rate was maximal at a separation of two-thirds of a tone. (From D. Deutsch, *Mapping of interactions in the pitch memory store*, *Science*, 1972, 175. Reprinted with permission.)

memory elements are inhibited they emit weaker signals, so that an increase in recognition errors results.

This hypothesis is strengthened by two further lines of evidence. First, the relative frequency range over which the disruptive effect operates corresponds well with the range over which centrally acting lateral inhibition has been found in physiological studies of the auditory system (Klinke, Boerger, & Gruber, 1969, 1970). Second, the error rate in this pitch recognition task cumulates when two critical tones bearing an inhibitory relationship to the first test tone are interpolated, one higher than the first test tone and the other lower (D. Deutsch, 1973a). Analogously, in lateral inhibitory networks there is also a cumulation of inhibition from stimuli that are placed on either side of the test stimulus (Ratcliff, 1965).

If a recurrent lateral inhibitory network were indeed involved here, we should also expect to find the phenomenon of disinhibition (see Note 14). More precisely, we should expect that, if a tone that was inhibiting memory for another tone was itself inhibited by a third tone, memory for the first tone should return.

Accordingly, D. Deutsch and Feroe (1975) performed the following experiment. Subjects compared two test tones for pitch with the tones separated by a sequence of six interpolated tones. In the second serial position of the interpolated sequence there was always placed a tone that was two-thirds of a tone removed from the first tone (i.e., in a maximally inhibiting relationship to the first test tone). Errors were plotted as a function of the pitch of a further tone. This tone was placed in the fourth serial position, and its relationship to the tone in the second serial position varied between identity and a whole tone separation. The results are shown in Figure 32.36. It can be seen that the predicted effect was indeed obtained. The error rate in sequences in which the second critical tone was identical to the first was significantly higher than baseline. Further, the error rate in sequences in which the two critical tones were separated by two-thirds of a tone was significantly lower than baseline. In a companion experiment using subjects selected on the same criterion as for the disinhibition study, a first-order inhibitory function was obtained. The theoretical disinhibition function was then calculated from this first-order function. The two are plotted in Figure 32.36, and it can be seen that there is a very good correspondence between the disinhibition function derived experimentally and the function derived theoretically. Strong evidence is therefore provided for the hypothesis that pitch memory elements are arranged as recurrent lateral inhibitory networks, similar to those observed in systems handling sensory information at the incoming level.

In summary, this set of studies demonstrates that a sequential pattern of pitches will be perceived to a greater or lesser degree of accuracy, depending in a precise and systematic fashion on the relationships between the individual pitches in the pattern. Further, such accuracy can be predicted in part from a mathematical model based on that describing lateral inhibitory interactions in sensory systems (see D. Deutsch & Feroe, 1975, for a detailed description of the present model).

There is a further effect that operates to cause disruption of pitch comparison judgments. When two test tones are separated by a sequence of interpolated tones, and a critical tone is interpolated that differs in pitch from the first test tone but that is identical in pitch to the second test tone, there is an increased tendency to misrecognize the pitch of the second test tone as identical to the first. This tendency is substantially greater when the critical tone is placed early in the interpolated

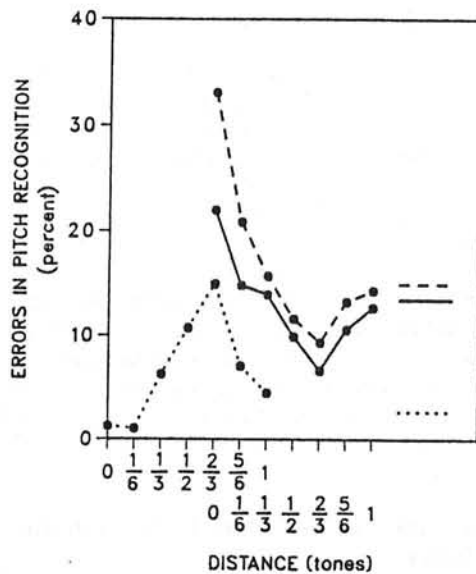


Figure 32.36. Percentage of errors in pitch recognition as a function of the pitch relationships between the first test tone and two critical interpolated tones. The dotted line plots percentage of errors in an experiment that varied the relationship between the first test tone and the critical tone. (The horizontal dotted line at right shows percentage of errors where no tones were interpolated within the critical range.) The solid line displays percentage of errors in an experiment in which a tone that was two-thirds of a tone removed from the first test tone was always interpolated. Errors are plotted as a function of the relationship between this tone and a second critical tone that was further removed along the pitch continuum. The dashed line displays percentage of errors for the same experimental conditions predicted theoretically from the lateral inhibition model. (The horizontal solid and dashed lines at right show percentage of errors obtained experimentally and assumed theoretically where no further critical tone was interpolated.) When the second critical tone was identical in pitch to the first, errors were significantly enhanced compared with the baseline condition where no further critical tone was interpolated. When the second critical tone was two-thirds of a tone removed from the first, errors were significantly reduced compared with this baseline condition. (From D. Deutsch & J. Feroe, *Disinhibition in pitch memory, Perception and Psychophysics*, 1975, 17. Reprinted with permission.)

sequence rather than late (D. Deutsch, 1972a). To explain this phenomenon, it was hypothesized that memory for the pitch of a tone is laid down both on a pitch continuum and on a temporal or order continuum. The distribution of this memory trace spreads in both directions as time proceeds, but particularly along the temporal or order continuum. Because of this spread, when a tone of the same pitch as the second test tone is included in the interpolated sequence, the subject sometimes concludes that this had been the first test tone. In other words, errors of misrecognition result from the subject's recognizing that a tone of identical pitch to the second test tone had occurred, but not being certain when it had occurred (D. Deutsch, 1972a). Further experiments have provided supporting evidence for this view (D. Deutsch, 1975d).

2.6.2. Facilitation Through Repetition in Short-Term Memory for Pitch. The effect of a critical interpolated tone on pitch comparison judgment need not be disruptive, but may instead be facilitatory. For example, if a tone whose pitch is identical to that of the first test tone is included in the interpolated sequence, comparison performance is enhanced. Subjects judge more accurately both that the two test tones are identical in pitch and also that they differ. This was demonstrated in an

experiment by D. Deutsch (1975b) in which there were three conditions. In the first, two test tones were compared for pitch when these were separated by a sequence of six interpolated tones. In the second, a sequence of four tones was interpolated instead. In the third, six tones were again interpolated, and a tone of identical pitch to the first test tone was placed in the second serial position of the interpolated sequence. The error rate was lowest in this third condition; indeed, it was significantly lower than in the condition where only four tones were interpolated.

A companion experiment showed that this facilitation effect was sensitive to the serial position of the repeated tone, being substantially greater when the repeated tone was presented early in the interpolated sequence than when it was presented late. For this and other reasons, it was concluded that the facilitation effect results from the same process as causes the errors of misrecognition described above, that is, a spread of memory distribution along a temporal or order continuum. It was hypothesized that when two such distributions overlap their overlapping portions sum, so that a stronger memory trace results. In any event, we can see that there is a strong perceptual advantage to repeating a tone in a pattern, particularly if the two occurrences of the tone are closely spaced.

2.6.3. The Influence of Relational Context on Pitch Comparison Judgments. A substrate for short-term memory for intervals was hypothesized by D. Deutsch (1975a). Such memory was assumed to be based on an array whose elements are activated by the presentation of simultaneous or successive pairs of tones. Tone pairs whose fundamental frequencies stand in the same ratio project out the same elements in the array, and tone pairs whose fundamental frequencies stand in closely similar ratios project out neighboring elements. Interactive effects take place along this array that are analogous to those occurring in the system that retains absolute pitch values. Such effects include facilitation through repetition and similarity-based interference.

This hypothesis received support from an experiment by D. Deutsch (1978b), which also showed that interval information can affect comparison judgments concerning the absolute pitches of tones with which the intervals are associated. In this experiment, subjects compared the pitches of two test tones that were both accompanied by tones of lower pitch. The test tones either were identical in pitch or differed by a semitone. The tone accompanying the first test tone was always identical in pitch to the tone accompanying the second test tone. The test tones were separated by a sequence of six interpolated tones. The tones in the second and fourth serial positions of the interpolated sequence were also accompanied by tones of lower pitch.

It was found that, when the intervals formed by the interpolated combinations were identical in size to the interval formed by the first test combination, the error rate was lower than when the sizes of the interpolated intervals were chosen at random. This shows that the system retaining interval information exhibits facilitation through repetition in the same way as the system retaining absolute pitch information. Further, when the intervals formed by the interpolated combinations differed in size by one semitone from the interval formed by the first test tone combination, the error rate was higher than when the intervals formed by the interpolated combinations were chosen at random. This shows that the system retaining interval information is also subject to similarity-based interference, analogous to that underlying memory for absolute pitch

values, and is consistent with the presence of the interval size array whose characteristics were hypothesized above.

The systems retaining absolute pitch information and interval information interact in determining pitch comparison judgments. For example, judgments of sameness or difference in the pitches of two test tones are biased by a sameness or difference in the harmonic intervals with which the test tones are associated. If the test tones differ, but are associated with the identical harmonic intervals, there is an increased tendency to misjudge them as identical. Similarly, if the test tones are identical, but are associated with different harmonic intervals, there is an increased tendency to misjudge them as different (D. Deutsch & Roll, 1974). An analogous effect holds for melodic intervals also (D. Deutsch, 1982a).

Pitch comparison judgments are influenced by relational context in yet another way. When two test tones are presented and separated by a sequence of interpolated tones, the entire configuration forms a framework of pitch relationships to which the test tones are anchored. Thus the firmer the processing of the melodic intervals within the configuration, the more accurate should be comparison judgments involving the test tones.

One would expect from findings described in Section 1.3 that melodic intervals would be more securely processed when these are of small size rather than large. It was therefore hypothesized that pitch comparison judgments would become more accurate as the average size of the melodic intervals formed by successive tones in the interpolated sequence was reduced. In an experiment to examine this prediction, two test tones were presented, separated by a sequence of six interpolated tones. In the first condition, the interpolated tones were chosen at random from a range of one octave, and they were also ordered at random. In the second condition, the interpolated tones were again chosen at random from a range of one octave, but they were arranged in monotonically ascending or descending order, so that the average size of the melodic intervals was reduced. In the third condition, the interpolated tones were chosen at random from a range of two octaves and were also ordered at random. In the fourth condition, the interpolated tones were chosen at random from a range of two octaves but were arranged in monotonically ascending or descending order. It was found, in accordance with the hypothesis, that as the average interval size formed by successive tones decreased the error rate in pitch comparison judgment also decreased (D. Deutsch, 1978a).

2.7. Contour as a Cue in Recognition of Pitch Patterns

Various studies have shown that melodies may be recognized on the basis of contour alone. Werner (1925) presented subjects with familiar melodies that were transformed onto very small scales, and he called these micromelodies. The subjects were able to recognize the micromelodies despite the fact that the interval sizes were drastically altered. White (1960) presented listeners with familiar melodies that were transformed by setting all the intervals to one semitone, so that recognition was mediated solely by the sequence of directions of pitch change. He found that above-chance performance was obtained under these conditions, showing that even preservation of relative interval size was not essential for melody recognition.

Dowling and Fujitani (1971) confirmed the role of contour in melody recognition using the following paradigm. Subjects were presented with a standard melody followed by a comparison melody. The comparison melody either was identical to the

standard, or had the same contour but was composed of different intervals, or was entirely different. The comparison melody either began on the same pitch as the standard or was transposed to a different pitch level. The authors found that, when the comparison melody was not transposed from the standard, recognition of a difference between the standard and comparison melodies was at a high level, both when contour was preserved and when it was not preserved. However, when the comparison melody was transposed from the standard, the subjects' performance levels did not differ depending on whether the transposition was exact or whether contour alone was preserved. The authors concluded that the subjects were basing their recognition judgments on contour rather than on a sameness or difference in interval size (see also Idson & Massaro, 1978; Kallman & Massaro, 1979).

2.8. Scale and Key Structure in Recognition of Pitch Patterns

A number of studies have shown that subjects will utilize their knowledge of scale structure in making recognition judgments concerning melodies. Frances (1958) reports that subjects were better able to detect a difference between two melodies when the notes of the first melody were all from one scale than when they were not. Dewar (1974) and Dewar, Cuddy, and Mewhort (1977) reached a similar conclusion. Dewar also noted that, when notes of the standard melody were all in the same scale, discrimination accuracy was poorer when all the notes of the comparison melody were also in the same scale as the notes of the standard melody than when they were not. This indicates that the subjects were basing their judgments in part on a sameness or difference in scale membership.

A related study was performed by Dowling (1978). He presented subjects with a standard melody, followed by a comparison melody, and required them to judge whether or not the comparison was a correct transposition of the standard. The comparison melody was related to the standard in any of four ways. The first type of comparison melody was an exact transposition of the standard. The second employed notes in the same diatonic scale as the standard, and for each successive relationship the number of steps up or down this scale was preserved, but the sizes of the intervals were sometimes altered in consequence. In the third type, the contour of the melody was preserved but the intervals between successive pitches were randomly selected without regard to scale. In the fourth type, the intervals between successive pitches were also randomly selected, and in addition the contour differed from that of the standard.

Dowling found that discrimination of a difference between the melodies was at a high level when the comparison melody consisted of intervals selected without regard to scale, particularly when contour was altered. However, subjects showed a strong tendency to judge as a correct transposition one in which the number of steps up or down the scale was preserved but the interval sizes were altered in consequence. This finding again shows that subjects were basing their judgments in part on a sameness or difference in scale membership.

In classical western music, when a melodic line is repeated at a different pitch level but remains in the same key, this repetition typically preserves the number of steps up or down the scale, with the result that exact interval sizes are often altered. This type of transposition generally appears correct to the listener, whereas an exact transposition that results in a departure from the scale appears incorrect (D. Deutsch, 1977).

This musical convention is likely to have evolved to exploit the tendency of the perceptual system to process pitches in terms of memberships of restricted, highly overlearned sets.

The influence of overlearned pitch structures on memory was demonstrated in another way by Krumhansl (1979). Subjects compared the pitches of two test tones that were separated by a sequence of eight interpolated tones. On half the trials the interpolated sequences were constructed so as to suggest a key. The results showed that, for such sequences, when the first test tone was in the same diatonic scale as the interpolated tones, recognition performance was better than when the first test tone was not in this scale. Krumhansl also noted that, in a tonal context, when the second test tone was not in the same diatonic scale as the interpolated tones, it was more frequently confused with a first test tone that was in this scale than when the reverse was true. She concluded that there is an instability of memory representation for tones outside a scale that has been established for the listener, so that these tones tend to become assimilated to tones that are in the scale.

Bharucha and Krumhansl (1983) studied memory for chords that were presented in sequence. They found that memory for sequences of chords that were chosen at random was poorer than that for sequences of chords that were drawn from a single key and that formed conventional harmonic progressions (see Note 15). Further, when the sequences were all in the same key, the substitution of one chord in the key for another was difficult to detect. However, a change from a chord in the key to a chord outside the key was easy to detect. In addition, more errors of confusion occurred when a chord outside the key was changed to a chord inside the key than when the opposite change occurred. This is in accordance with the idea that elements outside an established key are represented in memory in a less stable fashion and so tend to become assimilated to elements inside the key.

Differences between short-term and long-term memory in the processing of melodic information have been noted. Attneave and Olson (1971) showed that, when subjects were asked to transpose unfamiliar melodies to different pitch levels, performance was very poor, at least for those who were musically untrained. However, when a familiar sequence was employed instead, excellent performance was obtained. The finding that exact interval information is well retained in long-term memory is in accordance with general observation. Bartlett and Dowling (1980) also found experimentally that recognition of exact intervals was at a very high level for familiar melodies. In a further experiment, Dowling and Bartlett (1981) compared immediate with delayed recognition tests for a set of short, novel melodies containing tones that were not all in the same key. Although both interval and contour information were more difficult to retrieve following a delay, discrimination of exact transpositions from inexact transpositions that preserved contour did not decline with delay.

2.9. Memory for Hierarchically Organized Pitch Patterns

This section is concerned with memory for pitch information that is projected onto overlearned pitch alphabets or scales and is organized in the form of hierarchies (Section 2.5). It has been found using verbal materials that, when information was hierarchically structured and the observer was able to capitalize on this structure to produce a more efficient encoding, memory was enhanced. However, if the hierarchically structured infor-

mation was presented so as to prevent encoding in accordance with structure, an enhancement of memory did not result (Bower, 1972).

An analogous phenomenon was demonstrated in an experiment on memory for hierarchically structured tonal sequences (D. Deutsch, 1980b). Musically trained subjects listened to sequences of tones and recalled what they heard in musical notation. Examples of the presented sequences are shown in Figure 32.37. The sequence in Figure 32.37(a) consists of a higher-level subsequence of four elements that acts on a lower-level subsequence of three elements. The sequence in Figure 32.37(b) was constructed from the same set of tones as in Figure 32.37(a) but arranged in haphazard fashion. The sequences were each presented in three temporal configurations. In the first the tones were spaced at equal intervals, in the second they occurred in four groups of three (so that segmentation was in accordance with tonal structure), and in the third they occurred in three groups of four (so that segmentation was in conflict with tonal structure).

Figure 32.38 displays the percentage of tones that were correctly recalled in each serial position in the different experimental conditions. It can be seen that large effects of tonal structure and temporal segmentation were produced. For structured sequences that were segmented in accordance with structure, a very high level of recall was obtained. For structured sequences that were unsegmented, the level of recall was again very high, though slightly lower. But for structured sequences that were segmented in conflict with structure, the level of recall was considerably lower. For unstructured sequences, recall levels were lower still, but in the same range as for structured sequences that were segmented in conflict with structure. From the shape of the serial position functions, and also from analysis of transitional shift probabilities, it was demonstrated that the subjects were grouping the tones on the basis of temporal segmentation, even when this conflicted with tonal structure.

The experiment therefore demonstrates that listeners are well able to perceive hierarchical structures that are present in tonal sequences and to use such structures to produce a more efficient memory representation. However, temporal segmentation in conflict with tonal structure may destroy the capacity to exploit this information.

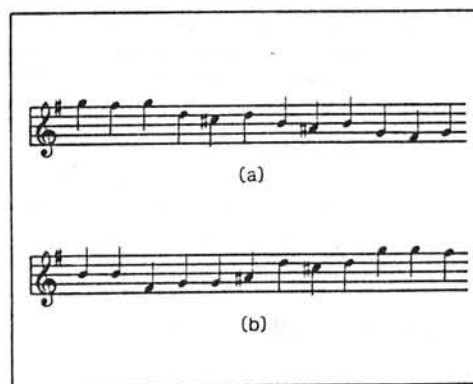


Figure 32.37. Examples of sequences employed in experiment on effects of tonal structure and temporal segmentation on recall of pitch sequences. Sequence (a) consists of a higher-level subsequence of four elements that acts on a lower-level subsequence of three elements. Sequence (b) consists of the same set of tones as in sequence (a), but arranged in haphazard fashion. (From D. Deutsch, *The processing of structured and unstructured tonal sequences*, *Perception and Psychophysics*, 1980, 28. Reprinted with permission.)

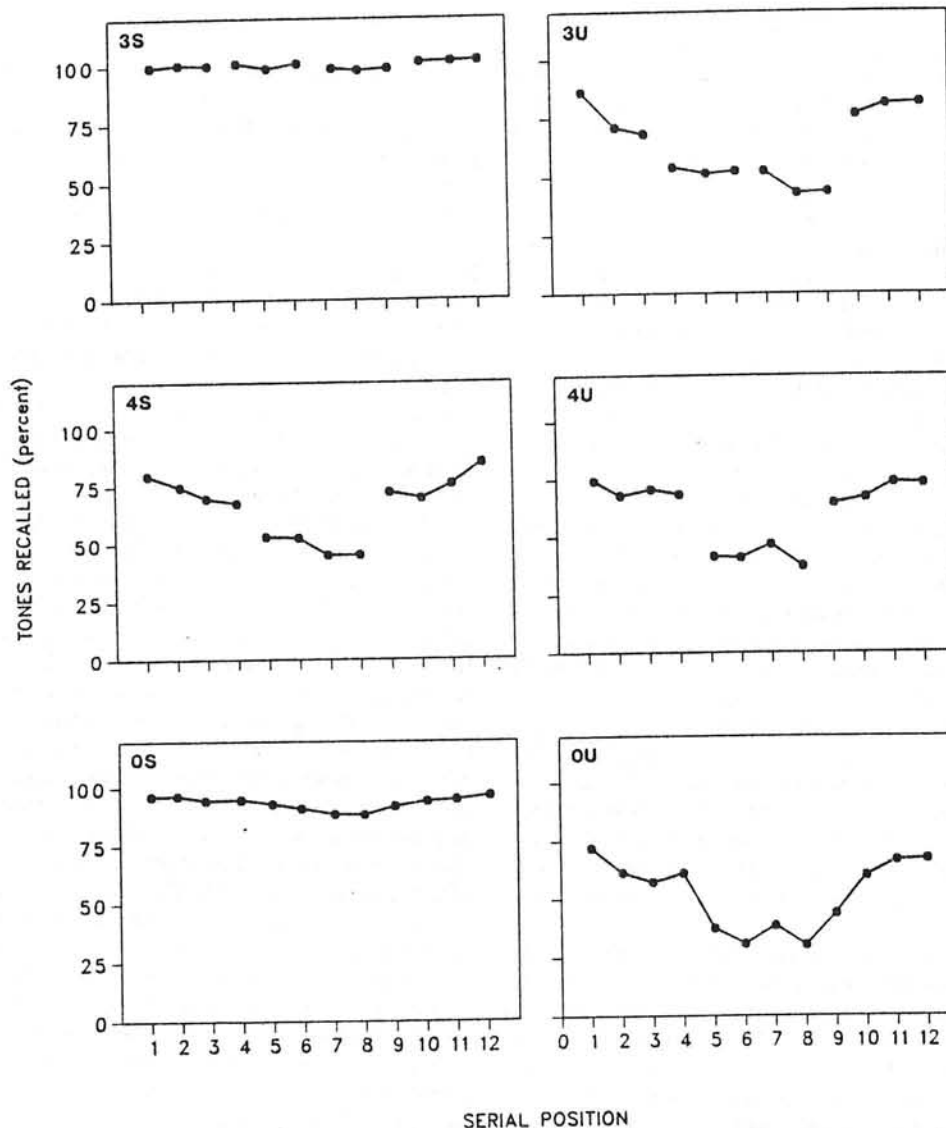


Figure 32.38. Results of experiment on the effects of structure and temporal segmentation on recall of pitch sequences. The percentages of tones correctly recalled at each serial position in the different conditions of the experiment are plotted. 3S: Structured in groups of three; segmented in groups of three. 4S: Structured in groups of three; segmented in groups of four. OS: Structured in groups of three; unsegmented. 3U: Unstructured; segmented in groups of three. 4U: Unstructured; segmented in groups of four. OU: Unstructured; unsegmented. Recall levels were very high for structured sequences, and for sequences that were segmented in accordance with structure. Recall levels were substantially lower for unstructured sequences, and for structured sequences that were segmented in conflict with structure. (From D. Deutsch, *The processing of structured and unstructured tonal sequences*, *Perception and Psychophysics*, 1980, 28. Reprinted with permission.)

3. ANALYSIS OF TIMBRE

Timbre may be described as that perceptual quality of a sound that distinguishes it from other sounds, when simple attributes such as pitch and loudness are held constant. The imprecision of this definition reflects the fact that timbre perception is a complex and little-understood phenomenon. This section focuses on the timbre of musical instrument tones, since it is with these that studies of timbre perception have been mostly concerned. However, the methods developed in such studies and the results so far obtained should ultimately prove of importance to understanding the perception of sound quality in general.

3.1. Timbre and Fourier Analysis

The classical view of timbre perception is that sound quality may be attributed entirely to the spectrum of the sound in steady state. Fourier's theorem states that a periodic waveform is defined by the amplitudes and phases of a harmonic series of spectral components. It was assumed that the ear is capable of performing such an analysis, except that it is insensitive to phase. However, others have argued that, given a periodic tone, a change in the phase relationships between the harmonics of the tone can alter perceived timbre (Mathes & R. L. Miller, 1947; Plomp & Steeneken, 1969), though this effect is generally

weak (Cabot, Mino, Dorans, Tackel, & Breed, 1976; Schroeder, 1975). Other effects of phase on the perception of tone complexes are described elsewhere in this chapter (Sections 1.1.1, 1.2.6).

In considering steady-state tones, one issue of importance is whether timbre is associated with the relationship between the frequency region of a formant and the fundamental (see Notes 2 and 16) or whether it depends on the absolute level of the formant, regardless of the frequency of the fundamental. Slawson (1968) had subjects make similarity judgments between pairs of tones in which were varied the fundamental frequencies, the two lower formant frequencies, and the higher formant frequencies. He found that, when the fundamental frequency of the second tone of each pair was an octave above the first, timbral quality was best preserved when the two lower formants were transposed by approximately 10% of the transposition of the fundamental. Plomp and Steeneken (Note 17) presented subjects with pulse trains through filters that had different center frequencies. Tones filtered at fixed frequencies were judged as more similar to each other than were tones filtered at frequencies relative to their pulse rates. Thus timbre appears to be related more to spectral envelope than to the amplitude relationships between the harmonics.

Other studies have been concerned with the role of the critical band (see Note 18) in timbre perception. Plomp and Mimpen (1968) and Plomp (1970) concluded that partials falling within the same critical band could not be distinguished from each other. For such reasons, spectra are sometimes displayed so as to take account of critical bands (Grey & Gordon, 1978; Zwicker, 1961; Zwicker & Scharf, 1965). Further, when many partials lie within the same critical band, the resultant sound is harsh (Risset & Wessel, 1982).

The classical approach assumes that timbre perception depends essentially on the spectra of tones in the steady state. A strong argument against this notion is that such spectra may be radically altered in various ways without much affecting perceived timbre. This happens, for example, when sounds are presented through a poor recording. Also, the frequency response of a normally reverberant room differs at different points in the room, with the result that sound spectra may be drastically changed. However, perceived timbre does not change dramatically as the listener shifts position in a room (Risset & Wessel, 1982).

For such reasons, recent studies of timbre perception have been concerned with the time-variant properties of tonal stimuli. Details of the initial portion of a tone, known as the attack, have been shown to exert a considerable influence on perceived timbre (Berger, 1964; Grey, 1975; Risset, 1966; Saldanha & Corso, 1964; Schaeffer, 1966; Wessel, 1973). Fluctuations in the steady-state portion and characteristics of the decay have also been found to exert an influence (Risset, 1966; Risset & Wessel, 1982; Schaeffer, 1966).

3.2. Investigation of Timbre by Analysis and Synthesis

Risset and Mathews (1969) pioneered an important technique in the study of timbre perception. Here, samples of natural instrument tones are digitized and analyzed by computer, and a set of physical parameters is thus extracted. Tones are then resynthesized by computer in accordance with these physical parameters. With this technique, the experimenter can vary systematically any parameters that he wishes and so examine the perceptual effects of these variations. For example, when

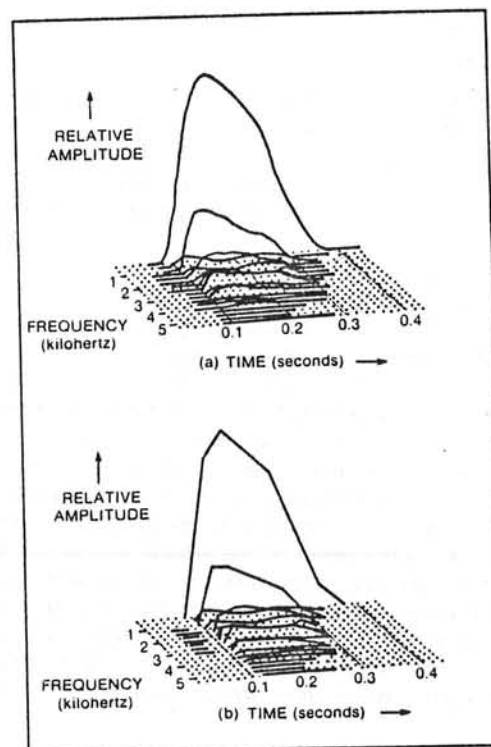


Figure 32.39. (a) Time-varying amplitude functions derived from heterodyne analysis for a bass clarinet tone, shown as an amplitude \times Frequency \times time perspective plot, with the fundamental harmonic plotted in the background. (b) Line-segment approximation to the functions plotted in (a). Both functions have been employed to resynthesize the tone, but form (b) provides considerable information reduction. Data from Grey and Moorer. (From J. C. Risset & D. L. Wessel, *Exploration of timbre by analysis and synthesis*, in D. Deutsch (Ed.), *Psychology of music*, Academic Press, Inc., 1982. Reprinted with permission.)

tones are resynthesized with a line-segment approximation to the time-varying amplitude and frequency function for the partials, very little loss of characteristic perceptual quality results, though there may be considerable information reduction (Grey & Moorer, 1977; Risset & Mathews, 1969). An example of a line-segment approximation is given on Figure 32.39.

3.3. Multidimensional Models of Timbre

Geometric models of subjective timbral space have been provided by multidimensional scaling techniques and have proved very effective. Subjects are asked to rate many pairs of tones for similarity, and their data are submitted to multidimensional scaling programs. J. R. Miller and Carterette (1975) have demonstrated that musical training affects timbral spaces in a complex fashion. In one experiment, the fundamental frequency was one of the dimensions varied. Due to the overwhelming salience of this dimension no differences were found that depended on training. However, when in a second experiment the fundamental frequency was held constant, differences that depended on training emerged.

Wessel (1973) employed tones of identical fundamental frequency and duration, which were taken from nine orchestral instruments. He concluded that instrumental timbre could be ordered along two perceptual dimensions. The first related to the distribution of energy in the steady state. Tones with more energy at high frequencies appeared at one end of this dimension,

and tones with more energy at low frequencies appeared at the other end. The second dimension related to tonal onset patterns. Tones whose low-order harmonics emerged more rapidly appeared at one end of this dimension, and tones whose high-order harmonics entered more rapidly appeared at the other end.

Grey (1975) performed an experiment that employed 16 instrument tones that were resynthesized by computer and equated for pitch, loudness, and duration. His data were most consistent with a three-dimensional solution. The first dimension related to the tones' spectral energy distribution. Tones with narrow bandwidths and a concentration of low-frequency energy appeared at one end, and tones with wide bandwidths and less concentration of low-frequency energy appeared at the other end. A second dimension was related to the distribution of energy in the attack segment. At one end, tones displayed high-frequency, low-amplitude energy in the attack, and at the other end there was no high-frequency precedent energy in the attack. For the third dimension, two alternative interpretations were proposed. The first was that this dimension related to the form of onset–offset patterns. The second hypothesis was that this was a cognitive dimension, along which the tonal stimuli were

arranged according to instrument family (e.g., brass, strings, woodwinds). This three-dimensional space is displayed in Figure 32.40.

3.4. Role of Context in Timbre Perception

The importance of a cue to timbre perception has been found to depend on the context in which this cue is embedded. In the perception of trumpet tones, details of the attack are more important for long tones than for short ones (Risset, 1966). In the perception of piano tones, the shape of the initiation of the decay is important to how the attack portion is perceived. Further, when tones are presented in close succession, their timbres are perceived differently than when these tones are presented in isolation. This was demonstrated by Grey (1978), who presented computer-synthesized tones either in isolation or in single or multivoiced musical contexts. He found that timbre discrimination was more difficult in multivoiced contexts, and that a single-voiced context caused a perceptual enhancement of spectral differences relative to isolated tones, while the presentation of isolated tones allowed listeners to compare temporal details more clearly.

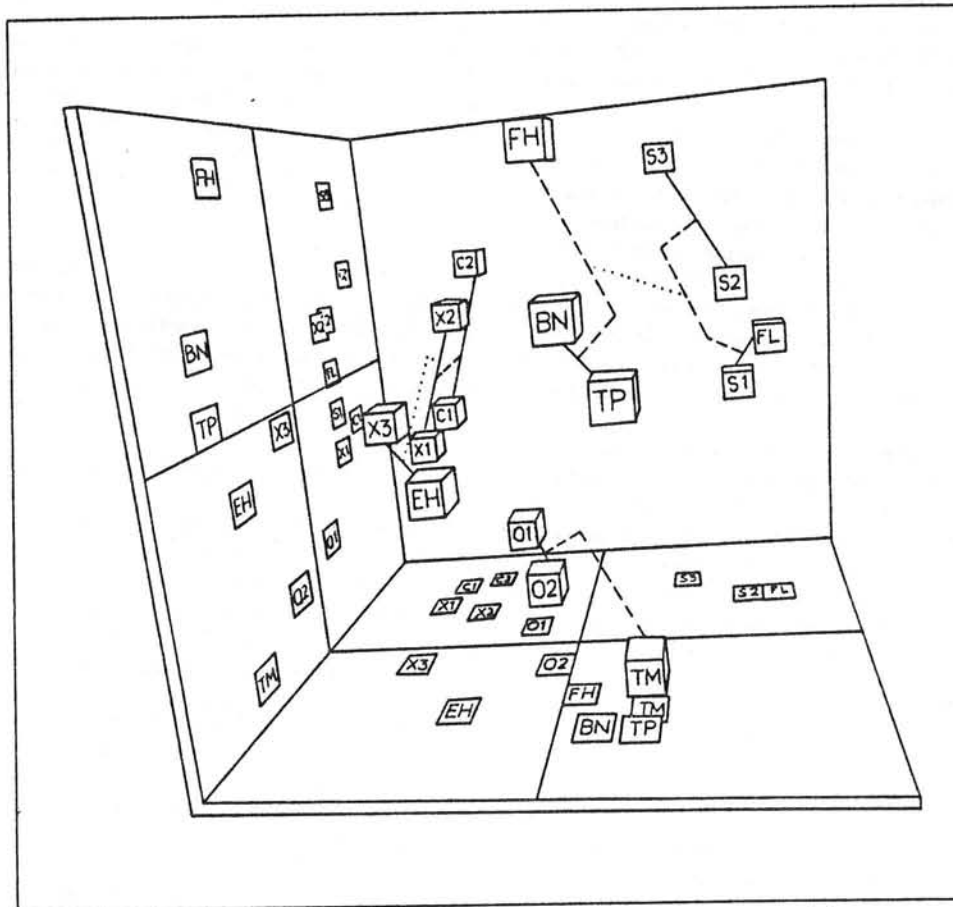


Figure 32.40. Three-dimensional display of similarities between different instrument timbres generated by multidimensional scaling. O1, O2, = oboe; C1, C2 = clarinets; X1, X2, X3 = saxophones; EH = English horn; FH = French horn; S1, S2, S3 = strings; TP = trumpet; TM = trombone; FL = flute; BN = bassoon. The proximities of the instruments to each other in this three-dimensional space indicate the extents of their perceived similarity. (From J. M. Grey, *Timbre discrimination in musical patterns*, *Journal of the Acoustical Society of America*, 1978, 64. Reprinted with permission.)

4. PERCEPTION OF TEMPORAL RELATIONSHIPS

This section is concerned with the ways in which the listener abstracts temporal relationships from patterns of sound. First, we shall examine the perception of temporal order of two or more events. Second, we shall consider the evidence concerning grouping mechanisms for temporally patterned stimuli. Third, we shall examine the encoding of rhythmic patterns.

4.1. Perception of Temporal Order

4.1.1. Modes of Order Perception. Following Hirsh (1974) and R. M. Warren (1974b), we may distinguish three basic modes of order perception in hearing, each loosely associated with a different range of temporal values. First, with very small time intervals separating the onsets of two events (under 10 msec), there results a single fused sound. Differences in the quality of this sound then serve as bases for temporal order judgment. For example, small interaural time differences between onsets of dichotically presented sounds give rise to lateralization cues (Babkoff, 1975). Also, with sounds presented monaurally or dichotically, spectral differences resulting from asynchrony of onset give rise to changes in sound quality (Patterson & Green, 1970). At somewhat longer time differences, order judgments may be based on the figural or Gestalt properties of a sound sequence, while the listener may still be unable to name the order of individual events within the sequence (R. M. Warren, 1974a). Finally, when sufficiently long time intervals separate the onsets of successive events, the listener can make order judgments by an item-by-item analysis of the pattern components.

Estimates of the ranges of temporal values associated with these three stages have been found to vary considerably depending on the training of the subject, the experimental paradigm, and the stimulus parameters employed (R. M. Warren, 1982). A major problem in providing such estimates is that one type of judgment can easily be disguised as another. For example, the subject can learn to associate one sound quality with the judgment *A followed by B*, and a different sound quality with the judgment *B followed by A*. Similarly, the subject can learn to attach the label *single sound event* to one sound quality, and the label *two sound events* to a different sound quality.

4.1.2. Perception of the Order of Two Events. Hirsh (1959) employed pairs of sounds drawn from a variety of tones, hisses, and clicks to establish the minimum time between the onsets of these sounds required for their order to be correctly reported 75% of the time. He concluded that this minimum was around 20 msec, with little variation due to the different types of sound presented, or to their levels. The subjects in this experiment were highly trained, and they were allowed to listen to the sound pairs as often as they wished before reaching a decision.

Later Hirsh and Sherrick (1961) investigated the ability to order two auditory events, two visual events, and two tactile events. Pairs of stimuli drawn from two sensory modes were also employed. As in the previous study, trained observers were used, and they were allowed to inspect the stimulus pairs as often as they wished. Thresholds of around 20 msec were again obtained, regardless of stimulus modality. The authors concluded that the value of approximately 20 msec represents a fundamental limit for the perception of temporal order when special modality-specific conditions are excluded.

Other studies have investigated the effect of repeated presentations and training on judgment of the order of tone pairs.

Hirsh and Fraisse (1964) employed untrained subjects and presented them with a single stimulus pair on each trial. For accurate identification of order, a difference of around 60 msec was required when a sound followed a light, and of around 100 msec when a light followed a sound. Later, Gengel and Hirsh (1970) found effects of both number of presentations and training. With untrained subjects, single trials yielded thresholds of about 45 msec, which decreased to below 30 msec following roughly 10 sessions. Repeated presentations yielded thresholds of about 25 msec which decreased with training to roughly 18 msec.

Broadbent and Ladefoged (1959) also reported differences depending on continued listening. Pairs of sounds presented were buzzes, hisses, or tones. During the first few presentations, correct ordering was not achieved even at 150 msec durations. However with repeated listening, accurate perception became possible at 30-msec durations. These authors remark that discriminations were made here on the basis of "quality" rather than "perceived order." R. M. Warren (1974b) has found that, with special training in which subjects were presented with sequences that speeded up gradually, correct naming of the temporal order of spectrally different sounds was possible with separations as low as 5 msec.

4.1.3. Perception of the Order of Three or More Events. Hirsh (1976) studied the perception of the order of three heterogeneous stimuli: a sound, a light, and a vibrotactile stimulus. In one experiment, subjects were asked to identify which stimulus occurred at the beginning of a pattern. Performance on three-element patterns was found to be poorer than on two-element patterns. Hirsh concluded that the third element had the effect of impairing judgment as to which of the two prior elements had come first. Analogously, Divenyi and Hirsh (1975) have found that identification of the temporal order of three 20-msec tones was depressed when a fourth tone, which was irrelevant to the task, was added to the sequence.

In a second experiment, Hirsh asked subjects to identify which of six possible permutations of a set of sounds had been presented. With 15 msec between onsets of successive sounds, performance was barely above chance; performance rose with 45 msec between onsets, and again with 150 msec between onsets.

In a third experiment, Hirsh studied the effects on order identification of repeating a stimulus pattern when there were clear breaks between repetitions. This procedure resulted in an improvement in performance. In a final experiment, the stimuli were presented in continuously cycling fashion, and performance levels here were substantially poorer. Phenomenologically, the pattern was perceived as three distinct trains or streams, corresponding to the three presentation modes. This finding is comparable to those on continuously cycling sound patterns to be described in Section 4.1.4.

4.1.4. Order Perception in Continuously Cycling Sound Patterns. When several disparate sound elements are cyclically presented, judgments of the orders of these elements become surprisingly difficult. This was shown in one situation by Bregman and Campbell (1971). They presented subjects with repeating sequences of six 100-msec tones, such that tones from a low frequency range alternated with tones from a high frequency range, with about 1 1/2 octaves separating the ranges. Following each sequence, a three-tone pattern was presented in isolation, and the subjects judged whether these three tones had occurred in the same order and spacing within the six-tone

sequence. Judgments were above chance only if the three test tones were all in the same frequency range.

A substantial difficulty in ordering appears when sounds of differing quality are presented in continuously cycling fashion. R. M. Warren, Obusek, Farmer, and R. P. Warren (1969) constructed such a sequence from a high tone, a hiss, a low tone, and a buzz. At presentation rates of 200 msec, subjects were unable to name the orders of these sounds. For correct ordering to be achieved, it was necessary to increase the duration of each sound beyond 500 msec.

R. M. Warren (1974a) showed, however, that, when subjects were not required to name the orders of the component sounds but rather to make "same-different" comparisons between two repeating sequences, performance levels were considerably superior. This was true even for unpracticed subjects. Such sequences can therefore be identified on a wholistic basis at presentation rates at which element-by-element order judgments cannot be made.

R. M. Warren (1974b) examined the ability of trained subjects to attach learned descriptive levels to cycling sequences of four auditory elements. The sequences were first presented at long durations so that correct naming was possible and were then gradually speeded up. By a series of such transfers, subjects were enabled to make correct judgments with durations as short as 10 msec per item. Correct naming was achieved here through a disguised wholistic pattern recognition. An effect of this sort could account for the finding that cycling sequences of verbal items are correctly ordered at considerably faster rates than cycling sequences of unrelated sounds (Dorman, Cutting, & Raphael, 1975; Thomas, Cetti, & Chase, 1971; Thomas, Hill, Carroll, & Garcia, 1970; R. M. Warren & R. P. Warren, 1970).

4.1.5. Theories of Order Perception. We may enquire into the nature of the process that enables the rapid reconstruction of the order of components of familiar sound sequences. R. M. Warren (1974b) proposed that such recognition is mediated by a two-stage process. At the first stage, the sequence is recognized in a wholistic fashion, that is, as a "temporal compound" that can be distinguished from other compounds without being analyzed in terms of its components. At the second stage, there occurs an item-by-item analysis of the components of this compound together with their orderings.

Another proposal was made by Wickelgren (1969b, 1976) for the case of speech sounds; however, it may be applied to nonspeech sounds also. He suggested that the correct ordering of the components of speech sounds is based on an encoding of a set of context-sensitive elements that however need not themselves be ordered. Thus he proposed for example that the word *struck* is encoded not as the ordered set of phonemes /s/, /t/, /r/, /u/, /k/, but rather as the unordered set of context-sensitive allophones /#^st/, /s^tr/, /t^ru/, /r^uk/, /u^k#/ . In this way, each of these context-sensitive elements contains some local information as to how this element is ordered in relation to the other elements. Information concerning the order of these elements can thus be derived from such an unordered set.

Sternberg and Knoll (1973) have proposed a model of temporal order judgment. This model deals only with the case in which information arrives via totally independent channels, and in which no relational information concerning successive stimuli is involved. According to this model, a "decision function" converts a difference between central arrival times of two sensory signals into a temporal order judgment. The psychometric function for order is regarded as a distribution function and is represented additively in terms of the central arrival latencies and

the decision function. As the authors point out, the assumption of channel independence is critical to the model, so that when there is any interaction between input channels the model cannot be applied.

4.2. Perception of Rhythm

4.2.1. Subjective Rhythmic Grouping. When asked to execute a repetitive response sequence, such as tapping, most individuals will perform the task at a characteristic rate, termed by Fraisse (1982) the "spontaneous tempo." This rate is most commonly around 600 msec, though large individual differences have been found, with values roughly in the range of 200–1400 msec (Fraisse, Pichot, & Clarouin, 1949). Individuals are generally very consistent in their rates of responding, both within and across trials.

The rate at which a sequence of auditory or visual events appears most natural to the observer is termed by Fraisse (1982) the "preferred tempo." Most frequently this rate has been determined to be around 600 msec, and individuals appear very consistent in their preferences (Fraisse, 1982). In a related study, Handel and Oshinsky (1981) presented subjects with polyrhythms consisting of two conflicting pulse trains and asked them to tap along with each pattern to indicate the beat (i.e., to pick one of the two pulse trains as the more salient). An interelement timing "window" of roughly 200–800 msec for the choice of pulse trains was observed; only rarely did subjects pick a pulse train that lay outside this temporal range (see also Handel, 1984).

A sequence of identical sounds that occur at regular intervals will appear to the observer as grouped into subsequences, each consisting of an accented element followed by one or more unaccented elements (Bolton, 1894; Woodrow, 1909). Subjective grouping of this nature occurs at presentation rates ranging from about ten per second to one every 2 sec (Fraisse, 1956, 1982; Vos, 1973). This range of temporal values correlates well with the distribution of melodic tempos in western traditional music. It is also interesting to note that tempo, measured by the number of consecutive notes per unit time, appears to be distributed within this temporal range in the music of widely divergent cultures (Figure 32.41).

4.2.2. Grouping by Temporal Proximity. The division of a sequence of elements into subsequences is readily achieved by increasing the size of temporal gap following the last element of each subsequence. Such grouping has a number of consequences. Povel and Okkerman (1981) presented subjects with sequences of pure tones of identical frequency, amplitude, and duration but separated by two alternating time intervals. Such sequences were perceived as repeating groups of two tones each. When the alternating intervals differed by roughly 5–10%, the first tone of each group was heard as accented. When this difference was increased, the accent was heard instead as on the second tone of each group, and the latter accent appeared stronger. Under such conditions, when subjects were asked to adjust the amplitude of the first tone so that the two tones in the group sounded equally loud, they increased this amplitude by roughly 4 dB. This effect was however produced only when the within-group interval was no longer than roughly 250 msec.

Grouping by temporal proximity has been shown to exert a strong influence on the perception of pitch patterns. Handel (1973) investigated the identification of repeating auditory patterns that consisted of dichotomous elements differing in pitch. These were temporally segmented, and it was found that com-

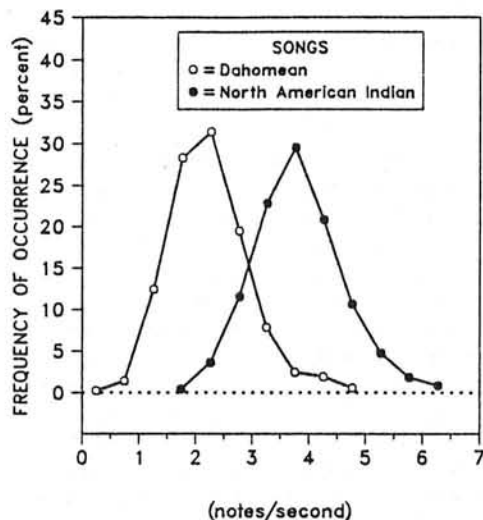


Figure 32.41. Relative frequencies of occurrence of tempos in the songs of two cultures that diverge extremely in their average tempos. Note that the shapes of the two distributions are remarkable similar, and that the total range covered is not much larger than the range over which spontaneous rhythmic groupings are formed. Data from Kolinski (1959). (From D. Deutsch, *The psychology of music*, in E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 10), Academic Press, Inc., 1978. Reprinted with permission.)

patible segmentation (e.g., an eight-element pattern segmented into groups of two) produced excellent performance, whereas incompatible segmentation (e.g., an eight-element pattern segmented into groups of three) produced poor performance (see also Handel & Yoder, 1975). Dowling (1973c) presented five-tone sequences that were separated by pauses, followed by a single five-tone sequence for recognition. Performance was superior when the sequence to be recognized had been presented in a single temporal segment than when it had not. Further, D. Deutsch (1980b) investigated recall of hierarchically structured tonal sequences. These were segmented by pauses, and it was found that, when the pauses were in accordance with tonal structure, performance levels were high. However, when the pauses conflicted with tonal structure, performance levels dropped considerably. Subjects were therefore shown to be grouping the sequences on the basis of temporal proximity rather than tonal structure when the two were placed in conflict. Such results parallel those obtained by others on recall of strings of verbal materials. When such strings are temporally segmented, recall tends to be in accordance with their temporal grouping, and this effect can be so strong as to mask grouping on the basis of meaning (Bower & Winzenz, 1969; McLean & Gregg, 1967; Mueller & Schumann, 1894).

When elements are grouped by temporal proximity, ease of processing may differ depending on the location of elements within the group. Divenyi and Hirsh (1978) presented subjects with rapid sequential patterns consisting of three tones. The tones within each pattern could occur in any of six permutations, and subjects were asked to identify on each trial which permutation had been presented. These three-tone patterns were embedded as subsequences in longer sequences consisting of seven or eight tones. It was found that identification performance was enhanced when the background and test patterns were in different frequency ranges. In addition, performance levels varied considerably depending on where in the sequence the test pattern was placed. Highest performance levels occurred

when the test pattern was at the end of the sequence. Performance levels were relatively high when the test pattern was at the beginning of the sequence; however, they were close to chance when the test pattern occurred in the middle of the sequence. Thus both frequency separation and temporal separation were found to reduce interference from the background tones. This may be related to an early finding that a single tone embedded in a sequence was particularly salient when it was the highest or the lowest in frequency, or when it was in the first or last temporal position (Ortmann, 1926).

4.2.3. Grouping by Accent. A second way in which a sequence of elements may be subdivided is by the imposition of accents. An element is perceived as accented when it is marked for attention in some fashion. For example, it might differ from other elements in loudness, in pitch, or in timbre. In general, accented elements combine with adjacent elements to form groupings, and they also combine with each other to form groupings at higher structural levels (see Section 4.2.6).

4.2.4. Grouping by Other Principles. As described in detail in Section 1.3, the division of sequences into subsequences has also been demonstrated along several other lines. For example, there is a strong tendency to group together sequentially presented elements that are proximal in pitch (Bregman, 1978; Bregman & Campbell, 1971; Dowling, 1973a; Van Noorden, 1975). Further, when adjacent elements in a sequence combine to form unidirectional pitch patterns, they are likely to be perceived as a group. This follows the principle of good continuation (Bregman & Dannenbring, 1973; Divenyi & Hirsh, 1974; Nickerson & Freeman, 1974; Van Noorden, 1975; R. M. Warren & Byrnes, 1975). Elements are also perceptually grouped by similarity of sound quality (R. M. Warren, 1974a; R. M. Warren, Obusek, Farmer, & R. P. Warren, 1969) and by amplitude (Dowling, 1973a; Van Noorden, 1975). Repetition of a subsequence within a sequence induces the listener to group the elements of the subsequence together. This is true even if the repetition is at an abstract level; for example, if a sequence of pitches is repeated in transposed form (D. Deutsch & Feroe, 1981; Simon & Sumner, 1968).

4.2.5. The Run Principle and the Gap Principle. Garner and his associates have examined perceptual organization of temporal patterns using the following paradigm. A basic pattern consisting of dichotomous elements (such as a high tone and a low tone) was repeated continuously without pause. This basic pattern thus gave rise to as many specific patterns with different starting points as were events within the patterns. Thus for example the pattern X X X O X O X O could alternatively be described as X X O X O X O X, or as X O X O X O X X, and so on (Garner, 1974). The issue of interest was which of these specific patterns the listener would tend to perceive.

Royer and Garner (1966) employed patterns consisting of eight events and found that the number of specific patterns perceived differed from one basic pattern to another, and further that the difficulty in pattern perception increased with the number of perceived alternatives. Another finding was that patterns beginning or ending with the longest run were always preferred. Later, Royer and Garner (1970) provided further clarification of these organizing principles, using patterns of nine events. When these patterns were described in terms of run lengths (e.g., the pattern 2115 was either X X O X O O O O or O O X O X X X X) it was found that the most preferred organizations were those that provided the best balance, with runs at both ends of the pattern (such as 2115 or 4113). Another

organizing principle was temporal progression of run length (such as 5211 or 1134). Furthermore, when a specific pattern was preferred, its temporal reversal was also preferred (e.g., 5211 and 1125). This was true both when the specific patterns and their reversals came from the same basic pattern and when they came from different basic patterns.

To analyze these principles further, Preusser, Garner, and Gottwald (1970) constructed patterns of events of a single type interleaved with gaps. A two-element pattern could thus be described as a composite of the two such one-element patterns. For example, the pattern X X X O O X O O O can be described as the composite of X X X ' ' X ' ' ' and ' ' O O ' O O O. Two principles were found to operate for such patterns: Preferred organizations either began with the longest run or ended with the longest gap. Further, when the run and the gap principles were placed in conflict, the gap principle dominated in determining which pattern was perceived. It was further noted that, with two-element patterns, subjects exhibited strong preferences for one element to serve as "figure" and the other element as "ground"; thus preferences for two-element patterns could be interpreted in terms of their associated one-element patterns (Garner, 1974).

An effect of presentation rate was also noted in this series of experiments (Garner & Gottwald, 1968). At slow rates, patterns starting at nonpreferred points were considerably more difficult to process than patterns starting at preferred points. This difference disappeared, however, at high presentation rates (Figure 32.42). The difficulty in processing patterns at slow rates was hypothesized to be due to an interference effect produced by verbal encoding of the patterns.

4.2.6. Rhythmic Hierarchies. When a listener spontaneously groups a sequence of regularly recurring events into

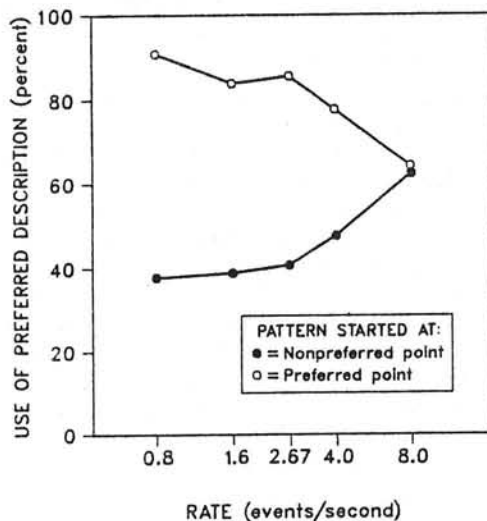


Figure 32.42. Use of the preferred description of two-element temporal patterns, as a function of presentation rate. Patterns started at the beginning of either the preferred description or a nonpreferred description. For patterns starting at the preferred point, use of the preferred description was highest at the slowest presentation rate and gradually declined as the presentation rate increased. For patterns starting at the nonpreferred point, use of the preferred description was lowest at the slowest presentation rate and rose as the presentation rate increased. (From W. R. Garner, *The processing of information and structure*, Lawrence Erlbaum Associates, 1974. Reprinted with permission.)

subsequences, such organization may occur simultaneously at more than one structural level (Vos, 1973; Woodrow, 1951). For example, the listener may perceive groups of four elements with the major accent on the first and a minor accent on the third. Such spontaneous organization indicates that the system underlying perception of rhythm is hierarchical in nature.

An experiment by Perkins (1974) provides further evidence for this view. Subjects were required to estimate the number of taps occurring in sequences in which the first of every four taps was stressed, and the first of every 16 taps was doubly stressed. Errors differed from correct responses more often by multiples of four and 16 than by adjacent numbers. Perkins concluded that the subjects were structuring the sequences hierarchically with reference to the imposed accents. A related study was performed by Sturges and Martin (1974). Here, subjects were presented with continuously repeating sequences of 14 or 16 dichotomous elements. These were patterns of seven or eight events that either repeated exactly or were altered slightly on repetition. The subjects were asked to recognize the sequences that contained exact repetitions. Patterns that exhibited a simple hierarchical structure were better recognized than those that did not. Further, eight-event patterns that were hierarchically structured were better recognized than seven-event patterns, even though the former patterns contained more events.

Other evidence derives from experiments in which subjects were required to generate temporally patterned sequences. Performance levels on such tasks vary substantially depending on the type of pattern to be produced. Isochronous sequences are generally produced at a very high degree of accuracy (N. R. Bartlett & S. C. Bartlett, 1959; Michon, 1967; M. Treisman, 1963; Wagner, 1971; Wing & Kristofferson, 1973a, 1973b). However, irregular sequences are typically produced only poorly, with subjects tending to generate interresponse intervals that are either approximately identical or stand roughly in a ratio of 2:1 or 3:1 (Fraisse, 1982; Montpellier, 1935). This again is evidence for hierarchical structuring of temporal relationships. Related evidence comes from Sternberg, Knoll, and Zukofsky (1982). When highly trained musicians were asked to subdivide a time interval repeatedly so as to produce a given fraction, high performance accuracy was achieved for a division of one-half, with poor performance accuracy for other divisions, such as one-eighth, one-seventh, and one-sixth. This is evidence for the subdivision of time intervals into simple fractions involving small integers.

Further supporting evidence for hierarchical structuring of temporal patterns comes from Summers (1975). Subjects were presented with continuously cycling sequences of nine lights, and they learned to respond to these by pressing nine keys. Later, they learned to produce this pattern with one of two temporal structures. For one group of subjects, this structure consisted of the repetitive presentation of two short intervals followed by a long interval. For another group of subjects, this consisted instead of two long intervals followed by a short interval. Later still, the subjects were told to respond as rapidly as possible. Under these conditions, the "short-short-long group" maintained the acquired temporal pattern in their responses. However, for the "long-long-short group" the original temporal pattern gradually disappeared. Since the short-short-long pattern had a simple hierarchical description but the long-long-short pattern had not, this result may be explained by assuming that the subjects were invoking hierarchical structure in performing the task.

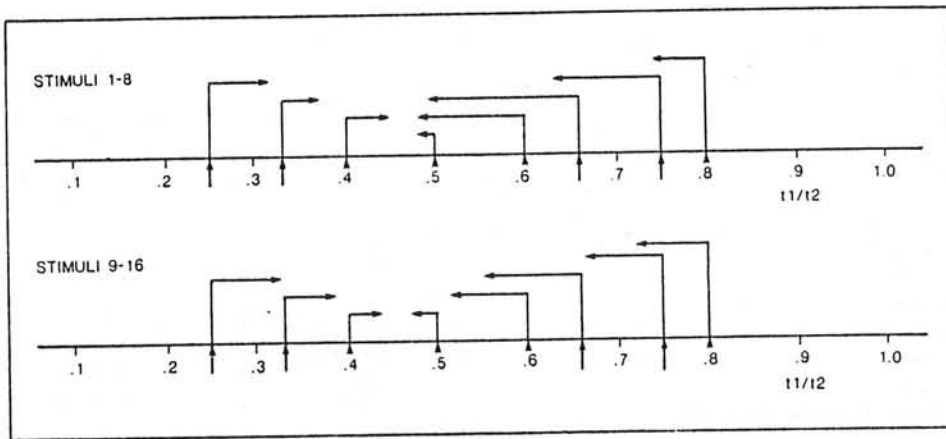


Figure 32.43. Distortion of duration ratios as found in reproduction of simple temporal patterns. Duration ratios of the stimuli are indicated by arrows pointing to the abscissa displaying the ratio continuum. The distortion, averaged over subjects, is indicated by the endpoints of the horizontal arrows. (From D. J. Povel, Internal representation of simple temporal patterns, *Journal of Experimental Psychology: Human Perception and Performance*, 7. Copyright 1981 by American Psychological Association. Reprinted with permission.)

Povel (1981) has studied the reproduction of repeating temporal patterns consisting of two or more intervals whose durations stood in various ratios to each other. He obtained substantial differences in reproduction accuracy depending on these ratios. When simple patterns were presented, those with durations standing in a ratio of 2:1 were accurately reproduced; those with durations standing in other ratios were reproduced less well. Systematic deviations in performance were found, which tended toward a ratio of 2:1 (Figure 32.43). When more complex patterns were presented, those that could be simply described in terms of a hierarchical model (to be described later this section) were well reproduced, and those that could not be so described were poorly reproduced. Povel concluded that the accurate production of temporal patterns requires that these be internally represented in accordance with his model.

On Povel's model, the first step in processing a temporal sequence consists of segmenting it into equal intervals (which he termed beat intervals) bordered by events. As a second step, the beat intervals may themselves be divided into equal intervals of two, three, or more units. These smaller intervals may in turn be subdivided into equal intervals, and so on. An illustration of this model is shown on Figure 32.44.

Povel's model is based on the formal description of rhythm in tonal music. As stated by Westergaard (1975), time in tonal music is best conceived of in terms of a set of equally spaced reference points, or beats. The time span between two primary beats is called a measure. The measure is itself divided into two or more equal time spans, bordered by secondary beats. The number of secondary beats dividing the measure defines the meter. For example, one secondary beat dividing the measure into equal parts produces duple meter, and two secondary beats produce triple meter. These smaller time spans may themselves be divided into equal parts; and so on. The symbols for tones and rests in tonal music were developed to reflect this hierarchical scheme. Thus ♩ denotes a whole note, ♪ a half note, ♫ a quarter note, ♮ an eighth note, ♯ a sixteenth note, and so on. It appears reasonable to assume that the hierarchical structure of rhythm in tonal music has evolved to capitalize on the characteristics of our processing mechanisms (see also Cooper & L. B. Meyer, 1960; Lerdahl & Jackendoff, 1983a; Yeston, 1976).

An important aspect of the perception of rhythmic patterns is generalization across tempo. Just as the equivalence of a transposed melody will be recognized provided that the frequency ratios formed by successive tones are preserved, so will the equivalence of a rhythm be recognized in the face of altered presentation rates provided that the ratios between successive temporal elements are preserved. Thus for example the pattern (in msec) 400–300–100–800 will be perceived as equivalent to the pattern 300–225–75–600. (Such generalization only holds true within a range of presentation rates; at very fast rates the elements fuse to produce timbral patterns; and at very slow rates temporal relationships are not securely perceived, as described in Section 4.2.1.) This invariance may be readily ex-

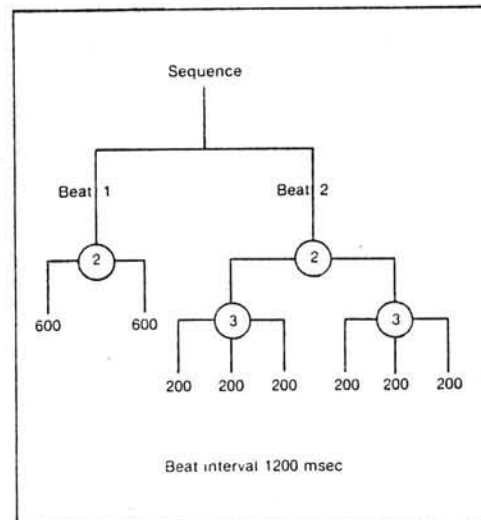


Figure 32.44. Proposed encoding of a temporal sequence. The sequence is segmented into equal intervals (called beat intervals) bordered by events. These intervals may themselves be divided into equal intervals, and so on. The model is based on the formal description of intervals in tonal music. (From D. J. Povel, Internal representation of simple temporal patterns, *Journal of Experimental Psychology: Human Perception and Performance*, 7. Copyright 1981 by American Psychological Association. Reprinted with permission.)

plained on a model of rhythmic processing that assumes that we encode relationships between temporal values across different hierarchical levels (Note 19).

Given that hierarchical structures are invoked in processing single rhythmic patterns, we may ask what happens when two patterns are processed simultaneously. To take the simplest case, we may ask whether two parallel isochronous sequences may be processed independently, or whether the observer synthesizes such sequences into a single hierarchical configuration. From one standpoint, since isochronous sequences are descriptively very simple, we might expect that the observer should find no difficulty in processing two in parallel. On the other hand, the experimental results showing the importance of hierarchical relationships in the processing of single sequences indicate that ease of processing of parallel sequences might also depend on the relationships between the temporal elements involved.

Let us take the simplest case, that of a 3-against-2 polyrhythm. If this pattern is repeated every 1200 msec, it may be described as consisting of a sequence of three 400-msec intervals together with a sequence of two 600-msec intervals. An alter-

native description may be advanced in terms of a hierarchy such as shown in Figure 32.45(a). Here the 1200-msec time span is first divided into three 400-msec segments, and these are each divided into two 200-msec segments. When the pattern (R/L, -) (R, L) (R/L, -) is associated with the lowest-level structure, a 3-against-2 polyrhythm is produced.

If this hierarchical model is correct, then ease of production of polyrhythms should depend on the complexities of their hierarchical representations. A 5-against-4 polyrhythm would thus be represented as in Figure 32.45(b). This representation is considerably more complex than the representation for the 3-against-2 polyrhythm, both in terms of the numbers of structures involved and in terms of the sizes of these structures.

An experiment was performed to test this hypothesis (D. Deutsch, 1983a). Two pulse trains were presented in parallel through earphones, one to each ear. The subjects were asked to tap with the right forefinger in synchrony with the pulse train delivered to the right ear, and with the left forefinger in synchrony with the pulse train delivered to the left ear. All pulse trains were isochronous, and the relative durations of the intervals associated with the right and left pulse trains were

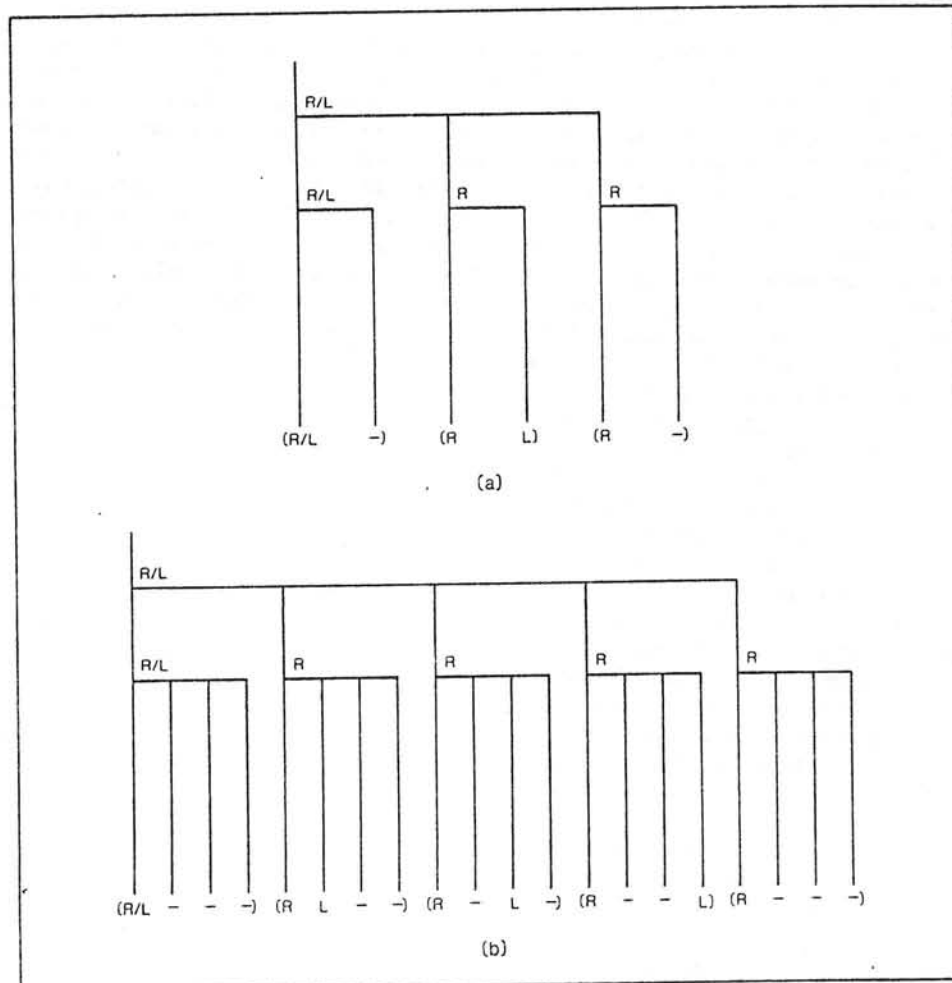


Figure 32.45. Representation of polyrhythms in terms of hierarchies. (a) A 3-against-2 polyrhythm; (b) a 5-against-4 polyrhythm. The representation in (a) is considerably more complex than that in (b). This accords with the finding that accuracy in generating polyrhythms declines with an increase in the complexity of their representation. (From D. Deutsch, *The generation of two isochronous sequences in parallel*, *Perception and Psychophysics*, 1983, 34. Reprinted with permission.)

systematically varied. A 1200 msec interval between pulse onsets served as the base interval. This interval was divided by 1 (resulting in a 1200-msec onset-to-onset interval), by 2 (resulting in a 600-msec onset-to-onset interval), by 3 (resulting in a 400-msec onset-to-onset interval), by 4 (resulting in a 300-msec onset-to-onset interval), and by 5 (resulting in a 240-msec onset-to-onset interval). Thus when these pulse trains were simultaneously presented, both simple rhythms (1 against 2, 1 against 3, and so on) and polyrhythms (2 against 3, 2 against 5, 3 against 4, 3 against 5, and 4 against 5) were produced.

It was found that, for each rate of tapping with one hand, accuracy was substantially dependent on the rate at which the other hand was tapping. When simple rhythms were produced, standard deviations were very low. They were significantly higher for the 3-against-2 polyrhythm, higher still for the 5-against-2 polyrhythm, and yet still higher for the more complex polyrhythms. It was concluded that, in processing two isochronous sequences in parallel, subjects combined these sequences to produce a single hierarchical configuration, and that accuracy in performance was a function of the complexity of this hierarchy.

5. SUMMARY

In this chapter, auditory pattern recognition has been examined from several points of view. First explored were the ways in which listeners group together components of a complex sound spectrum so that they perceive either a single sound or multiple simultaneous sounds. Next explored were the principles by which listeners form groupings from a series of single sounds that occur in rapid succession. The role of selective attention in the formation of auditory groupings was discussed in this context.

The second part of the chapter was concerned with the principles whereby auditory shapes are analyzed by listeners so as to give rise to perceptual equivalences and similarities. The evidence for low-level perceptual features based on octave equivalence and interval equivalence was discussed, as was the evidence for contour as a perceptual feature. The ways in which such low-level features are combined at higher levels were then explored in detail. The encoding of pitch patterns in the form of hierarchies was examined, as was the involvement of short-term and long-term memory in the recognition of pitch patterns.

The third part of the chapter concerned the principles underlying recognition of timbre or sound quality. Emphasis was placed on the time-variant properties of the signal, in contrast with earlier approaches that attempted to define timbre in terms of the sound spectrum in steady state. Multidimensional models of timbre perception that emphasize time-variant characteristics were described.

The fourth part of the chapter examined the perception of temporal relationships in patterns of sound. Evidence concerning perception of the order of two or more sound events was reviewed, as were theories of how such order perception might be accomplished. Next explored were the ways in which listeners group patterns of sound in time. Finally, the evidence for hierarchical encoding of rhythmic patterns was examined.

It is clear from the numerous examples presented in this chapter that the auditory system is capable of a very high level of information abstraction, provided that it is given appropriate input. Knowledge of the properties of this system should therefore prove of considerable value to those wishing to exploit its capabilities in future technological developments.

NOTES

1. The view taken here is that the concept of "unconscious inference" serves as a useful heuristic to guide the experimenter in the study of grouping mechanisms, and that this is also true of principles of figural goodness. However, it is not assumed that these concepts serve as "explanations" of grouping phenomena in any fundamental sense. Rather, it is expected that the behavioral phenomena described here will ultimately be explained in terms of underlying physiological mechanisms. Where possible, relevant neurophysiological findings are presented here; however, the neurophysiological bases of most of the phenomena described in this chapter are presently unknown.
2. A periodic sound can be described in terms of its sinusoidal frequency components, or partials. When the frequencies of the upper partials are integral multiples of the frequency of the lowest partial (or fundamental), the sound is termed *harmonic*. When this is not the case, the sound is termed *nonharmonic*. Sounds are termed *fused* when they appear as single sounds and are termed *unfused* when they appear as several simultaneous sounds.
3. Kubovy, M. *The ear is not phase deaf*. Paper presented at the nineteenth annual meeting of the Psychonomic Society, San Antonio, November, 9-11, 1978.
4. Kubovy, M. *The sound of silence: A new pitch-segregation phenomenon*. Paper presented at the seventeenth annual meeting of the Psychonomic Society, St. Louis, November 11-13, 1976.
5. A major scale consists of seven tones per octave. When a scale is presented in ascending order, the intervals formed by successive tones are two major seconds, a minor second, three major seconds, and a minor second. (See Figure 32.24.) When the scale begins on the note C it is called the C-major scale. When it begins on the note D it is called the D-major scale, and so on. Any scale consisting of tones related by this pattern of intervals is called a diatonic scale.
6. A contrapuntal pattern results when two or more melodic lines are presented in parallel.
7. First-order localization cues are provided by differences in amplitude and in time of arrival of the sounds at the two ears.
8. The fact that following the nondominant ear required a larger amplitude difference between the ears in the second experiment compared with the first simply reflects the fact that different subjects were employed in these two experiments, as large differences exist between subjects on this measure.
9. Reverberation times in enclosed spaces are of the order of seconds, which is consistent with the time courses observed here.
10. Erickson, R. *LOOPS, an informal timbre experiment*. Unpublished manuscript, Center for Music Experiment, University of California, San Diego, 1974.
11. Neurons in the auditory system have been found that exhibit peaks of sensitivity at octave multiples (Evans, 1974; Suga & Jen, 1976). Such neurons would mediate the hypothesized convergence of input based on the octave relation.
12. Neurons in the auditory system have been found whose firing is facilitated when certain harmonically related tones are presented together (Suga, O'Neill, & Manabe, 1979). The characteristics of these units are as hypothesized to exist at one stage along the channel mediating transposition.
13. The chromatic scale consists of tones that are spaced in semitone increments. There are 12 such tones within the octave.
14. For a description of lateral inhibitory networks see Ratliff (1965).
15. A harmonic progression may be broadly defined as a series of chords that bear certain relationships to each other. For a detailed description see Forte (1974).
16. A formant is a fixed frequency region in which the partials of a tone are prominent, regardless of the frequency of the fundamental.
17. Plomp, R., & Steeneken, H. J. M. *Pitch versus timbre*. Paper presented at the Seventh International Congress on Acoustics, Budapest, 1971.

18. The critical band is that band of frequencies within which the loudness of a band of noise of constant sound pressure is independent of bandwidth.
19. Deutsch, D., & Feroe, J. *The internal representation of rhythmic patterns*. In preparation.

REFERENCES

- Allen, D. Octave discriminability of musical and non-musical subjects. *Psychonomic Science*, 1967, 7, 421-422.
- Attneave, F., & Olson, R. K. Pitch as a medium: A new approach to psychophysical scaling. *Journal of Psychology*, 1971, 84, 147-165.
- Babbitt, M. Twelve-tone invariants as compositional determinants. *The Musical Quarterly*, 1960, 46, 246-259.
- Babbitt, M. The structure and function of musical theory. *College Music Symposium*, 1965, 5, 10-21.
- Babkoff, H. Diotic temporal interactions: Fusion and temporal order. *Perception and Psychophysics*, 1975, 18, 267-272.
- Bachem, A. Note on Neu's review of the literature on absolute pitch. *Psychological Bulletin*, 1948, 45, 161-162.
- Bachem, A. Time factors in relative and absolute pitch; Studies in psychology. *Journal of the Acoustical Society of America*, 1954, 26, 751-753.
- Baird, J. W. Memory for absolute pitch: Studies in psychology. In *Titchener commemorative volume*. Worcester, 1917.
- Bartlett, J. C., & Dowling, W. J. Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, 6, 501-515.
- Bartlett, N. R., & Bartlett, S. C. Synchronization of a motor response with an anticipated sensory event. *Psychological Review*, 1959, 66, 203-218.
- Berger, K. W. Some factors in the recognition of timbre. *Journal of the Acoustical Society of America*, 1964, 36, 1888-1891.
- Bharucha, J., & Krumhansl, C. L. The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, 1983, 13, 63-102.
- Bjork, R. A. All-or-none subprocesses in the learning of complex sequences. *Journal of Mathematical Psychology*, 1968, 5, 182-195.
- Blackwell, H. R., & Schlosberg, H. Octave generalization, pitch discrimination, and loudness thresholds in the white rat. *Journal of Experimental Psychology*, 1942, 33, 407-419.
- Bolton, T. L. Rhythm. *American Journal of Psychology*, 1894, 6, 145-238.
- Bower, G. H. A selective review of organizational factors in memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic, 1972.
- Bower, G. H., & Winzeng, D. Group structure, coding and memory for digit series. *Journal of Experimental Psychology Monographs*, 1969, 80 (2, Pt. 2), 1-17.
- Bregman, A. S. The formation of auditory streams. In J. Requin (Ed.), *Attention and performance* (Vol. 7). Hillsdale, N.J.: Erlbaum, 1978.
- Bregman, A. S. Asking the "what for" question in auditory perception. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, N.J.: Erlbaum, 1981.
- Bregman, A. S., & Campbell, J. Primary auditory stream segregation and perception of order in rapid sequence of tones. *Journal of Experimental Psychology*, 1971, 89, 244-249.
- Bregman, A. S., & Dannenbring, G. L. The effect of continuity on auditory stream segregation. *Perception and Psychophysics*, 1973, 13, 308-312.
- Bregman, A. S., & Pinker, S. Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, 1978, 32, 20-31.
- Bregman, A. S., & Rudnicki, A. I. Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance*, 1975, 104, 263-267.
- Broadbent, D. E. The role of auditory localization in attention and memory span. *Journal of the Acoustical Society of America*, 1954, 47, 191-196.
- Broadbent, D. E. *Perception and communication*. London: Pergamon, 1958.
- Broadbent, D. E., & Ladefoged, P. Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 1959, 31, 1539-1540.
- Browne, R. Review of *The structure of atonal music* by A. Forte. *Journal of Music Theory*, 1974, 18, 390-415.
- Burns, E. M. In search of the shruti. *Journal of the Acoustical Society of America*, 1974a, 56 (Suppl.), S26.
- Burns, E. M. Octave adjustment by non-western musicians. *Journal of the Acoustical Society of America*, 1974b, 56 (Suppl.), S25-S26.
- Burns, E. M. *The perception of musical intervals (frequency ratios)*. Unpublished doctoral thesis, University of Minnesota, 1977.
- Burns, E. M., & Ward, W. D. Categorical perception of musical intervals. *Journal of the Acoustical Society of America*, 1974, 55, 456.
- Burns, E. M., & Ward, W. D. Categorical perception—Phenomenon or epiphenomenon: Evidence from experiments on the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, 1978, 63, 456-468.
- Burns, E. M., & Ward, W. D. Intervals, scales, and tuning. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic, 1982.
- Butler, D. A further study of melodic channeling. *Perception and Psychophysics*, 1979, 25, 264-268.
- Cabot, R. C., Mino, M. G., Dorans, D. A., Tackel, I. S., & Breed, H. E. Detection of phase shifts in harmonically related tones. *Journal of the Audio Engineering Society*, 1976, 24, 568-571.
- Cherry, E. C. Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 1953, 25, 975-979.
- Cherry, E. C., & Taylor, W. K. Some further experiments upon the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 1954, 26, 554-559.
- Chrisman, R. Identification and correlation of pitch-sets. *Journal of Music Theory*, 1971, 15, 58-83.
- Cooper, G. W., & Meyer, L. B. *The rhythmic structure of music*. Chicago: University of Chicago Press, 1960.
- Corteen, R. S., & Wood, B. Autonomic responses to shock-associated words in an unattended channel. *Journal of Experimental Psychology*, 1972, 94, 308-313.
- Craig, J. D. The effect of musical training and cerebral asymmetries in perception of an auditory illusion. *Cortex*, 1979, 15, 671-677.
- Dannenbring, G. L. Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian Journal of Psychology*, 1976, 30, 99-114.
- Dannenbring, G. L., & Bregman, A. S. Stream segregation and the illusion of overlap. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 544-555.
- Dannenbring, G. L., & Bregman, A. S. Streaming versus fusion of sinusoidal components of complex tones. *Perception and Psychophysics*, 1978, 24, 369-376.
- de Boer, E. On the "residue" and auditory pitch perception. In W. D. Keidel & W. D. Neff (Eds.), *Handbook of sensory physiology* (Vol. V/3). Vienna: Springer-Verlag, 1976.
- Deutsch, D. Music recognition. *Psychological Review*, 1969, 76, 300-307.
- Deutsch, D. Tones and numbers: Specificity of interference in short-term memory. *Science*, 1970, 68, 1604-1605.
- Deutsch, D. Effect of repetition of standard and comparison tones on recognition memory for pitch. *Journal of Experimental Psychology*, 1972a, 93, 156-162.
- Deutsch, D. Mapping of interactions in the pitch memory store. *Science*, 1972b, 175, 1020-1022.
- Deutsch, D. Octave generalization and tune recognition. *Perception and Psychophysics*, 1972c, 11, 411-412.
- Deutsch, D. Interference in memory between tones adjacent in the musical scale. *Journal of Experimental Psychology*, 1973a, 100, 228-231.

- Deutsch, D. Octave generalization of specific interference effects in memory for tonal pitch. *Perception and Psychophysics*, 1973b, 13, 271-275.
- Deutsch, D. An auditory illusion. *Nature* (London), 1974, 252, 307-309.
- Deutsch, D. Auditory memory. *Canadian Journal of Psychology*, 1975a, 29, 87-105.
- Deutsch, D. Facilitation by repetition in recognition memory for tonal pitch. *Memory and Cognition*, 1975b, 3, 263-266.
- Deutsch, D. Musical illusions. *Scientific American*, 1975c, 233, 92-104.
- Deutsch, D. The organization of short-term memory for a single acoustic attribute. In D. Deutsch & J. A. Deutsch (Eds.), *Short term memory*. New York: Academic, 1975d.
- Deutsch, D. Two-channel listening to musical scales. *Journal of the Acoustical Society of America*, 1975e, 57, 1156-1160.
- Deutsch, D. Lateralization by frequency in dichotic tonal sequences as a function of interaural amplitude and time differences. *Journal of the Acoustical Society of America*, 1976, 60 (Suppl.), S50.
- Deutsch, D. Memory and attention in music. In M. Critchley & R. A. Henson (Eds.), *Music and the brain*. London: Heinemann, 1977.
- Deutsch, D. Delayed pitch comparisons and the principle of proximity. *Perception and Psychophysics*, 1978a, 23, 227-230.
- Deutsch, D. Interactive effects in memory for harmonic intervals. *Perception and Psychophysics*, 1978b, 24, 7-10.
- Deutsch, D. Lateralization by frequency for repeating sequences of dichotic 400-Hz and 800-Hz tones. *Journal of the Acoustical Society of America*, 1978c, 63, 184-186.
- Deutsch, D. Octave generalization and melody identification. *Perception and Psychophysics*, 1978e, 23, 91-92.
- Deutsch, D. The psychology of music. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 10). New York: Academic, 1978d.
- Deutsch, D. Binaural integration of melodic patterns. *Perception and Psychophysics*, 1979a, 25, 399-405.
- Deutsch, D. Octave generalization and the consolidation of melodic information. *Canadian Journal of Psychology*, 1979b, 33, 201-204.
- Deutsch, D. Ear dominance and sequential interactions. *Journal of the Acoustical Society of America*, 1980a, 67, 220-228.
- Deutsch, D. The processing of structured and unstructured tonal sequences. *Perception and Psychophysics*, 1980b, 28, 381-389.
- Deutsch, D. The octave illusion and auditory perceptual integration. In J. V. Tobias & E. D. Schubert (Eds.), *Hearing research and theory* (Vol. 1). New York: Academic, 1981.
- Deutsch, D. The influence of melodic context on pitch recognition judgment. *Perception and Psychophysics*, 1982a, 31, 407-410.
- Deutsch, D. The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic, 1982b.
- Deutsch, D. The generation of two isochronous sequences in parallel. *Perception and Psychophysics*, 1983a, 34, 331-337.
- Deutsch, D. The octave illusion in relation to handedness and familial handedness background. *Neuropsychologia*, 1983b, 21, 289-293.
- Deutsch, D., & Boulanger, R. C. Octave equivalence and the immediate recall of pitch sequences. *Music Perception*, 1984, 2, 40-51.
- Deutsch, D., & Feroe, J. Disinhibition in pitch memory. *Perception and Psychophysics*, 1975, 17, 320-324.
- Deutsch, D., & Feroe, J. The internal representation of pitch sequences in tonal music. *Psychological Review*, 1981, 88, 503-522.
- Deutsch, D., & Roll, P. L. Error patterns in delayed pitch comparison as a function of relational context. *Journal of Experimental Psychology*, 1974, 103, 1027-1034.
- Deutsch, D., & Roll, P. L. Separate "what" and "where" decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 23-29.
- Deutsch, J. A., & Deutsch, D. Attention: Some theoretical considerations. *Psychological Review*, 1963, 70, 80-90.
- Dewar, K. M. *Context effects in recognition memory for tones*. Unpublished doctoral dissertation, Queens University, Kingston, Ontario, 1974.
- Dewar, K. M., Cuddy, C. L., & Mewhort, D. J. K. Recognition memory for single tones with and without context. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, 3, 60-67.
- Divenyi, P. L., & Hirsh, I. J. Discrimination of the silent gap in two-tone sequences of different frequencies. *Journal of the Acoustical Society of America*, 1972, 52, 166-167.
- Divenyi, P. L., & Hirsh, I. J. Identification of temporal order in three-tone sequences. *Journal of the Acoustical Society of America*, 1974, 56, 144-151.
- Divenyi, P. L., & Hirsh, I. J. The effect of blanking on the identification of temporal order in the three-tone sequences. *Perception and Psychophysics*, 1975, 17, 246-252.
- Divenyi, P. L., & Hirsh, I. J. Some figural properties of auditory patterns. *Journal of the Acoustical Society of America*, 1978, 64, 1369-1385.
- Dorman, M. F., Cutting, J. E., & Raphael, L. J. Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, 104, 121-129.
- Dowling, W. J. Recognition of melodic transformations: Inversion, retrograde, and retrograde-inversion. *Perception and Psychophysics*, 1972, 12, 417-421.
- Dowling, W. J. The perception of interleaved melodies. *Cognitive Psychology*, 1973a, 5, 322-337.
- Dowling, W. J. The 1215-cent octave: Convergence of western and non-western data on pitch scaling. *Journal of the Acoustical Society of America*, 1973b, 53, 373.
- Dowling, W. J. Rhythmic groups and subjective chunks in memory for melodies. *Perception and Psychophysics*, 1973c, 4, 37-40.
- Dowling, W. J. Scaling and contour: Two components of a theory of memory for melodies. *Psychological Review*, 1978, 85, 342-354.
- Dowling, W. J., & Bartlett, J. C. The importance of interval information in long-term memory for melodies. *Psychomusicology*, 1981, 1, 30-49.
- Dowling, W. J., & Fujitani, D. S. Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, 1971, 49, 1525-1531.
- Dowling, W. J., & Hollombe, A. W. The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception and Psychophysics*, 1977, 21, 60-64.
- Drobisch, M. W. Über die mathematische bestimmung der musikalischen intervale, 1846. (Cited by C. A. Ruckmick, A new classification of tonal qualities. *Psychological Review*, 1929, 36, 172-180.)
- Erickson, R. *Sound structure in music*. Berkeley: University of California Press, 1975.
- Estes, W. K. An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington, D. C.: Winston, 1972.
- Evans, E. F. Neural processes for the detection of acoustic patterns and for sound localization. In F. O. Schmitt & F. T. Worden (Eds.), *The neurosciences, third study program*. Cambridge, Mass.: M.I.T. Press, 1974.
- Fitzgibbon, P. J., Pollatsek, A., & Thomas, I. B. Detection of temporal gaps within and between perceptual tonal groups. *Perception and Psychophysics*, 1974, 16, 522-528.
- Forte, A. *The structure of atonal music*. New Haven, Ct.: Yale University Press, 1973.
- Forte, A. *Tonal harmony in concept and practice* (2nd ed.). New York: Holt, Rinehart, and Winston, 1974.
- Fraisse, P. *Les structures rythmiques*. Louvain: Editions Universitaires, 1956.
- Fraisse, P. Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic, 1982.
- Fraisse, P., Pichot, P., & Clairouin, G. Les aptitudes rythmiques. Etude comparee des oligophrenes et des enfants normaux. *Journal de Psychologie Normal et Pathologique*, 1949, 42, 309-330.
- Frances, R. *La perception de la musique*. Paris: Vrin, 1958.
- Garner, W. R. *The processing of information and structure*. Hillsdale, N.J.: Erlbaum, 1974.
- Garner, W. R., & Gottwald, R. L. The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology*, 1968, 20, 97-109.

- Geffen, G., & Reynolds, N. Pure-tone perception and ear advantages in dichotic listening. *Perception and Psychophysics*, 1982, 31, 68-75.
- Gengel, R. W., & Hirsh, I. J. Temporal order: The effect of single versus repeated presentations, practice, and verbal feedback. *Perception and Psychophysics*, 1970, 7, 209-211.
- Goodglass, H., & Quadfasel, F. A. Language laterality in left-handed aphasics. *Brain*, 1954, 77, 521-543.
- Gray, J. A., & Wedderburn, A. A. I. Grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, 1960, 12, 180-184.
- Greeno, J. G., & Simon, H. A. Processes for sequence production. *Psychological Review*, 1974, 81, 187-196.
- Gregory, R. L. *The intelligent eye*. New York: McGraw-Hill, 1970.
- Grey, J. M. *An exploration of musical timbre*. Unpublished doctoral dissertation, Stanford University, 1975.
- Grey, J. M. Timbre discrimination in musical patterns. *Journal of the Acoustical Society of America*, 1978, 64, 467-472.
- Grey, J. M., & Gordon, J. W. Perceptual effects of spectral modifications in musical timbres. *Journal of the Acoustical Society of America*, 1978, 63, 1493-1500.
- Grey, J. M., & Moorer, J. A. Perceptual evaluation of synthesized musical instrument tones. *Journal of the Acoustical Society of America*, 1977, 62, 454-462.
- Haas, H. Über den einfluss eines Einfachechos auf die Hørsamkeit von Sprache. *Acustica*, 1951, 1, 49-52.
- Hall, D. E. Quantitative evaluation of musical scale tunings. *American Journal of Physics*, 1974, 42, 543-552.
- Hall, D. E. *Musical acoustics: An introduction*. Belmont, Cal.: Wadsworth, 1980.
- Handel, S. Temporal segmentation of repeating auditory patterns. *Journal of Experimental Psychology*, 1973, 101, 46-54.
- Handel, S. Using polyrhythms to study rhythm. *Music Perception*, 1984, 1, 465-484.
- Handel, S., & Oshinsky, J. S. The meter of syncopated auditory polyrhythms. *Perception and Psychophysics*, 1981, 30, 1-9.
- Handel, S., & Yoder, D. The effects of intensity and rhythm intervals on the perception of auditory and visual temporal patterns. *Quarterly Journal of Experimental Psychology*, 1975, 27, 111-122.
- Hanson, A. R., & Riseman, E. M. (Eds.). *Computer vision systems*. New York: Academic, 1978.
- Harris, J. D. The decline of pitch discrimination with time. *Journal of Experimental Psychology*, 1952, 43, 96-99.
- Hécaen, H., & de Ajureaguerra, J. *Left handedness*. New York: Grune and Stratton, 1964.
- Hécaen, H., & Piercy, M. Paroxysmal dysphasia and the problem of cerebral dominance. *Journal of Neurology and Neurological Psychiatry*, 1956, 19, 194-201.
- Heise, G. A., & Miller, G. A. An experimental study of auditory patterns. *American Journal of Psychology*, 1951, 64, 68-77.
- Helmholtz, V. H. *On the sensations of tones as a physiological basis for the theory of music* (2nd English ed.). New York: Dover, 1954. (Originally published, 1859.)
- Hirsh, I. J. Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 1959, 31, 759-767.
- Hirsh, I. J. Temporal order and auditory perception. In H. R. Moskowitz, B. Scharf, & J. C. Stevens (Eds.), *Sensation and measurement*. Dordrecht, Netherlands: Reidel, 1974.
- Hirsh, I. J. Order of events in three sensory modalities. In S. K. Hirsh, D. H. Eldridge, I. J. Hirsh, & S. R. Silverman (Eds.), *Essays honoring Hallowell Davis*. St. Louis, Mo.: Washington University Press, 1976.
- Hirsh, I. J., & Fraisse, P. Simultaneite et succession de stimuli heterogenes. *Annee Psychologique*, 1964, 64, 1-19.
- Hirsh, I. J., & Sherrick, C. E. Perceived order in different sense modalities. *Journal of Experimental Psychology*, 1961, 62, 423-432.
- Hochberg, J. Organization and the Gestalt tradition. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 1). New York: Academic, 1974.
- Howe, H. S. Some combinatorial properties of pitch structures. *Journal of Music Theory*, 1965, 4, 45-61.
- Huggins, A. W. F. Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America*, 1964, 36, 1055-1064.
- Hulse, S. H., Humpal, J., & Cynx, J. Discrimination and generalization of rhythmic and arrhythmic sound patterns by European starlings (*Sturnus vulgaris*). *Music Perception*, 1984, 1, 442-464.
- Humphreys, L. F. Generalization as a function of method of reinforcement. *Journal of Experimental Psychology*, 1939, 25, 361-372.
- Idson, W. L., & Massaro, D. W. A bidimensional model of pitch in the recognition of melodies. *Perception and Psychophysics*, 1978, 24, 551-565.
- Judd, T. Comments on Deutsch's musical scale illusion. *Perception and Psychophysics*, 1979, 26, 85-92.
- Julesz, B. *Foundations of cyclopean perception*. Chicago: University of Chicago Press, 1971.
- Julesz, B., & Hirsh, I. J. Visual and auditory perception—An essay of comparison. In E. E. Davis & P. B. Denes (Eds.), *Human communication: A unified view*. New York: McGraw-Hill, 1972.
- Kahneman, D. *Attention and effort*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.
- Kallman, H. J. Octave equivalence as measured by similarity ratings. *Perception and Psychophysics*, 1982, 32, 37-49.
- Kallman, H. J., & Massaro, D. W. Tone chroma is functional in melody recognition. *Perception and Psychophysics*, 1979, 26, 32-36.
- Keele, S. W., & Neill, W. T. Mechanisms of attention. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 9). New York: Academic, 1979.
- Keiler, A. On some properties of Schenker's pitch derivations. *Music Perception*, 1983, 1, 200-228.
- Klinke, R., Boerger, G., & Gruber, J. Alteration of afferent, tone-evoked activity of neurons of the cochlear nucleus following acoustic stimulation of the contralateral ear. *Journal of the Acoustical Society of America*, 1969, 45, 788-789.
- Klinke, R., Boerger, G., & Gruber, J. The influence of the frequency relation in dichotic stimulation upon the cochlear nucleus activity. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing*. Sijthoff, Netherlands: Leiden, 1970.
- Koester, T. The time error in pitch and loudness discrimination as a function of time interval and stimulus level. *Archives of Psychology*, 1945, 297, entire issue.
- Koffka, K. *Principles of Gestalt psychology*. New York: Harcourt, 1935.
- Kolinski, M. The evaluation of tempo. *Ethnomusicology*, 1959, 3, 45-57.
- Kotovsky, K., & Simon, H. A. Empirical tests of a theory of human acquisition of concepts of sequential events. *Cognitive Psychology*, 1973, 4, 399-424.
- Krumhansl, C. L. The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, 1979, 11, 346-374.
- Krumhansl, C. L. Perceptual structures for tonal music. *Music Perception*, 1983, 1, 28-62.
- Krumhansl, C. L., Bharucha, J. J., & Kessler, E. J. Perceived harmonic structure of chords in three related musical keys. *Journal of Experimental Psychology: Human Perception and Performance*, 1982, 8, 24-36.
- Krumhansl, C. L., & Kessler, E. J. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 1982, 89, 334-368.
- Krumhansl, C. L., & Shepard, R. N. Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, 5, 579-594.
- Kubovy, M. Concurrent pitch-segregation and the theory of indispensable attributes. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, N.J.: Erlbaum, 1981.
- Kubovy, M., Cutting, J. E., & McGuire, R. M. Hearing with the third ear: Dichotic perception of a melody without monaural familiarity cues. *Science*, 1974, 186, 272-274.
- Kubovy, M., & Howard, F. P. Persistence of a pitch-segregating echoic memory. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 531-537.

- Kubovy, M., & Jordan, R. Tone-segregation by phase: On the phase sensitivity of the single ear. *Journal of the Acoustical Society of America*, 1979, 66, 100-106.
- Leeuwenberg, E. L. A perceptual coding language for visual and auditory patterns. *American Journal of Psychology*, 1971, 84, 307-349.
- Lerdahl, F., & Jackendoff, R. *A generative theory of tonal music*. Cambridge, Mass.: M.I.T. Press, 1983a.
- Lerdahl, F., & Jackendoff, R. An overview of hierarchical structure in music. *Music Perception*, 1983b, 1, 229-252.
- Lewin, D. The intervallic content of a collection of notes. *Journal of Music Theory*, 1960, 4, 98-101.
- Lewin, D. A theory of segmental association in twelve-tone music. *Perspectives of New Music*, 1962, 1, 89-116.
- Lewis, J. L. Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, 1970, 85, 225-228.
- Locke, S., & Kellar, L. Categorical perception in a non-linguistic mode. *Cortex*, 1973, 9, 355-368.
- Longuet-Higgins, H. C. Letter to a musical friend. *Music Review*, 1962a, 23, 244-248.
- Longuet-Higgins, H. C. Second letter to a musical friend. *Music Review*, 1962b, 23, 271-280.
- Longuet-Higgins, H. C. The perception of music. *Interdisciplinary Science Reviews*, 1978, 3, 148-156.
- Mach, E. *The analysis of sensations and the relation of the physical to the psychological* (C. M. Williams, trans.; W. Waterlow, review and supplement). New York: Dover, 1959. (Originally published in German, 1906.)
- Mathes, R. C., & Miller, R. L. Phase effects in monaural perception. *Journal of the Acoustical Society of America*, 1947, 19, 780-797.
- Mathews, M. V., & Pierce, J. R. Harmony and nonharmonic partials. *Journal of the Acoustical Society of America*, 1980, 68, 1252-1257.
- McAdams, S. Spectral fusion and the creation of auditory images. In M. Clynes (Ed.), *Music, mind and brain*. New York: Plenum, 1982.
- McClurkin, R. H., & Hall, J. Pitch and timbre in a two-tone dichotic auditory illusion. *Journal of the Acoustical Society of America*, 1981, 69, 592-594.
- McLean, R. S., & Gregg, L. W. Effects of induced chunking on temporal aspects of serial retention. *Journal of Experimental Psychology*, 1967, 74, 455-459.
- McNally, K. A., & Handel, S. Effects of element composition on streaming and the ordering of repeating sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, 3, 451-460.
- Meyer, L. B. *Emotion and meaning in music*. Chicago: University of Chicago Press, 1956.
- Meyer, L. B. *Music, the arts and ideas*. Chicago: University of Chicago Press, 1960.
- Meyer, L. B. *Explaining music: Essays and explorations*. Berkeley: University of California Press, 1973.
- Meyer, M. On the attributes of the sensations. *Psychological Review*, 1904, 11, 83-103.
- Meyer, M. Review of G. Revesz, "Zur Grundlegung der Tonpsychologie." *Psychological Bulletin*, 1914, 11, 349-352.
- Michon, J. A. Magnitude scaling of short durations with closely spaced stimuli. *Psychonomic Science*, 1967, 9, 359-360.
- Miller, G. A., & Heise, G. A. The trill threshold. *Journal of the Acoustical Society of America*, 1950, 22, 637-638.
- Miller, G. A., & Licklider, J. C. R. The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 1950, 22, 167-173.
- Miller, J. R., & Carterette, E. C. Perceptual space for musical structures. *Journal of the Acoustical Society of America*, 1975, 58, 711-720.
- Miller, J. R., Wier, C., Pastore, R., Kelly, W., & Dooling, R. Discrimination and labeling of noise-buzz sequences with varying noise-level times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, 60, 410-417.
- Milner, B., Branch, C., & Rasmussen, T. Evidence for bilateral speech representation in some nonrighthanders. *Transactions of the American Neurological Association*, 1966, 91, 306-308.
- Montpellier, G. de. *Les alterations morphologiques des mouvements rapides*. Louvain, France: Institut Supérieur de Philosophie, 1935.
- Mueller, G. E., & Schumann, F. Experimentelle Beiträge zur Untersuchung des Gedächtnisses. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 1894, 6, 81-190; 257-339.
- Nabelek, I. V., Nabelek, A. K., & Hirsh, I. J. Pitch of sound bursts with continuous or discontinuous change of frequency. *Journal of the Acoustical Society of America*, 1973, 53, 1305-1312.
- Narmour, E. *Beyond Schenkerism*. Chicago: University of Chicago Press, 1977.
- Narmour, E. Some major theoretical problems concerning the concept of hierarchy in the analysis of tonal music. *Music Perception*, 1983, 1, 129-199.
- Neisser, U. *Cognitive psychology*. New York: Appleton, 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 and Psychophysics*, 1974, 16, 471-477.
- Ortmann, O. On the melodic relativity of tones. *Psychological Monographs*, 1926, 35 (Whole No. 162).
- Patterson, J. H., & Green, D. M. Discrimination of transient signals having identical energy spectra. *Journal of the Acoustical Society of America*, 1970, 48, 898-905.
- Perkins, D.N. Coding position in a sequence by rhythmic grouping. *Memory and Cognition*, 1974, 2, 219-223.
- Perle, G. *Serial composition and atonality* (3rd ed.). Berkeley: University of California Press, 1972.
- Perle, G. *Twelve-tone tonality*. Berkeley: University of California Press, 1977.
- Piston, W. *Harmony* (2nd ed.). London: Norton, 1948.
- Plomp, R. Timbre as a multidimensional attribute of complex tones. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing*. Sijthoff, Netherlands: Leiden, 1970.
- Plomp, R. Perception of sound signals at low signal-to-noise ratios. In D. J. Getty & J. H. Howard, Jr. (Eds.), *Auditory and visual pattern recognition*. Hillsdale, N.J.: Erlbaum, 1981.
- Plomp, R., & Mimpen, A. M. The ear as frequency analyzer II. *Journal of the Acoustical Society of America*, 1968, 43, 764-767.
- Plomp, R., & Steeneken, H. J. M. Effect of phase on the timbre of complex tones. *Journal of the Acoustical Society of America*, 1969, 46, 409-421.
- Plomp, R., Wagenaar, W. A., & Mimpen, A. M. Musical interval recognition with simultaneous tones. *Acustica*, 1973, 29, 101-109.
- Povel, D. J. Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 1981, 7, 3-18.
- Povel, D. J., & Okkerman, H. Accents in equitone sequences. *Perception and Psychophysics*, 1981, 30, 565-572.
- Preusser, D., Garner, W. R., & Gottwald, R. L. Perceptual organization of two-element temporal patterns as a function of their component one-element patterns. *American Journal of Psychology*, 1970, 83, 151-170.
- Rakowski, A. Tuning of isolated musical intervals. *Journal of the Acoustical Society of America*, 1976, 59 (Suppl.), 550.
- Rameau, J. P. *Treatise on harmony* (P. Gosset, trans.). New York: Dover, 1971. (Originally published, 1722.)
- Rasch, R. A. The perception of simultaneous notes such as in polyphonic music. *Acustica*, 1978, 40, 1-72.
- Ratcliff, F. *Mach bands: Quantitative studies of neural networks in the retina*. San Francisco: Holden Day, 1965.
- Restle, F. Theory of serial pattern learning: Structural trees. *Psychological Review*, 1970, 77, 481-495.
- Restle, F., & Brown, E. Organization of serial pattern learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 4). New York: Academic, 1970.
- Revesz, G. *Zur Grundlegung der Tonpsychologie*. Leipzig: Feit, 1913.
- Risset, J. C. *Computer study of trumpet tones*. Murray Hill, N.J.: Bell Laboratories, 1966.
- Risset, J. C. Musical acoustics. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 4). New York: Academic, 1978.
- Risset, J. C., & Mathews, M. V. Analysis of musical instruments tones.

- Physics Today*, 1969, 22(2), 23-30.
- Risset, J. C., & Wessel, D. L. Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *Psychology of music*. New York: Academic, 1982.
- Royer, F. L., & Garner, W. R. Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception and Psychophysics*, 1966, 1, 41-47.
- Royer, F. L., & Garner, W. R. Perceptual organization of nine-element auditory temporal patterns. *Perception and Psychophysics*, 1970, 7, 115-120.
- Ruckmick, C. A. A new classification of tonal qualities. *Psychological Review*, 1929, 36, 172-180.
- Saldanha, E. L., & Corso, J. F. Timbre cues for the recognition of musical instruments. *Journal of the Acoustical Society of America*, 1964, 36, 2021-2026.
- Salzer, F. *Structural hearing*. New York: Dover, 1962.
- Schackford, C. Some aspects of perception. I. *Journal of Music Theory*, 1961, 5, 162-202.
- Schackford, C. Some aspects of perception. II. *Journal of Music Theory*, 1962, 6, 66-90.
- Schaeffer, P. *Traite des objets musicaux*. Paris: Ed. du Seuil, 1966. (With three records of sound examples.)
- Schenker, H. *Neus musikalische theorien un phantasien: Der freie satz*. Vienna, Austria: Universal Edition, 1956.
- Schenker, H. *Harmony*. (O. Jonas, Ed.; E. M. Borgese, trans.) Cambridge, Mass.: M.I.T. Press, 1973.
- Schoenberg, A. *Style and idea*. London: Williams & Norgate, 1951.
- Schouten, J. F. On the perception of sound and speech: Subjective time analysis. *Fourth International Congress on Acoustics, Copenhagen Congress Report II*, 1962, pp. 201-203.
- Schroeder, M. R. Models of hearing. *Proceedings of the Institute of Electrical and Electronics Engineers*, 1975, 63, 1332-1350.
- Schubert, E. D., & Parker, C. D. Addition to Cherry's findings on switching speech between two ears. *Journal of the Acoustical Society of America*, 1956, 27, 792-794.
- Shepard, R. N. Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America*, 1964, 36, 2345-2353.
- Shepard, R. N. Structural representations of musical pitch. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic, 1982.
- Shiffrin, R. M., & Schneider, W. Toward a unitary model for selective attention, memory scanning and visual search. In S. Dornic (Ed.), *Attention and performance* (Vol. 6). Hillsdale, N.J.: Erlbaum, 1977.
- Siegel, J. A., & Siegel, W. Absolute identification of notes and intervals by musicians. *Perception and Psychophysics*, 1977a, 21, 143-152.
- Siegel, J. A., & Siegel, W. Categorical perception of tonal intervals: Musicians can't tell sharp from flat. *Perception and Psychophysics*, 1977b, 21, 399-407.
- Simon, H. A. Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 1972, 79, 369-382.
- Simon, H. A., & Kotovsky, K. Human acquisition of concepts for sequential patterns. *Psychological Review*, 1963, 70, 534-546.
- Simon, H. A., & Sumner, R. K. Pattern in music. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley, 1968.
- Slawson, A. W. Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *Journal of the Acoustical Society of America*, 1968, 43, 87-101.
- Small, A. M. An objective analysis of violin performance. *University of Iowa Studies in the Psychology of Music*, 1936, 4, 172-231.
- Steiger, H., & Bregman, A. S. Capturing frequency components of glided tones: Frequency separation, orientation, and alignment. *Perception and Psychophysics*, 1981, 30, 425-431.
- Sternberg, S., & Knoll, R. L. The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance* (Vol. 4). New York: Academic, 1973.
- Sternberg, S., Knoll, R. L., & Zukofsky, P. Timing by skilled musicians. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic, 1982.
- Studdert-Kennedy, M., Liberman, A. M., Harris, K., & Cooper, F. S. The motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, 77, 234-249.
- Stumpf, C., & Meyer, M. Maassbestimmungen uber die Reinheit consonanter Intervalle. *Beitrage zur Akustik und Musik*, 1898, 2, 84-167.
- Sturges, P. T., & Martin, J. G. Rhythmic structures in auditory temporal pattern perception and immediate memory. *Journal of Experimental Psychology*, 1974, 102, 377-383.
- Subirana, A. Handedness and cerebral dominance. In P. J. Vinken & G. W. Bruyn (Eds.), *Handbook of clinical neurology*, 1969, 4, 248-272.
- Suga, N., & Jen, P. H. S. Disproportionate tonotopic representation for processing CF-FM sonar signals in the mustache bat auditory cortex. *Science*, 1976, 194, 542-544.
- Suga, N., O'Neill, W. E., & Manabe, T. Harmonic-sensitive neurons in the auditory cortex of the mustache bat. *Science*, 1979, 203, 270-274.
- Summers, J. J. The role of timing in motor program representation. *Journal of Motor Behavior*, 1975, 7, 229-241.
- Sundberg, J. E. F., & Lindquist, J. Musical octaves and pitch. *Journal of the Acoustical Society of America*, 1973, 54, 922-929.
- Sutherland, N. S. Object recognition. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 3). New York: Academic, 1973.
- Terhardt, E. Pitch shifts of harmonics, an explanation of the octave enlargement phenomenon. *Proceedings of the Seventh International Congress on Acoustics*, Budapest, 1971.
- Thomas, I. B., Cetti, R. P., & Chase, P. W. Effects of silent intervals on the perception of temporal order for vowels. *Journal of the Acoustical Society of America*, 1971, 49, 584.
- Thomas, I. B., Hill, P. B., Carroll, F. S., & Garcia, B. Temporal order in the presentation of vowels. *Journal of the Acoustical Society of America*, 1970, 48, 1010-1013.
- Thurlow, W. R. An auditory figure-ground effect. *American Journal of Psychology*, 1957, 70, 653-654.
- Thurlow, W. R., & Erchul, W. P. Judged similarity in pitch of octave multiples. *Perception and Psychophysics*, 1977, 22, 177-182.
- Tobias, J. V., & Schubert, E. D. Effective onset duration of auditory stimuli. *Journal of the Acoustical Society of America*, 1959, 31, 1595-1605.
- Treisman, A. M. Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 1960, 12, 242-248.
- Treisman, A. M. Selective attention in man. *British Medical Bulletin*, 1964, 20, 12-16.
- Treisman, A. M. Shifting attention between the ears. *Quarterly Journal of Experimental Psychology*, 1971, 23, 157-167.
- Treisman, M. Temporal discrimination and the indifference interval: Implications for a model of the "internal clock." *Psychological Monographs*, 1963, 77 (13, Whole No. 576).
- Van Noorden, L. P. A. S. *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Technische Hogeschool, Eindhoven, the Netherlands, 1975.
- Vicario, G. L'effetto tunnel acustico. *Revista di Psicologia*, 1960, 54, 41-52.
- Vitz, P. C., & Todd, T. C. A model of learning for simple repeating binary patterns. *Journal of Experimental Psychology*, 1967, 75, 108-117.
- Vitz, P. C., & Todd, T. C. A coded element model of the perceptual processing of sequential stimuli. *Psychological Review*, 1969, 76, 433-449.
- Von Ehrenfels, C. *Über Gestaltqualitäten Vierteljahrsschrift für Wissenschaftliche Philosophie*, 1890, 14, 249-292.
- Vos, P. G. *Pattern perception in metrical tone sequences*. Unpublished thesis, University of Nijmegen, Nijmegen, the Netherlands, 1973.
- Wagner, C. The influence of the tempo of playing on the rhythmic structure studied at pianists playing scales. In E. Jokl & H. Hebelinck (Eds.), *Medicine and sports* (Vol. 6). Basel, Switzerland: Karger, 1971.
- Wallach, H., Newman, E. B., & Rosenzweig, M. R. The precedence effect

- in sound localization. *American Journal of Psychology*, 1949, 62, 315-336.
- Ward, W. D. Subjective musical pitch. *Journal of the Acoustical Society of America*, 1954, 26, 369-380.
- Warren, R. M. Perceptual restoration of missing speech sounds. *Science*, 1970, 167, 392-393.
- Warren, R. M. Auditory pattern discrimination by untrained listeners. *Perception and Psychophysics*, 1974a, 15, 495-500.
- Warren, R. M. Auditory temporal discrimination by trained listeners. *Cognitive Psychology*, 1974b, 6, 237-256.
- Warren, R. M. *Auditory perception: A new synthesis*. New York: Pergamon, 1982.
- Warren, R. M., & Byrnes, D. L. Temporal discrimination of recycled tonal sequences: Pattern matching and naming of order by untrained listeners. *Journal of the Acoustical Society of America*, 1975, 18, 273-280.
- Warren, R. M., Obusek, C. J., & Ackroff, J. M. Auditory induction: Perceptual synthesis of absent sounds. *Science*, 1972, 176, 1149-1151.
- Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. Auditory sequence: Confusions of patterns other than speech or music. *Science*, 1969, 164, 586-587.
- Warren, R. M., & Warren, R. P. Auditory illusions and confusions. *Scientific American*, 1970, 223, 30-36.
- Werner, H. Uber Mikromelodik und Mikroharmonik. *Zeitschrift fur Psychologie*, 1925, 98, 74-89.
- Wertheimer, M. Untersuchung zur Lehre von der Gestalt II. *Psychologische Forschung*, 1923, 4, 301-350.
- Wessel, D. L. Psychoacoustics and music. *Bulletin of the Computer Arts Society*, 1973, 1, 30-31.
- Wessel, D. L. Timbre space as a musical control structure. *Computer Music Journal*, 1979, 3, 45-52.
- Westergaard, P. *An introduction to tonal theory*. New York: Norton, 1975.
- White, B. Recognition of distorted melodies. *American Journal of Psychology*, 1960, 73, 100-107.
- Wickelgren, W. A. Consolidation and retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, 1966, 72, 250-259.
- Wickelgren, W. A. Rehearsal grouping and the hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology*, 1967, 19, 97-102.
- Wickelgren, W. A. Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology*, 1969a, 6, 13-61.
- Wickelgren, W. A. Context-sensitive coding, associative memory and serial order in (speech) behavior. *Psychological Review*, 1969b, 76, 1-15.
- Wickelgren, W. A. Phonetic coding and serial order. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 7). New York: Academic, 1976.
- Williams, K. N., & Perrott, D. R. Temporal resolution of tonal pulses. *Journal of the Acoustical Society of America*, 1972, 51, 644-647.
- Wing, A. M., & Kristofferson, A. B. Response delays and the timing of discrete motor sequences. *Perception and Psychophysics*, 1973a, 14, 5-12.
- Wing, A. M., & Kristofferson, A. B. The timing of interresponse intervals. *Perception and Psychophysics*, 1973b, 13, 455-460.
- Woodrow, H. A quantitative study of rhythm. *Archive of Psychology*, 1909, 18, 1.
- Woodrow, H. Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
- Yeston, M. *The stratification of musical rhythm*. New Haven: Yale University Press, 1976.
- Zangwill, O. L. *Cerebral dominance and its relation to psychological function*. Edinburgh, Scotland: Oliver and Boyd, 1960.
- Zatorre, R. J., & Halpern, A. R. Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception and Psychophysics*, 1979, 26, 384-395.
- Zwicker, E. Subdivision of the audible frequency range into critical bands. *Journal of the Acoustical Society of America*, 1961, 33, 248.
- Zwicker, E., & Scharf, B. A model of loudness summation. *Psychological Review*, 1965, 72, 3-26.

