In S. Yantis & H. Pashler (eds.). Stevens' Handbook of Experimental Psychology, vol I: Sensation and Perception. Wiley, New York, pp. 397–452 (2002)

CHAPTER 10

Auditory Perception and Cognition

STEPHEN MCADAMS AND CAROLYN DRAKE

The sound environment in which we live is extraordinarily rich. As we scurry about in our little animal lives, we perceive sound sources and the sequences of events they emit and must adapt our behavior accordingly. How do we extract and make use of the information available in this highly structured acoustic array?

Sounds arise in the environment from the mechanical excitation of bodies that are set into vibration. These bodies radiate some of their vibratory energy into the surrounding air (or water) through which this energy propagates, getting bounced off some objects and partially absorbed by different materials. The nature of the acoustic wave arising from a source depends on the mechanical properties both of that source and of the interaction with other objects that set it into vibration. Many of these excitatory actions are extended through time, and this allows a listener to pick up important information concerning both the source and the action through an analysis of the sequences of events produced. Further, most environments contain several vibrating structures, and the acoustic waves impending on the eardrums represent the sum of many sources, some near, others farther away.

To perceive what is happening in the environment and adjust its behavior appropriately to the sound sources present, a listening organism must be able to disentangle the acoustic information from the many sources and evaluate the properties of individual events or sequences of events arising from a given source. At a more cognitive level, it is also useful to process the temporal relations among events in more lengthy sequences to understand the nature of actions on objects that are extended in time and that may carry important cultural messages such as in speech and music for humans. Finally, in many cases, with so much going on, listening must be focused on a given source of sound. Furthermore, this focusing process must possess dynamic characteristics that are tuned to the temporal evolution of the source that is being tracked in order to understand its message.

Aspects of these complex areas are addressed in this chapter to give a nonexhaustive flavor for current work in auditory perception and cognition. We focus on auditory scene analysis, timbre and sound source perception, temporal pattern processing, and attentional processes in hearing and finish with a consideration of developmental issues concerning these areas. The reader may wish to consult several general texts for additional information and inspiration (Bregman, 1990; Handel, 1989; McAdams & Bigand, 1993; Warren, 1999, with an accompanying CD), as well as compact discs of audio demos (Bregman & Ahad, 1995; Deutsch, 1995; Houtsma, Rossing & Wagenaars, 1987).

AUDITORY SCENE ANALYSIS

It is useful for an organism to build a mental representation of the acoustic environment in terms of the behavior of sound sources (objects set into vibration by actions upon them) in order to be able to structure its behavior in relation to them. We can hear in the same room and at the same time the noise of someone typing on a keyboard, the sound of someone walking, and the speech of someone talking in the next room. From a phenomenological point of view, we hear all of these sounds as if they arrive independently at our ears without distortion or interference among them, unless, of course, one source is much more intense than the others, in which case it would mask them, making them inaudible or at least less audible.

The acoustic waves of all sources are combined linearly in the atmosphere, and the composite waveform is then analyzed as such by the peripheral auditory system (Figure 10.1; see Chap. 9, this volume). Sound events are not opaque like most visual objects are. The computational problem is thus to interpret the complex waveform as a combination of sound-producing events. This process is called *auditory scene analysis* (Bregman, 1990) by analogy with the analysis of a visual scene in terms of objects (see Chap. 5, this volume, for a comparison of how these two sensory systems have come to solve analogous problems). Contrary to vision, in which a contiguous array of stimulation of the sensory organ corresponds to an object (although this is not always the case, as with partially occluded or transparent objects), in hearing the stimulation is a distributed frequency array mapped onto the basilar membrane. For a complex sound arising from a given source, the auditory system must thus reunite the sound components coming from the same source that have previously been channeled into separate auditory nerve fibers on the basis of their frequency content. Further, it must separate the information coming from distinct sources that contain close frequencies that would stimulate the same auditory nerve fibers. This is the problem of *concurrent organization*. The problem of *sequential organization* concerns perceptually connecting (or binding) over time successive events emitted by the same source and segregating events

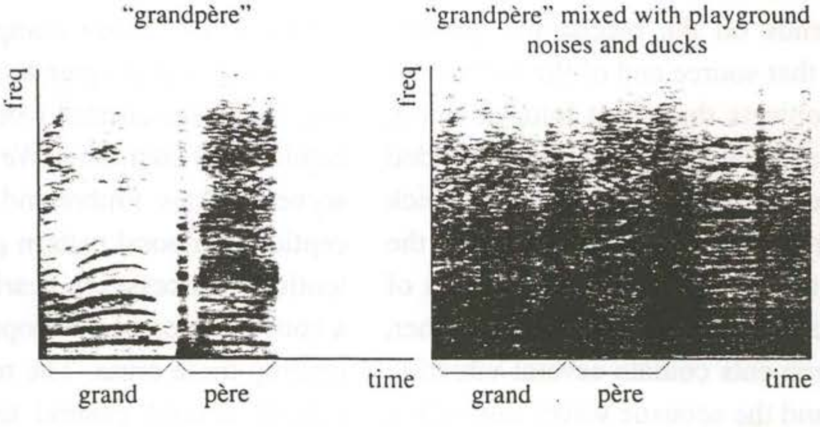


Figure 10.1 Spectrogram of (a) a target sound—the word *grandpère* ("grandfather" in French)—and (b) the target sound embedded in a noisy environment (a children's playground with voices and ducks).
NOTE: A spectrogram represents time on the horizontal axis and frequency on the vertical axis. The level at a given frequency is coded by the darkness of the spectrographic trace. Note that in many places in the mixture panel, the frequency information of the target sound is strongly overlapped by that of the noisy environment. In particular, the horizontal lines representing harmonic frequency components of the target word become intermingled with those of other voices in the mixture.

coming from independent sources in order to follow the message of only one source at a time.

This section examines the mechanisms that are brought into play by the auditory system to analyze the acoustic events and the behavior over time of sound sources. The ultimate goal of such a system would be to segregate perceptually actions that occur simultaneously; to detect new actions in the environment; to follow actions on a given object over time; to compute the properties of sources to feed into categorization, recognition, identification, and comprehension processes; and to use knowledge of derived attributes to track and extract sources and messages. We consider in order the processes involved in auditory event formation (concurrent grouping), the distinction of new event arrival from change of an ongoing event, auditory stream formation (sequential grouping), the interaction of concurrent and sequential grouping factors, the problem posed by the transparency of auditory events, and, finally, the role of schema-based processes in auditory organization.

Auditory Event Formation (Concurrent Grouping)

The processes of concurrent organization result either in the *perceptual fusion* or *grouping* of components of the auditory sensory representation into a single auditory event or in their *perceptual segregation* into two or more distinct events that overlap in time. The nature of these components of the sensory representation depends on the dual coding scheme in the auditory periphery. On the one hand, different parts of the acoustic frequency spectrum are represented in separate anatomical locations at many levels of the auditory system, a representation that is called tonotopic (see Chap. 9, this volume). On the other hand, even within a small frequency range in which all the acoustic information is carried by a small number of adjacent auditory nerve fibers, different periodicities in the stimulating waveform can be discerned on the basis of the temporal pattern of neural discharges that are time-locked to the stimulating waveform (see Chap. 9, this volume). The term *auditory event* refers to the unity and limited temporal extent that are experienced when, for example, a single sound source is set into vibration by a time-limited action on it. Some authors use the term *auditory objects,* but we prefer to distinguish objects (as vibrating physical sources) from perceptual events. A single source can produce a series of events.

A relatively small number of acoustic cues appear to signal either common behavior among acoustic components (usually arising from a single source) or incoherent behavior between components arising from distinct sources. The relative contribution of a given cue for scene analysis, however, depends on the perceptual task in which the listener is engaged: Some cues are more effective in signaling grouping for one attribute, such as identifying the pitch or vowel quality of a sound, than for another attribute, such as judging its position in space. Furthermore, some cues are more resistant than are others to environmental transformations of the acoustic waves originating from a vibrating object (reflections, reverberation, filtering by selective absorption, etc.).

Candidate cues for increasing segregation of concurrent sounds include inharmonicity, irregularity of spacing of frequency components, asynchrony of onset or offset of components, incoherence of change over time of level and frequency of components, and differences in spatial position.

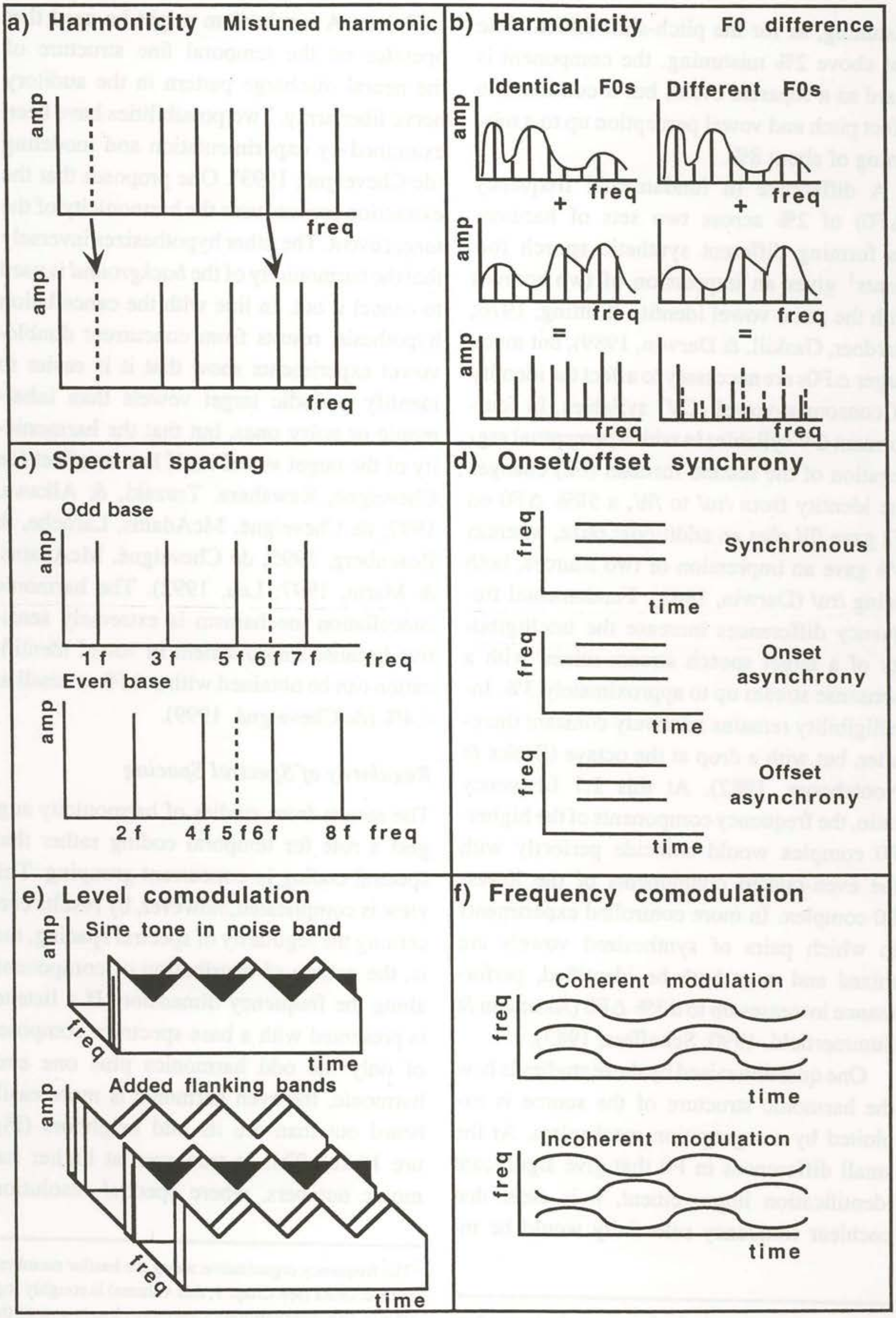
Harmonicity

In the environment two unrelated sounds rarely have frequency components that line up such that each frequency is an integer multiple of the fundamental frequency (F0), which is called a harmonic series. It is even less likely that they will maintain this relation with changes in frequency over time. A mechanism that is sensitive to deviations from harmonicity and groups components having harmonic relations could be useful for grouping acoustic components across the spectrum that arise from a single source and for segregating those that arise from distinct sources.

Two main classes of stimuli have been used to study the role of harmonicity in concurrent grouping: harmonic complexes with a single component mistuned from its purely harmonic relation and complexes composed of two or more sets of harmonic components with a difference in fundamental frequency (Figures 10.2a–b).

Listeners report hearing out a single, mistuned harmonic component from the rest of the complex tone if its harmonic rank is low and the mistuning is around 2% of its nominal frequency (Moore, Peters, & Glasberg, 1985). If mistuning is sufficient, listeners can match the pitch of the segregated harmonic, but this ability deteriorates at component frequencies above approximately 2000 Hz, where temporal information in the neural discharge pattern is no longer reliably related to waveform periodicities (Hartmann, McAdams, & Smith, 1990). A mistuned harmonic can also affect the virtual pitch (see Chap. 11, this volume) of the whole complex, pulling it in the direction of mistuning. This pitch shift increases for mistunings up to 3% and then decreases beyond that, virtually disappearing beyond about 8% (Hartmann, 1988; Hartmann et al., 1990; Moore, Glasberg, & Peters, 1985). This relation between mistuning and pitch shift suggests a harmonic-template model with a tolerance function on the harmonic sieve (Duifhuis, Willems, & Sluyter, 1982) or a time-domain autocoincidence processor (de Cheveigné, 1993) with a temporal margin of error. Harmonic mistuning can also affect vowel perception by influencing whether the component frequency is integrated into the computation of the spectral envelope that determines the vowel identity (Darwin & Gardner, 1986). By progressively mistuning this harmonic, a change in vowel percept has been recorded up to about 8%

Figure 10.2 Stimuli used to test concurrent grouping cues.
NOTE: a) Harmonicity tested with the mistuned harmonic paradigm. A harmonic stimulus without the fundamental frequency still gives a pitch at that frequency (dashed line). A shift of at least 2% but no more than 8% in the frequency of the fourth harmonic causes the harmonic to be heard separately but still contributes to a shift in the pitch of the complex sound. b) Harmonicity tested with the concurrent vowel paradigm. In the left column two vowels (indicated by the spectral envelopes with formant peaks) have the same fundamental frequency (F0). The resulting spectrum is the sum of the two, and the new spectral envelope does not correspond to either of the vowels, making them difficult to identify separately. In the right column, the F0 of one of the vowels is shifted, and two separate groups of harmonics are represented in the periodicity information in the auditory nerve, making the vowels more easily distinguished. c) Spectral spacing. An even harmonic in an odd-harmonic base, or vice versa, is easier to hear out than are the harmonics of the base. d) Onset/offset asynchrony. When harmonics start synchronously, they are fused perceptually into a single perceptual event. An asynchrony of the onset of at least 30–50 ms makes the harmonic easier to hear out. An asynchrony of the offset has a relatively weak effect on hearing out the harmonic. e) Level comodulation (comodulation masking release). The amplitude envelopes of a sine tone (black) and a narrow band noise with a modulating envelope (white) are shown. The masking threshold of the sine tone in the noise is measured. When flanking noise bands with amplitude envelopes identical to that of the on-signal band are added, the masked threshold of the sine tone decreases by about 3 dB. f) Frequency comodulation. A set of harmonics that are coherently modulated in frequency (with a sinusoidal vibrato in this example) are heard as a single event. Making the modulation incoherent on one of the harmonics makes it easier to hear out because of the time-varying inharmonicity that is created.


a) Harmonicity Mistuned harmonic
amp
freq
amp
freq
b) Harmonicity F0 difference
Identical F0s
Different F0s
amp
freq
+
freq
+
amp
freq
freq
=
=
amp
freq
freq
c) Spectral spacing
Odd base
amp
1 f 3 f 5 f 6 f 7 f freq
Even base
amp
2 f 4 f 5 f 6 f 8 f freq
d) Onset/offset synchrony
freq
Synchronous
time
freq
Onset
asynchrony
time
freq
Offset
asynchrony
time
e) Level comodulation
amp
Sine tone in noise band
freq
time
amp
Added flanking bands
freq
time
f) Frequency comodulation
Coherent modulation
freq
time
Incoherent modulation
freq
time

mistuning, as for the pitch-shift effect. Note that above 2% mistuning, the component is heard as a separate event, but it continues to affect pitch and vowel perception up to a mistuning of about 8%.

A difference in fundamental frequency (ΔF0) of 2% across two sets of harmonics forming different synthetic speech formants[1] gives an impression of two sources with the same vowel identity (Cutting, 1976; Gardner, Gaskill, & Darwin, 1989), but much larger ΔF0s are necessary to affect the identity of consonant-vowel (CV) syllables. In four-formant CV syllables in which perceptual segregation of the second formant (F2) changes the identity from /ru/ to /li/, a 58% ΔF0 on F2 gave /li/ plus an additional buzz, whereas 4% gave an impression of two sources, both being /ru/ (Darwin, 1981). Fundamental frequency differences increase the intelligibility of a target speech stream mixed with a nonsense stream up to approximately 3%. Intelligibility remains relatively constant thereafter, but with a drop at the octave (Brokx & Nooteboom, 1982). At this 2:1 frequency ratio, the frequency components of the higher-F0 complex would coincide perfectly with the even-ranked components of the lower-F0 complex. In more controlled experiments in which pairs of synthesized vowels are mixed and must both be identified, performance increases up to a 3% ΔF0 (Assmann & Summerfield, 1990; Scheffers, 1983).

One question raised by these studies is how the harmonic structure of the source is exploited by a segregation mechanism. At the small differences in F0 that give significant identification improvement, it is clear that cochlear frequency selectivity would be insufficient. A mechanism might be used that operates on the temporal fine structure of the neural discharge pattern in the auditory nerve fiber array. Two possibilities have been examined by experimentation and modeling (de Cheveigné, 1993). One proposes that the extraction process uses the harmonicity of the *target* event. The other hypothesizes inversely that the harmonicity of the *background* is used to cancel it out. In line with the cancellation hypothesis, results from concurrent double-vowel experiments show that it is easier to identify periodic target vowels than inharmonic or noisy ones, but that the harmonicity of the target vowel itself has no effect (de Cheveigné, Kawahara, Tsuzaki, & Aikawa, 1997; de Cheveigné, McAdams, Laroche, & Rosenberg, 1995; de Cheveigné, McAdams, & Marin, 1997; Lea, 1992). The harmonic cancellation mechanism is extremely sensitive because improvement in vowel identification can be obtained with a ΔF0 as small as 0.4% (de Cheveigné, 1999).

Regularity of Spectral Spacing

The results from studies of harmonicity suggest a role for temporal coding rather than spectral coding in concurrent grouping. This view is complicated, however, by results concerning the regularity of spectral spacing, that is, the pattern of distribution of components along the frequency dimension. If a listener is presented with a base spectrum composed of only the odd harmonics plus one even harmonic, the even harmonic is more easily heard out than are its odd neighbors (Figure 10.2c). This is true even at higher harmonic numbers, where spectral resolution[2]

[1] Formants are regions in the frequency spectrum where the energy is higher than in adjacent regions. They are due to the resonance properties of the vocal tract and determine many aspects of consonant and vowel identity (see Chap. 12, this volume).

[2] The frequency organization along the basilar membrane in the cochlea (see Chap. 9, this volume) is roughly logarithmic, so higher harmonics are more closely spaced than are lower harmonics. At sufficiently high ranks, adjacent harmonics no longer stimulate separate populations of auditory nerve fibers and are thus "unresolved" in the tonotopic representation.

is reduced. Note that the even harmonic surrounded by odd harmonics would be less resolved on the basilar membrane than would either of its neighbors. Contrary to the ΔF0 cue, harmonic sieve and autocoincidence models cannot account for these results (Roberts & Bregman, 1991). Nor does the underlying mechanism involve a cross-channel comparison of the amplitude modulation envelope in the output of the auditory filter bank, because minimizing the modulation depth or perturbing the modulation pattern by adding noise does not markedly reduce the difference in hearing out even and odd harmonics (Roberts & Bailey, 1993). However, perturbing the regularity of the base spectrum by adding extraneous components or removing components reduces the perceptual "popout" of even harmonics (Roberts & Bailey, 1996), confirming the spectral pattern hypothesis.

Onset and Offset Asynchrony

Unrelated sounds seldom start or stop at exactly the same time. Therefore, the auditory system assumes that synchronous components are part of the same sound or were caused by the same environmental event. Furthermore, the auditory system is extremely sensitive to small asynchronies in analyzing the auditory scene. A single frequency component in a complex tone becomes audible on its own with an asynchrony as small as 35 ms (Rasch, 1978). Onset asynchronies are more effective than offset asynchronies are in creating segregation (Figure 10.2d; Dannenbring & Bregman, 1976; Zera & Green, 1993). When a component is made asynchronous, it also contributes less to the perceptual properties computed from the rest of the complex. For example, a 30-ms asynchrony can affect timbre judgments (Bregman & Pinker, 1978). Making a critical frequency component that affects the estimation of a vowel sound's spectral envelope asynchronous by 40 ms changes the vowel identity (Darwin, 1984). Furthermore, the asynchrony effect is abolished if the asynchronous portion of the component (i.e., the part that precedes the onset of the vowel complex) is grouped with another set of components that are synchronous with it alone and that have a common F0 that is different from that of the vowel. This result suggests that it is indeed a grouping effect, not the result of adaptation (Darwin & Sutherland, 1984).

The effect of a mistuned component on the pitch of the complex (discussed earlier) is increasingly reduced for asynchronies from 80 to 300 ms (Darwin & Ciocca, 1992). This latter effect is weakened if another component groups with a preceding portion of the asynchronous component (Ciocca & Darwin, 1993). Note that in these results the asynchronies necessary to affect pitch perception are much greater than are those that affect vowel perception (Hukin & Darwin, 1995a).

Coherence of Change in Level

From Gestalt principles such as common fate (see Chap. 5, this volume), one might expect that common direction of change in level would be a cue for grouping components together; inversely, independent change would signal that segregation was appropriate. The evidence that this factor is a grouping cue, however, is rather weak. In experiments by Hall and colleagues (e.g., Hall, Grose, & Mendoza, 1995), a phenomenon called *comodulation masking release* is created by placing a narrow-band noise masker centered on a target frequency component (sine tone) that is to be detected (Figure 10.2e). The masked threshold of the tone is measured in the presence of the noise. Then, noise bands with similar or different amplitude envelopes are placed in more distant frequency regions. The presence of similar envelopes (i.e., comodulation) makes it possible to detect the tone in the noise at a level of about 3 dB lower than in their absence. The masking seems to be released to some extent by the presence of comodulation on the distant noise bands. Some authors have attributed this phenomenon to the grouping of the noise bands into a single auditory image that then allows the noise centered on the tone to be interpreted as part of a different source, thus making detection of the tone easier (Bregman, 1990, chap. 3). Others, however, consider either that cross-channel detection of the amplitude envelope simply gives a cue to the auditory system concerning when the masking noise should be in a level dip, or that the flanking maskers suppress the on-signal masker (Hall et al., 1995; McFadden & Wright, 1987).

Coherence of Change in Frequency

For sustained complex sounds that vary in frequency, there is a tendency for all frequencies to change synchronously and to maintain the frequency ratios. As such, one might imagine that frequency modulation coherence would be an important cue in source grouping (Figure 10.2f). The effects of frequency modulation incoherence may have two origins: within-channel cues and cross-channel cues. Within-channel cues would result from the interactions of unresolved components that changed frequency incoherently over time, creating variations in beating or roughness in particular auditory channels. They could signal the presence of more than one source. Such cues are detectable for both harmonic and inharmonic stimuli (McAdams & Marin, 1990) but are easier to detect for the former because of the reliability of within-channel cues for periodic sounds. Frequency modulation coherence is not, however, detectable across auditory channels (i.e., in distant frequency regions) above and beyond the mistuning from harmonicity that they create (Carlyon, 1991, 1992, 1994). Although frequency modulation increases vowel prominence when the $\Delta F0$ is already large, there is no difference between coherent and incoherent modulation across the harmonics of several vowels either on vowel prominence (McAdams, 1989) or on vowel identification (Summerfield & Culling, 1992). However, frequency modulation can help group together frequency components for computing pitch. In a mistuned harmonic stimulus, shifts in the perceived pitch of the harmonic complex continue to occur at greater mistunings when all components are modulated coherently than when they are unmodulated (Darwin, Ciocca, & Sandell, 1994).

Spatial Position

It was thought early on that different spatial positions should give rise to binaural cues that could be used to segregate temporally and spectrally overlapping sound events. Although work on speech comprehension in noisy environments (e.g., Cherry's 1953 "cocktail party effect") emphasized spatial cues to allow listeners to ignore irrelevant sources, the evidence in support of such cues for grouping is in fact quite weak. An interaural time difference (ITD) is clearly a powerful cue for direction (see Chap. 9, this volume), but it is remarkably ineffective as a cue for grouping simultaneous components that compose a particular source (Culling & Summerfield, 1995; Hukin & Darwin, 1995b).

The other principal grouping cues generally override spatial cues. For example, the detection of changes in ITD on sine components across two successive stimulus intervals is similar when they are presented in isolation or embedded within an inharmonic complex. However, detection performance is much worse when they are embedded within a harmonic complex; thus harmonicity overrides spatial incoherence (Buell & Hafter, 1991). Furthermore, mistuning a component can affect its lateralization (Hill & Darwin, 1996), suggesting that grouping takes place on the basis of harmonicity, and only *then* is the spatial position computed on the basis of the lateralization cues for the set of components that have been grouped together (Darwin & Ciocca, 1992).

Lateralization effects may be more substantial when the spatial position is attended to over an extended time, as would be the case in paying sustained attention to a given sound source in a complex environment (Darwin & Carlyon, 1995). Listeners can attend across time to one of two spoken sentences distinguished by small differences in ITD, but they do not use such continuity of ITD to determine which individual frequency components should form part of a sentence. These results suggest that ITD is computed on the peripheral representation of the frequency components in parallel to a grouping of components on the basis of harmonicity and synchrony. Subsequently, direction is computed on the grouped components, and the listener attends to the direction of the grouped object (Darwin & Hukin, 1999).

General Considerations Concerning Concurrent Grouping

Note that there are several possible cues for grouping and segregation, which raises the possibility that what the various cues signal in terms of source structures in the environment can diverge. For example, many kinds of sound sources are not harmonic, but the acoustic components of the events produced by them would still start and stop at the same time and probably have a relatively fixed spatial position that could be attended to. In many cases, however, redundancy of segregation and integration cues works against ambiguities in inferences concerning grouping on the basis of sensory information. Furthermore, the cues to scene analysis are not all-or-none. The stronger they are, the more they affect grouping, and the final perceptual result is the best compromise on the basis of both the strength of the evidence available and the perceptual task in which the listener is engaged (Bregman, 1993). As many of the results cited earlier demonstrate, the grouping and segregation of information in the auditory sensory representation precedes and thus determines the perceptual properties of a complex sound source, such as its spatial position, its pitch, or its timbre. However, the perceived properties can in turn become cues that facilitate sustained attending to, or tracking of, sound sources over time.

New Event Detection versus Perception of a Changing Event

The auditory system appears to be equipped with a mechanism that triggers event-related computation when a sudden change in the acoustic array is detected. The computation performed can be a resampling of some property of the environment, such as the spatial position of the source, or a grouping process that results in the decomposition of an acoustic mixture (Bregman, 1991). This raises the questions of what constitutes a sudden change indicating the arrival of a new event and how it can be distinguished from a more gradual change that results from an evolution of an already present event.

An example of this process is binaural adaptation and the recovery from such adaptation when an acoustic discontinuity is detected. Hafter, Buell, & Richards (1988) presented a rapid (40/s) series of clicks binaurally with an interaural time difference that gave a specific lateralization of the click train toward the leading ear (Figure 10.3a). As one increases the number of clicks in the train, accuracy in discriminating the spatial position between two successive click trains increases, but the improvement is progressively less (according to a compressive power function) as the click train is extended in duration. The binaural system thus appears to become

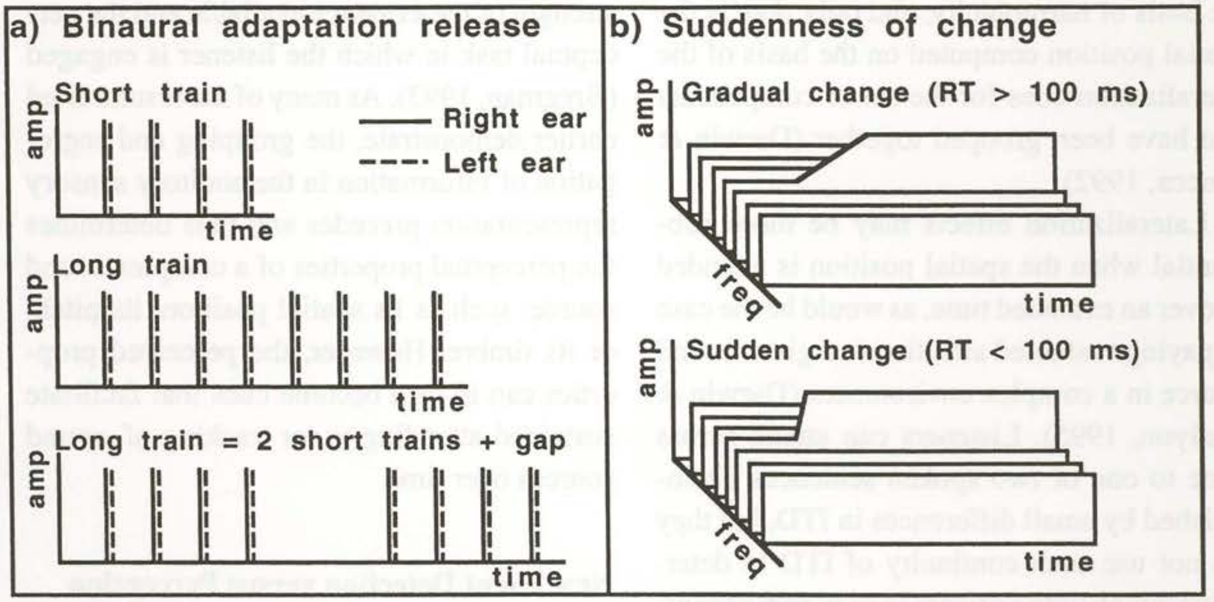


Figure 10.3 Stimuli used to test the resetting of auditory sampling of the environment upon new event detection.
NOTE: a) Binaural adapatation release. A train of clicks (interclick separation = 2.5 ms) is sent to the two ears with a small interaural time difference (ITD) that displaces the perceived lateralization of the sound toward the leading ear (the right ear in this example). The just noticeable ITD decreases as a function of the number of clicks in the train, but the relative contribution of later clicks is lesser than is that of the earlier clicks, indicating binaural adaptation. Release from adaptation is triggered by the detection of a new event, such as a discontinuity in the click train (e.g., a silent gap of 7.5 ms). b) Suddenness of change. The amplitude envelopes on harmonic components are shown. All harmonics are constant in level except one, which increases in level in the middle. A slow change in level (>100 ms) is heard as a change in the timbre of the event, whereas a sudden change (<100 ms) is heard as a new (pure-tone) event.

progressively quiet beyond stimulus onset for constant stimulation. However, if some kind of discontinuity is introduced in the click train (a longer or shorter gap between clicks, or a brief sound with a sudden onset in a remote spectral region, even of fairly low intensity), the spatial environment is suddenly resampled at the moment of the discontinuity. A complete recovery from the process of binaural adaptation appears to occur in the face of such discontinuities and indicates that the auditory system is sensitive to perturbations of regularity. Hafter and Buell (1985) proposed that at a fairly low level in the auditory system, multiple bands are monitored for changes in level that might accompany the start of a new signal or a variation in the old one. Sudden changes cause the system to resample the binaural inputs and to update its spatial map at the time of the restart, suggesting that knowledge about the direction of a source may rely more on memory than on the continual processing of ongoing information.

Similarly, an acoustic discontinuity can provoke the emergence of a new pitch in an otherwise continuous complex tone. A sudden interaural phase disparity or frequency disparity in one component of a complex tone can create successive-difference cues that make the component emerge (Kubovy, 1981; Kubovy, Cutting, & McGuire, 1974). In this case, the successive disparity triggers a recomputation of which pitches are present. Thus, various sudden changes trigger resampling. But how fast a change is "sudden"? If listeners must identify the direction of change in pitch for successive pure-tone events added in phase to a continuous harmonic complex (Figure 10.3b), performance is a monotone decreasing function of rise time; that is, the more sudden the change, the more the change is perceived as a new event with its own pitch, and the better is the performance. From these results Bregman, Ahad, Kim, and Melnerich (1994) proposed that "sudden" can be defined as basically less than 100 ms for onsets.

Auditory Stream Formation (Sequential Grouping)

The processes of sequential organization result in the perceptual integration of successive events into a single auditory stream or their perceptual segregation into two or more streams. Under everyday listening conditions, an auditory stream corresponds to a sequence of events emitted by a single sound source.

General Considerations Concerning Sequential Grouping

Several basic principles of auditory stream formation emerge from research on sequential grouping. These principles reflect regularities in the physical world that shaped the evolution of the auditory mechanisms that detect them.

1. *Source properties change slowly.* Sound sources generally emit sequences of events that are transformed in a progressive manner over time. Sudden changes in event properties are likely to signal the presence of several sources (Bregman, 1993).
2. *Events are allocated exclusively to streams.* A given event is assigned to one or another stream and cannot be perceived as belonging to both simultaneously (Bregman & Campbell, 1971), although there appear to be exceptions to this principal in interactions between sequential and concurrent grouping processes and in duplex perception (discussed later).
3. *Streaming is cumulative.* The auditory system appears by default to assume that a sequence of events arises from a single source until enough evidence to the contrary can be accumulated, at which point segregation occurs (Bregman, 1978b). Also, if a cyclical sequence is presented over a long period of time (several tens of seconds), segregation tends to increase (Anstis & Saida, 1985).
4. *Sequential grouping precedes stream attribute computation.* The perceptual properties of sequences depend on what events are grouped into streams, as was shown for concurrent grouping and event attributes. A corollary of this point is the fact that the perception of the order of events depends on their being assigned to the same stream: It is easier to judge temporal order on within-stream patterns than on across-stream patterns that are perceptually fragmented (Bregman & Campbell, 1971; van Noorden, 1975).

The cues that determine sequential auditory organization are closely related to the Gestalt principles of proximity and similarity (see Chap. 5, this volume). The notion of proximity in audition is limited here to the temporal distance between events, and similarity encompasses the acoustic similarity of successive events. Given the intrinsically temporal nature of acoustic events, grouping is considered in terms of continuity and rate of change in acoustic properties between successive events. In considering the acoustic factors that affect grouping in the following, keep in mind that not all acoustic differences are equally important in determining segregation (Hartmann & Johnson, 1991).

Frequency Separation and Temporal Proximity

A stimulus sequence composed of two alternating frequencies in the temporal pattern

ABA—ABA— (where "—" indicates a silence) is heard as a galloping rhythm if the tones are integrated into a single stream and as two isochronous sequences (A—A—A—A— and B———B———) if they are segregated. At slower tempos and smaller frequency separations, integration tends to occur, whereas at faster tempos and larger frequency separations, segregation tends to occur. Van Noorden (1975) measured the frequency separation at which the percept changes from integration to segregation or vice versa for various event rates. If listeners are instructed to try to hear the gallop rhythm or conversely to focus on one of the isochronous sequences, temporal coherence and fission boundaries are obtained, respectively (see Figure 10.4). These functions do not have the same form. The fission boundary is limited by the frequency resolution of the peripheral auditory system and is relatively unaffected by the event rate. The temporal coherence boundary reflects the limits of inevitable segregation and strongly depends on tempo. Between the two is an ambiguous region where the listener's perceptual intent plays a strong role.

Streaming is not an all-or-none phenomenon with clear boundaries between integration and segregation along a given sensory continuum, however. In experiments in which the probability of a response related to the degree of segregation was measured (Brochard, Drake, Botte, & McAdams, 1999), the probability varied continuously as a function of frequency separation. This does not imply that the percept is ambiguous. It is either one stream or two streams, but the probability of hearing one or the other varies for a given listener and across listeners.

It is not the absolute frequency difference that determines which tones are bound together in the same stream, but rather the relative differences among the frequencies. Bregman (1978a), for example, used a sequential tone pattern ABXY. If A and B are within a critical band (i.e., they stimulate overlapping sets of auditory nerve fibers) in a high frequency region, and if X and Y are within a critical band in a low frequency region, then A and B form one stream, and X and Y form another stream (see Figure 10.5). If X and Y are now moved to the same frequency region as A and B such that A and X are close and B and Y are close, without changing the frequency ratios between A and B nor between X and Y, then the relative frequency differences predominate and streams of A-X and B-Y are obtained.

The abruptness of transition from one frequency to the next also has an effect on stream segregation. In the studies just cited, one tone stops on one frequency, and the next tone begins at a different frequency. In many sound sources that produce sequences of events and vary the fundamental frequency, such as the voice, such changes may be more gradual. Bregman & Dannenbring (1973) showed that the inclusion of frequency ramps (going toward the next tone at the end and coming from the previous tone at the beginning) or even complete frequency glides between tones yielded greater integration of the sequence into a single stream.

The Cumulative Bias toward Greater Segregation

Anstis and Saida (1985) showed that there is a tendency for reports of a segregated percept to increase over time when listening to alternating-tone sequences. This stream biasing decays exponentially when the stimulus sequence is stopped and has a time constant of around 4 s on average (Beauvois & Meddis, 1997). Anstis and Saida proposed a mechanism involving the fatigue of frequency jump detectors to explain this phenomenon, but Rogers and Bregman (1993a) showed that an inductor sequence with a single tone could induce a bias toward streaming in the absence of jumps. The biasing mechanism requires

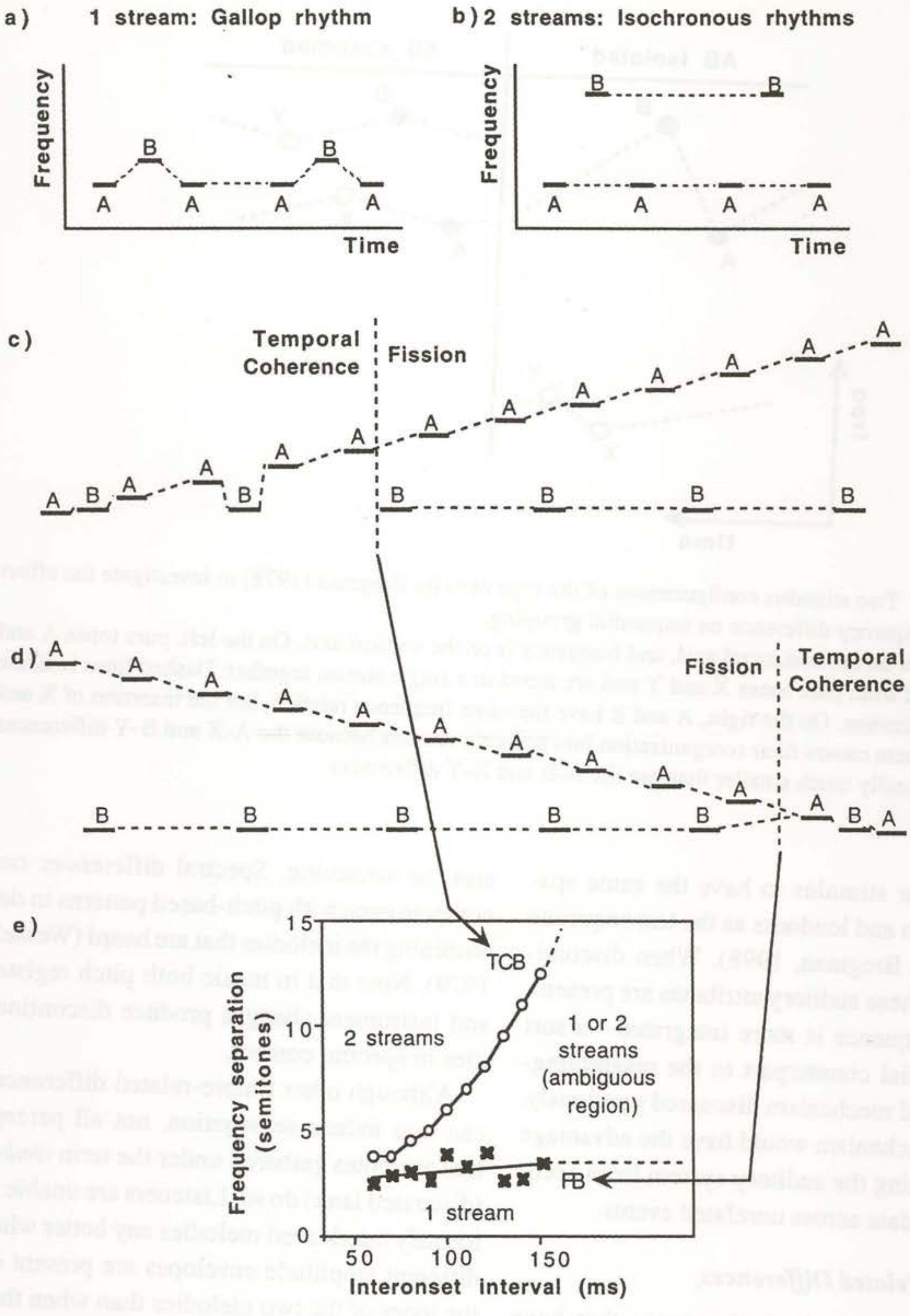


Figure 10.4 Van Noorden's temporal coherence and fission boundaries.

NOTE: A repeating "ABA—" pattern can give a percept of either a) a gallop rhythm or b) two isochronous sequences, depending on the presentation rate and AB frequency difference. c) To measure the temporal coherence boundary (TCB), the initial frequency difference is small and increases while the listener attempts to hold the gallop percept. d) To measure the fission boundary (FB), the initial difference is large and is decreased while the listener tries to focus on a single isochronous stream. In both cases, the frequency separation at which the percept changes is recorded. The whole procedure is repeated at different interonset intervals, giving the curves shown in (e).

SOURCE: Adapted from van Noorden (1975, Figure 2.7). Copyright © 1975 by Leon van Noorden. Adapted with permission.

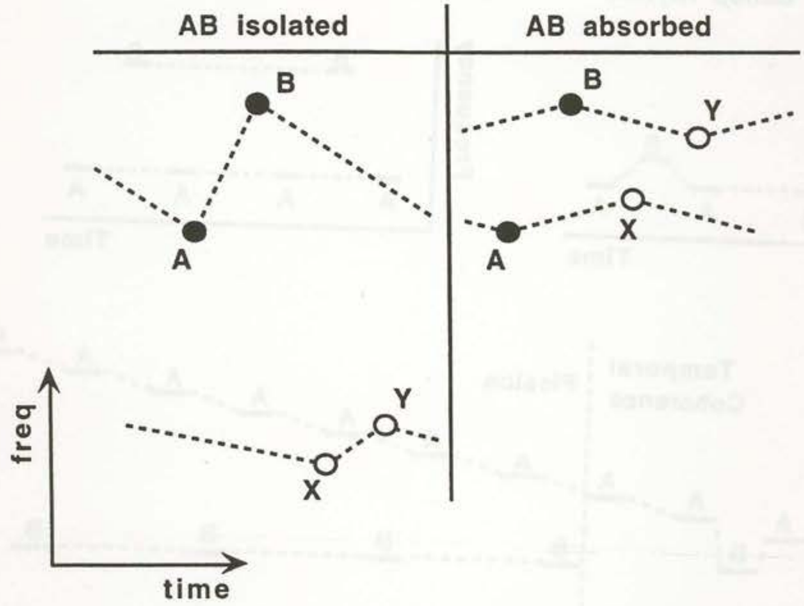


Figure 10.5 Two stimulus configurations of the type used by Bregman (1978) to investigate the effect of relative frequency difference on sequential grouping.
NOTE: Time is on the horizontal axis, and frequency is on the vertical axis. On the left, pure tones A and B are isolated from pure tones X and Y and are heard in a single stream together. Dashed lines indicate stream organization. On the right, A and B have the same frequency relation, but the insertion of X and Y between them causes their reorganization into separate streams because the A-X and B-Y differences are proportionally much smaller than are the A-B and X-Y differences.

the inductor stimulus to have the same spatial location and loudness as the test sequence (Rogers & Bregman, 1998). When discontinuities in these auditory attributes are present, the test sequence is more integrated—a sort of sequential counterpart to the resampling-on-demand mechanism discussed previously. Such a mechanism would have the advantage of preventing the auditory system from accumulating data across unrelated events.

Timbre-Related Differences

Sequences with alternating tones that have the same fundamental frequency (i.e., same virtual pitch) but that are composed of differently ranked harmonics derived from that fundamental (i.e., different timbres) tend to segregate (Figure 10.6a; van Noorden, 1975). Differences in spectral content can thus cause stream segregation (Hartmann & Johnson, 1991; Iverson, 1995; McAdams & Bregman, 1979). It is therefore not pitch per se that creates the streaming. Spectral differences can compete even with pitch-based patterns in determining the melodies that are heard (Wessel, 1979). Note that in music both pitch register and instrument changes produce discontinuities in spectral content.

Although other timbre-related differences can also induce segregation, not all perceptual attributes gathered under the term *timbre* (discussed later) do so. Listeners are unable to identify interleaved melodies any better when different amplitude envelopes are present on the tones of the two melodies than when they are absent, and differences in auditory roughness are only weakly useful for melody segregation, and only for some listeners (Hartmann & Johnson, 1991). However, dynamic (temporal) cues can contribute to stream segregation. Iverson (1995) played sequences that alternated between different musical instruments at the same pitch and asked listeners for ratings of the degree of segregation.

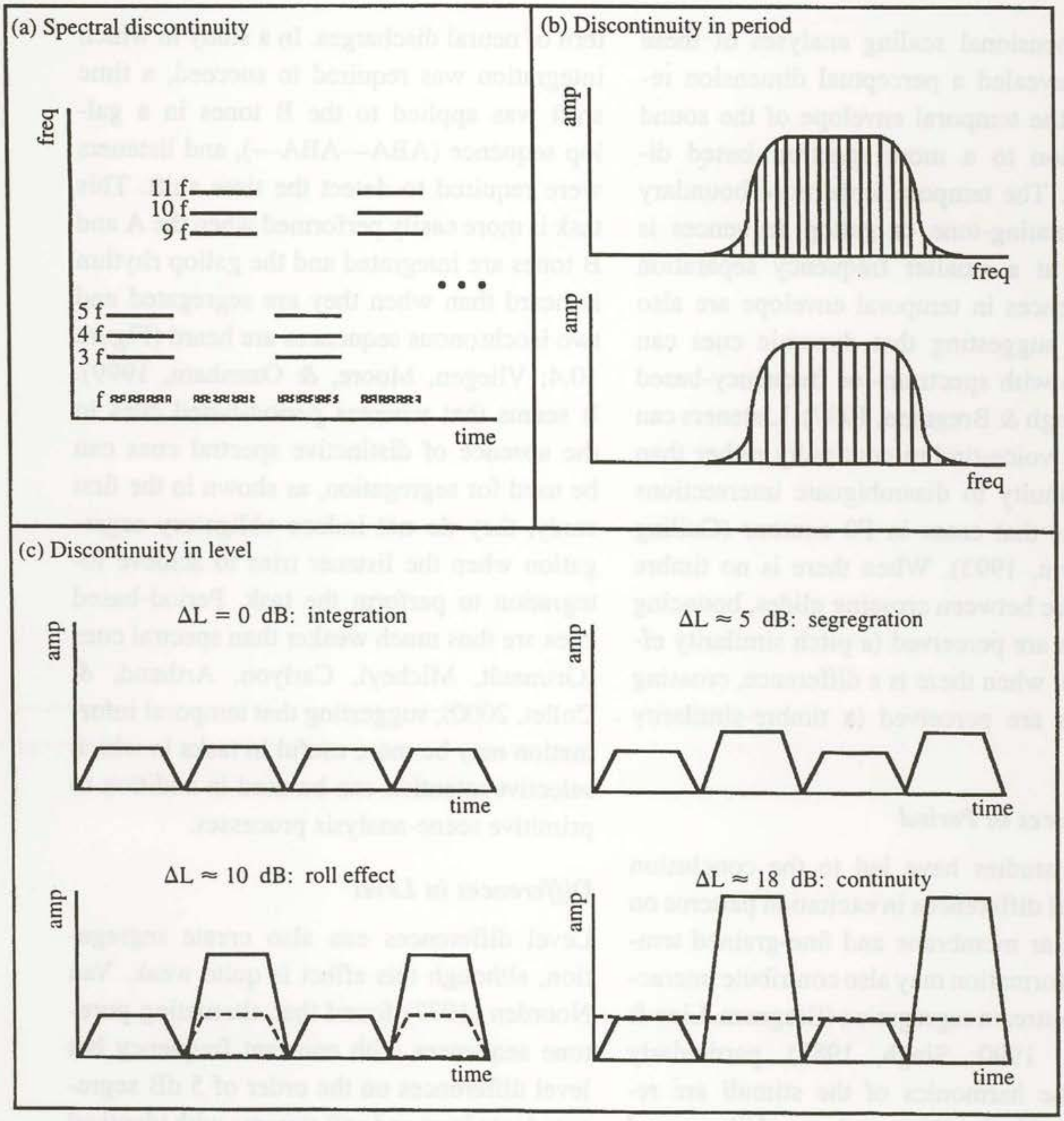


Figure 10.6 Stimuli used to test sequential grouping cues.

NOTE: a) Spectral discontinuity. An alternating sequence of tones with identical fundamental frequencies gives rise to a perception of constant pitch but differing timbres when the spectral content of the tones are different. This discontinuity in spectral content also creates a perceptual segregation into two streams. b) Discontinuity in period. A harmonic complex that is filtered in the high-frequency region gives rise to a uniform pattern of excitation on the basilar membrane, even if the period of the waveform (the fundamental frequency) is changed. The upper diagram has a lower F0 than has the lower diagram. There can be no cue of spectral discontinuity in a sequence of tones that alternates between these two sounds, yet segregation occurs on the basis of the difference in period, presumably carried by the temporal pattern of neural discharges in the auditory nerve. c) Discontinuity in level. A sequence of pure tones of constant frequency but alternating in level gives rise to several percepts depending on the relative levels. A single stream is heard if the levels are close. Two streams at half the tempo are heard if the levels differ by about 5 dB. A roll effect in which a louder half-tempo stream is accompanied by a softer full-tempo stream is obtained at certain rapid tempi when the levels differ by about 10 dB. Finally, at higher tempi and large differences in level, a louder pulsing stream is accompanied by a softer continuous tone.

Multidimensional scaling analyses of these ratings revealed a perceptual dimension related to the temporal envelope of the sound in addition to a more spectrum-based dimension. The temporal coherence boundary for alternating-tone or gallop sequences is situated at a smaller frequency separation if differences in temporal envelope are also present, suggesting that dynamic cues can combine with spectrum- or frequency-based cues (Singh & Bregman, 1997). Listeners can also use voice-timbre continuity rather than F0 continuity to disambiguate intersections in voices that cross in F0 contour (Culling & Darwin, 1993). When there is no timbre difference between crossing glides, bouncing contours are perceived (a pitch similarity effect), but when there is a difference, crossing contours are perceived (a timbre-similarity effect).

Differences in Period

Several studies have led to the conclusion that local differences in excitation patterns on the basilar membrane and fine-grained temporal information may also contribute interactively to stream segregation (Bregman, Liao & Levitan, 1990; Singh, 1987), particularly when the harmonics of the stimuli are resolved on the basilar membrane. Vliegen and Oxenham (1999) used an interleaved melody recognition task in which the tones of a target melody were interleaved with those of a distractor sequence. If segregation does not occur at least partially, recognition of the target is nearly impossible. In one condition, they applied a band-pass filter that let unresolved harmonics through (Figure 10.6b). Because the harmonics would not be resolved in the peripheral auditory system, there would be no cue based on the tonotopic representation that could be used to segregate the tones. However, segregation did occur, most likely on the basis of cues related to the periods of the waveforms carried in the temporal pattern of neural discharges. In a study in which integration was required to succeed, a time shift was applied to the B tones in a gallop sequence (ABA—ABA—), and listeners were required to detect the time shift. This task is more easily performed when the A and B tones are integrated and the gallop rhythm is heard than when they are segregated and two isochronous sequences are heard (Figure 10.4; Vliegen, Moore, & Oxenham, 1999). It seems that whereas period-based cues in the absence of distinctive spectral cues can be used for segregation, as shown in the first study, they do not induce obligatory segregation when the listener tries to achieve integration to perform the task. Period-based cues are thus much weaker than spectral cues (Grimault, Micheyl, Carlyon, Arthaud, & Collet, 2000), suggesting that temporal information may be more useful in tasks in which selective attention can be used in addition to primitive scene-analysis processes.

Differences in Level

Level differences can also create segregation, although this effect is quite weak. Van Noorden (1977) found that alternating pure-tone sequences with constant frequency but level differences on the order of 5 dB segregated into loud and soft streams with identical tempi (Figure 10.6c). Hartmann and Johnson (1991) also found a weak effect of level differences on interleaved melody recognition performance. When van Noorden increased the level difference and the sequence rate was relatively fast (greater than 13 tones/s), other perceptual effects began to emerge. For differences of around 10 dB, a percept of a louder stream at one tempo accompanied by a softer stream at twice that tempo was obtained. For even greater differences (> 18 dB), a louder intermittent stream was accompanied by a continuous softer stream. In both cases, the more intense event would seem to be interpreted as being composed of two events of identical spectral content. These percepts are examples of what Bregman (1990, chap. 3) has termed the *old-plus-new heuristic* (discussed later).

Differences in Spatial Location

Dichotically presented alternating-tone sequences do not tend to integrate into a trill percept even for very small frequency separations (van Noorden, 1975). Similarly, listeners can easily identify interleaved melodies presented to separate ears (Hartmann & Johnson, 1991). Ear of presentation is not, however, a sufficient cue for segregation. Deutsch (1975) presented simultaneously ascending and descending musical scales such that the notes alternated between ears; that is, the frequencies sent to a given ear hopped around (Figure 10.7). Listeners reported hearing an up-down pattern in one ear and a down-up pattern in the other, demonstrating an organization based on frequency proximity despite the alternating ear of presentation. An interaural time difference is slightly less effective in creating segregation than is dichotic presentation (Hartmann & Johnson, 1991).

Figure 10.7 Melodic patterns of the kind used by Deutsch (1975).

NOTE: a) Crossing scales are played simultaneously over headphones. In each scale, the tones alternate between left (L) and right (R) earpieces. b) The patterns that would be heard if the listener focused on a given ear. c) The patterns reported by listeners.

Interactions between Concurrent and Sequential Grouping Processes

Concurrent and sequential organization processes are not independent. They can interact and even enter into competition, the final perceptual result depending on the relative organizational strength of each one. In the physical environment, there is a fairly good consensus among the different concurrent and sequential grouping cues. However, under laboratory conditions or as a result of compositional artifice in music, they can be made to conflict with one another. Bregman and Pinker (1978) developed a basic stimulus (Figure 10.8) for testing the situation in which a concurrent organization (fusion or segregation of B and C) and a sequential organization (integration or segregation of A and B) were in competition for the same component (B). When the sequential organization is reinforced by the frequency proximity of A and B and the concurrent organization is weakened by the asynchrony of B and C, A and B form a single stream, and C is perceived with a pure timbre. If the concurrent organization is reinforced by synchrony while the sequential organization is weakened by separating A and B in frequency, A forms a stream by itself, and B fuses with C to form a second stream with a richer timbre.

The Transparency of Auditory Events

In line with the belongingness principle of the Gestalt psychologists, Bregman (1990,

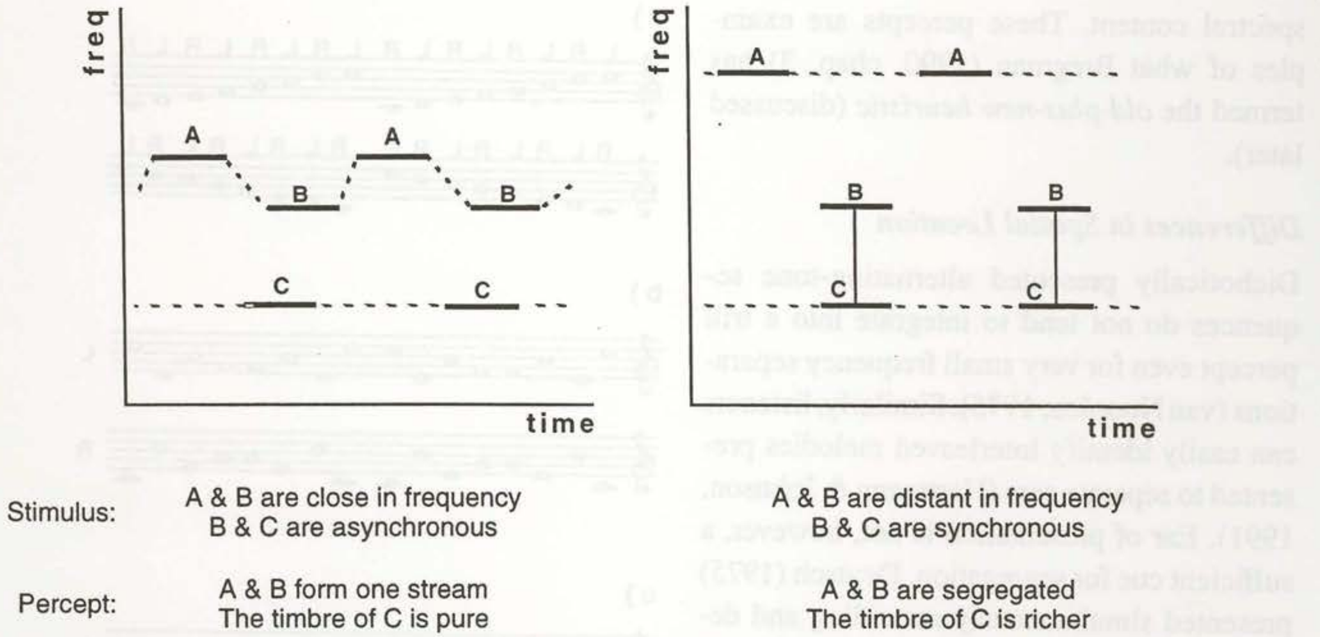


Figure 10.8 Schematic representative of some of the stimulus configurations used by Bregman and Pinker (1978) to study the competition between concurrent and sequential grouping processes.
NOTE: Pure tone A alternates with a complex tone composed of pure tones B and C. The relative frequency proximity of A and B and the asynchrony of B and C are varied. When A and B are close in frequency and B and C are sufficiently asynchronous (left diagram), an AB stream is formed, and C is perceived as having a pure timbre. When A and B are distant in frequency and B and C are synchronous (right diagram), A forms a stream by itself, and B and C fuse into a single event with a richer timbre.

chap. 7) has proposed the principle of exclusion allocation: A given bit of sensory information cannot belong to two separate perceptual entities simultaneously. In general, this principle seems to hold: Parts of a spectrum that do not start at the same time are exhaustively segregated into temporally overlapping events, and tones presented sequentially are exhaustively segregated into streams. There are, however, several examples of both speech and nonspeech sounds that appear to violate this principle.

Duplex Perception of Speech

If the formant transition specifying a stop consonant such as /b/ (see Chap. 12, this volume) is excised from a consonant-vowel syllable and is presented by itself, a brief chirp sound is heard. The remaining base part of the original sound without the formant transition gives a /da/ sound. If the base and transition are remixed in the same ear, a /ba/ sound results. However, when the formant transition and base sounds are presented to opposite ears, listeners hear both a /ba/ sound in the ear with the base (integration of information from the two ears to form the syllable) and a simultaneous chirp in the other ear (Cutting, 1976; Rand, 1974). The formant transition thus contributes both to the chirp and to the /ba/—hence the term *duplex*. It is not likely that this phenomenon can be explained by presuming that speech processing is unconstrained by primitive scene analysis mechanisms (Darwin, 1991).

To account for this apparent paradox, Bregman (1990, chap. 7) proposes a two-component theory that distinguishes sensory evidence from perceptual descriptions. One component involves primitive scene analysis processes that assign links of variable strength among parts of the sensory evidence. The link strength depends both on the sensory evidence (e.g., the amount of asynchrony or mistuning for concurrent grouping, or the degree of temporal proximity and spectral dissimilarity for sequential grouping) and on competition among the cues. The links are evidence for belongingness but do not necessarily create disjunct sets of sensory information; that is, they do not provide an all-or-none partitioning. A second component then builds descriptions from the sensory evidence that *are* exhaustive partitionings for a given perceptual situation. Learned schemas can intervene in this process, making certain descriptions more likely than others, perhaps as a function of their frequency of occurrence in the environment. It is at this latter level that evidence can be interpreted as belonging to more than one event in the global description. But why should one allow for this possibility in auditory processing? The reason is that acoustic events do not occlude other events in the way that most (but not all) objects occlude the light reflected from other objects that are farther from the viewer. The acoustic signal arriving at the ears is the weighted sum of the waveforms radiating from different vibrating objects, where the weighting is a function of distance and of various transformations of the original waveform due to the properties of the environment (reflections, absorption, etc.). It is thus possible that the frequency content of one event coincides partially with that of another event. To analyze the properties of the events correctly, the auditory system must be able to take into account this property of sound, which, by analogy with vision, Bregman has termed *transparency.*

This theory presumes (a) that primitive scene analysis is performed on the sensory input prior to the operation of more complex pattern-recognition processes, (b) that the complex processes that build perceptual descriptions are packaged in schemas embodying various regularities in the sensory evidence, (c) that higher-level schemas can build from regularities detected in the descriptions built by lower-level schemas, (d) that the descriptions are constrained by criteria of consistency and noncontradiction, and (e) that when schemas (including speech schemas) make use of the information that they need from a mixture, they do not remove it from the array of information that other description-building processes can use (which may give rise to duplex-type phenomena). Although many aspects of this theory have yet to be tested empirically, some evidence is consistent with it, such as the fact that duplex perception of speech can be influenced by primitive scene-analysis processes. For example, sequential organization of the chirp component can remove it from concurrent grouping with the base stimulus, suggesting that duplex perception occurs in the presence of conflicting cues for the segregation and the integration of the isolated transition with the base (Ciocca & Bregman, 1989).

Auditory Continuity

A related problem concerns the partitioning on the basis of the surrounding context of sensory information present within overlapping sets of auditory channels. The *auditory continuity phenomenon,* also called auditory induction, is involved in perceptual restoration of missing or masked sounds in speech and music interrupted by a brief louder sound or by an intermittent sequence of brief, loud sound bursts (for reviews, see Bregman, 1990, chap. 3; Warren, 1999, chap. 6). If the waveform of a speech stream is edited such that chunks of it are removed and other chunks are left, the speech is extremely difficult to understand (Figure 10.9a). If the silent periods are replaced with noise that is loud enough to have masked the missing speech, were it present, and whose spectrum includes that of the original speech, listeners claim to hear continuous speech (Warren, Obusek, &

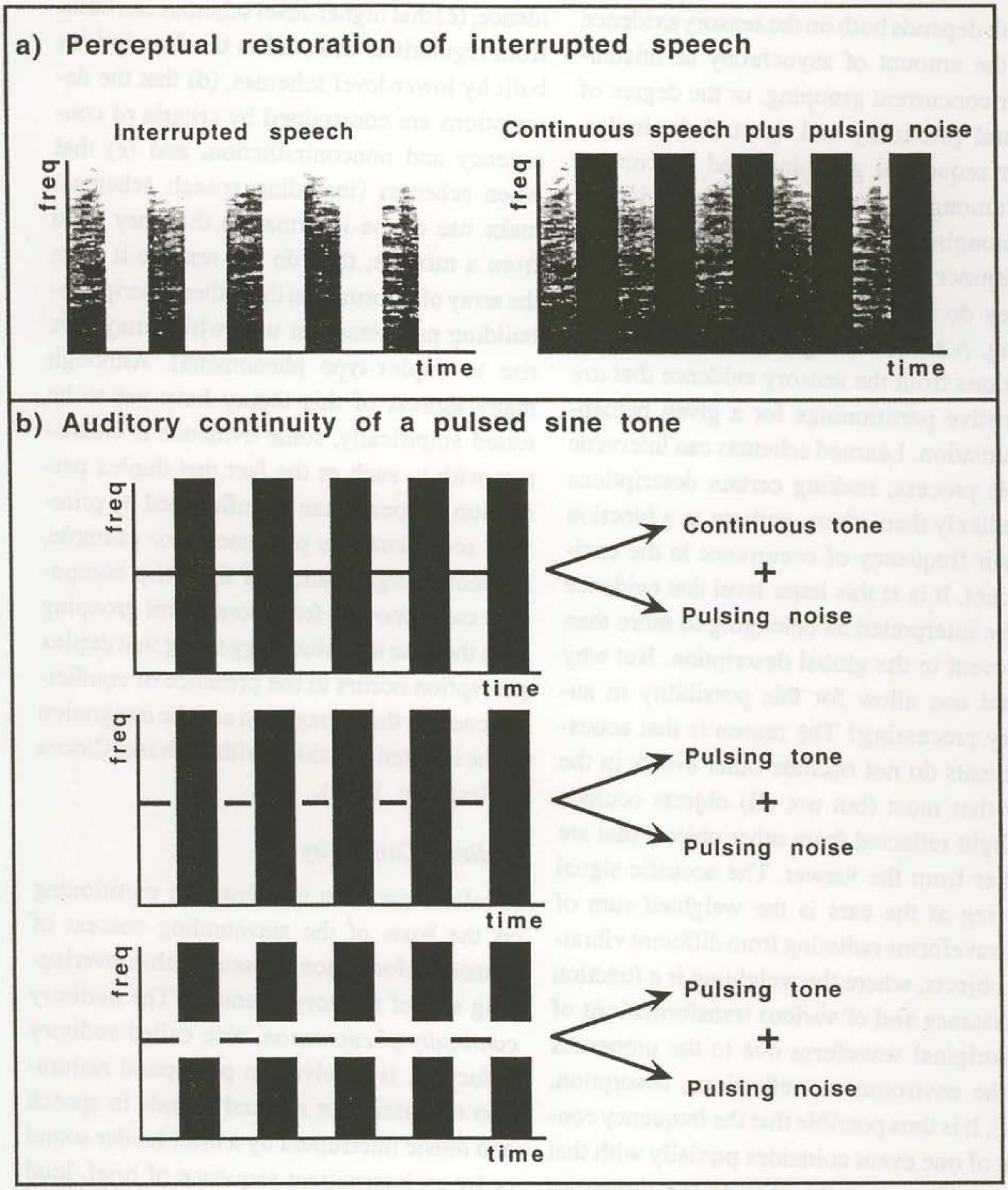


Figure 10.9 Stimuli used to test the auditory continuity phenomenon.

NOTE: a) Speech that is interrupted by silences is heard as such and is difficult to understand. If the silences are filled with a noise of bandwidth and level sufficient to have masked the absent speech signal, a pulsing noise is heard accompanied by an apparently continuous speech stream. b) Auditory continuity can be demonstrated also with a pulsed sine tone. When the silent gaps are filled with noise, a continuous tone is heard along with the pulsing noise. However, if small silent gaps of several milliseconds separate the tone and noise bursts, indicating to the auditory system that the tone actually ceased, then no continuity is obtained. Furthermore, the continuity effect does not occur if the noise does not have any energy in the frequency region of the tone.

Ackroff, 1972). Speech intelligibility can even improve if contextual information that facilitates identification of key words is present (Warren, Hainsworth, Brubaker, Bashford, & Healy, 1997).

Similar effects of continuity can be demonstrated with nonspeech stimuli, such as a sine tone interrupted by noise (Figure 10.9b) or by a higher-level sine-tone of similar frequency. An intermittent sequence superimposed on a continuous sound is heard, as if the more intense event were being partitioned into two entities, one that was the continuation of the lower-level sound preceding and following the higher-level event and another that was a sound burst. This effect works with pure tones, which indicates that it can be a completely within-channel operation. However, if any evidence exists that the lower-level sound stopped (such as short, silent gaps between the sounds), two series of intermittent sounds are heard. Furthermore, the spectrum of the interrupting sound must cover that of the interrupted sound for the phenomenon to occur; that is, the auditory system must have evidence that the interrupting sound could have masked the softer sound (Figure 10.9).

The partitioning mechanism has been conceived by Bregman (1990) in terms of an "old-plus-new" heuristic. The auditory system performs a subtraction operation on the high-level sound. A portion of the energy equivalent to that in the lower-level sound is assigned to the continuous stream, and the rest is left to form the intermittent stream. Indeed, the perceived levels of the continuous sound and intermittent sequence depend on the relative level change and are consistent with a mechanism that partitions the energy (Warren, Bashford, Healy, & Brubaker, 1994). However, the perceived levels are not consistent with a subtraction performed either in units of loudness (sones) or in terms of physical pressure or power (McAdams, Botte, & Drake, 1998). Furthermore, changes occur in the timbre of the high-level sounds in the presence of the low-level sounds compared to when these are absent (Warren et al., 1994). The relative durations of high- and low-level sounds are crucial to the phenomenon. The continuity effect is much stronger when the interrupting event is short compared to the uninterrupted portion. The perceived loudness is also a function of the relative levels of high and low portions, their relative durations, and the perceptual stream to which attention is being directed (Drake & McAdams, 1999). Once again, this continuity phenomenon demonstrates the existence of a heuristic for partitioning acoustic mixtures (if there is sufficient sensory evidence that a mixture indeed exists). It provides the listener with the ability to deal efficiently and veridically with the stimulus complexity resulting from the transparency of auditory events.

Schema-Based Organization

Much mention has been made of the possibility that auditory stream formation is affected by conscious, controlled processes, such as searching for a given source or event in the auditory scene. Bregman (1990) proposed a component that he termed *schema-based scene analysis* in which specific information is selected on the basis of attentional focus and previously acquired knowledge, resulting in the popout of previously activated events or the extraction of sought-after events. Along these lines, van Noorden's (1975) ambiguous region is an example in which what is heard depends in part on what one tries to hear. Further, in his interleaved melody recognition experiments, Dowling (1973a) observed that a verbal priming of an interleaved melody increased identification performance.

Other top-down effects in scene analysis include the role of pattern context (good continuation in Gestalt terms) and the use of previous knowledge to select target information from the scene. For example, a competition between good continuation and frequency proximity demonstrates that melodic pattern can affect the degree of streaming (Heise & Miller, 1951). Frequency proximity alone cannot explain these results.

Bey (1999; Bey & McAdams, in press) used an interleaved melody recognition paradigm to study the role of schema-based organization. In one interval an interleaved mixture of target melody and distractor sequence was presented, and in another interval an isolated comparison melody was presented. Previous presentation of the isolated melody gave consistently better performance than when the mixture sequence was presented before the comparison melody. Furthermore, if the comparison melody was transposed by 12, 13, or 14 semitones—requiring the listener to use a pitch-interval-based representation instead of an absolute-pitch representation to perform the task—performance was similar to when the isolated comparison melody was presented after the mixture. These results suggest that in this task an absolute-pitch representation constitutes the "knowledge" used to extract the melody. However, performance varied as a function of the frequency separation of the target melody and distractor sequence, so performance depended on both sensory-based organizational constraints and schema-based information selection.

TIMBRE PERCEPTION

Early work on timbre perception paved the way to the exploration of sound source perception. The word *timbre* gathers together a number of auditory attributes that until recently have been defined only by what they are not: Timbre is what distinguishes two sounds coming from the same position in space and having the same pitch, loudness, and subjective duration. Thus, an oboe and a trumpet playing the same note, for example, would be distinguished by their timbres. This definition indeed leaves everything to be defined. The perceptual qualities grouped under this term are multiple and depend on several acoustic properties (for reviews, see Hajda, Kendall, Carterette, & Harshberger, 1997; McAdams, 1993; Risset & Wessel, 1999). In this section, we examine spectral profile analysis, the perception of auditory roughness, and the multidimensional approach to timbre perception.

Spectral Profile Analysis

The sounds that a listener encounters in the environment have quite diverse spectral properties. Those produced by resonating structures of vibrating objects have more energy near the natural frequencies of vibration of the object (string, plate, air cavity, etc.) than at more distant frequencies. In a frequency spectrum in which amplitude is plotted as a function of frequency, one would see peaks in some frequency regions and dips in others. The global form of this spectrum is called the spectral envelope. The extraction of the spectral envelope by the auditory system would thus be the basis for the evaluation of constant resonance structure despite varying fundamental frequency (Plomp & Steeneken, 1971; Slawson, 1968) and may possibly contribute to source recognition. This extraction is surely strongly involved in vowel perception, the quality of which is related to the position in the spectrum of resonance regions called *formants* (see Chap. 12, this volume). Spiegel and Green (1982) presented listeners with complex sounds in which the amplitudes were equal on all components except one. The level of this component was increased to create a bump in the spectral envelope. They showed that a listener is able to discriminate these spectral envelopes despite random variations complementary fields in the hopes that the resulting juxtaposition of ideas will facilitate and enable future research to fill in the gaps.

REFERENCES

Abdala, C., & Folsom, R. C. (1995). The development of frequency resolution in human as revealed by the auditory brainstem response recorded with notched noise masking. *Journal of the Acoustical Society of America, 98*(2), 921–930.

Alain, C. A. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience, 5,* 202–212.

Alain, C. A., Arnott, S. R., & Picton T. W. (2001). Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance, 27*(5), 1072–1089.

Alain, C. A., Woods, D. L., & Ogawa, K. H. (1994). Brain indices of automatic pattern processing. *NeuroReport, 6,* 140–144.

Allan, L. (1992). The internal clock revisited. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 191–202). Dordrecht: Kluwer Academic.

Allen, P., & Wightman, F. (1995). Effects of signal and masker uncertainty on children's detection. *Journal of Speech and Hearing Research, 38*(2), 503–511.

Andrews, M. W., & Dowling, W. J. (1991). The development of perception of interleaved melodies and control of auditory attention. *Music Perception, 8*(4), 349–368.

Anstis, S., & Saida, S. (1985). Adaptation of auditory streaming to frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 257–271.

Assmann, P. F., & Summerfield, Q. (1990). Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *Journal of the Acoustical Society of America, 88,* 680–697.

Aures, W. (1985). Ein Berechnungsverfahren der Rauhigkeit [A roughness calculation method]. *Acustica, 58,* 268–281.

Bamberger, J. (1980). Cognitive structuring in the apprehension and description of simple rhythms. *Archives of Psychology, 48,* 177–199.

Baruch, C. (2001). L'audition du bébé et du jeune enfant. *Année Psychologique, 101,* 91–124.

Baruch, C., & Drake, C. (1997). Tempo discrimination in infants. *Infant Behavior and Development, 20*(4), 573–577.

Beauvois, M. W., & Meddis, R. (1997). Time decay of auditory stream biasing. *Perception and Psychophysics, 59,* 81–86.

Berg, K. M. (1993). A comparison of thresholds for 1/3-octave filtered clicks and noise burst in infants and adults. *Perception and Psychophysics, 54*(3), 365–369.

Berg, K. M., & Smith, M. C. (1983). Behavioral thresholds for tones during infancy. *Journal of Experimental Child Psychology, 35,* 409–425.

Bertrand, D. (1999). *Groupement rythmique et représentation mentale de mélodies chez l'enfant.* Liège, Belgium: Université de Liège.

Bey, C. (1999). *Reconnaissance de mélodies intercalées et formation de flux auditifs: Analyse fonctionnelle et exploration neuropsychologique [Recognition of interleaved melodies and auditory stream formation: Functional analysis and neuropsychological exploration].* Unpublished doctoral dissertation, Ecole des Hautes Etudes en Sciences Sociales (EHESS), Paris.

Bey, C., & McAdams, S. (in press). Schema-based processing in auditory scene analysis. *Perception and Psychophysics.*

Bigand, E., McAdams, S., & Forêt, S. (2000). Divided attention in music. *International Journal of Psychology, 35,* 270–278.

Block, R. A. (1990). Models of psychological time. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 1–35). Hillsdale, NJ: Erlbaum.

Boltz, M. G. (1994). Changes in internal tempo and effects on the learning and remembering of event durations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1154–1171.

Boltz, M. G. (1995). Effects of event structure on retrospective duration judgments. *Perception & Psychophysics, 57,* 1080–1096.

Boltz, M. G. (1998). Task predictability and remembered duration. *Perception and Psychophysics, 60*(5), 768–784.

Boltz, M. G., Schulkind, M., & Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Memory and Cognition, 19*(6), 593–606.

Botte, M. C., Drake, C., Brochard, R., & McAdams, S. (1997). Perceptual attenuation of nonfocused auditory stream. *Perception and Psychophysics, 59*(3), 419–425.

Bregman, A. S. (1978a). Auditory streaming: Competition among alternative organizations. *Perception and Psychophysics, 23,* 391–398.

Bregman, A. S. (1978b). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 380–387.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge: MIT Press.

Bregman, A. S. (1991). Using brief glimpses to decompose mixtures. In J. Sundberg, L. Nord, & R. Carleson (Eds.), *Music, language, speech and brain* (pp. 284–293). London: Macmillan.

Bregman, A. S. (1993). Auditory scene analysis: Hearing in complex environments. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 10–36). Oxford: Oxford University Press.

Bregman, A. S., & Ahad, P. (1995). *Demonstrations of auditory scene analysis: The perceptual organization of sound.* Montréal, Québec, Canada: McGill University. Compact disc available at http://www.psych.mcgill.ca/labs/auditory/bregmancd.html.

Bregman, A. S., Ahad, P., Kim, J., & Melnerich, L. (1994). Resetting the pitch-analysis system: 1. Effects of rise times of tones in noise backgrounds or of harmonics in a complex tone. *Perception and Psychophysics, 56,* 155–162.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89,* 244–249.

Bregman, A. S., & Dannenbring, G. L. (1973). The effect of continuity on auditory stream segregation. *Perception and Psychophysics, 13,* 308–312.

Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology, 44,* 400–413.

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology, 32,* 19–31.

Broadbent, D. E. (1958). *Perception and communication.* London: Pergamon.

Brochard, R., Drake, C., Botte, M.-C., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1742–1759.

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics, 10,* 23–36.

Buell, T. N., & Hafter, E. R. (1991). Combination of binaural information across frequency bands. *Journal of the Acoustical Society of America, 90,* 1894–1900.

Cabe, P. A., & Pittenger, J. B. (2000). Human sensitivity to acoustic information from vessel filling. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 313–324.

Carello, C., Anderson, K. A., & Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychological Science, 9,* 211–214.

Carlyon, R. P. (1991). Discriminating between coherent and incoherent frequency modulation of complex tones. *Journal of the Acoustical Society of America, 89,* 329–340.

Carlyon, R. P. (1992). The psychophysics of concurrent sound segregation. *Philosophical*

Transactions of the Royal Society, London B, 336, 347–355.

Carlyon, R. P. (1994). Detecting mistuning in the presence of synchronous and asynchronous interfering sounds. *Journal of the Acoustical Society of America, 95*, 2622–2630.

Carterette, E. C., & Kendall, R. A. (1999). Comparative music perception and cognition. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 725–791). San Diego: Academic Press.

Chaigne, A., & Doutaut, V. (1997). Numerical simulations of xylophones: I. Time-domain modeling of the vibrating bars. *Journal of the Acoustical Society of America, 101*, 539–557.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America, 25*, 975–979.

Church, R. M. (1984). Properties of the internal clock. In J. Gibbon & L. Allan (Eds.), *Timing and time perception* (Vol. 423, pp. 566–582). New York: New York Academy of Sciences.

Ciocca, V., & Bregman, A. S. (1989). The effects of auditory streaming on duplex perception. *Perception and Psychophysics, 46*, 39–48.

Ciocca, V., & Darwin, C. J. (1993). Effects of onset asynchrony on pitch perception: Adaptation or grouping? *Journal of the Acoustical Society of America, 93*, 2870–2878.

Clarke, E. F., & Krumhansl, C. (1990). Perceiving musical time. *Music Perception, 7*(3), 213–252.

Clarkson, M. G., & Clifton, R. K. (1995). Infant's pitch perception: Inharmonic tonal complexes. *Journal of the Acoustical Society of America, 98*(3), 1372–1379.

Clarkson, M. G., Martin, R. L., & Miciek, S. G. (1996). Infants perception of pitch: Number of harmonics. *Infant Behavior and Development, 19*, 191–197.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraint within the human information-processing system. *Psychological Bulletin, 104*, 163–191.

Creelman, C. D. (1962). Human discrimination of auditory duration. *Journal of the Acoustical Society of America, 34*, 582–593.

Crowder, R. G. (1993). Auditory memory. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 113–145). Oxford: Oxford University Press.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics, 5*, 365–373.

Culling, J. F., & Darwin, C. J. (1993). The role of timbre in the segregation of simultaneous voices with intersecting Fo contours. *Perception and Psychophysics, 34*, 303–309.

Culling, J. F., & Summerfield, Q. (1995). Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. *Journal of the Acoustical Society of America, 98*, 758–797.

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review, 83*, 114–140.

Daniel, P., & Weber, R. (1997). Psychoacoustical roughness: Implementation of an optimized model. *Acustica, 83*, 113–123.

Dannenbring, G. L., & Bregman, A. S. (1976). Stream segregation and the illusion of overlap. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 544–555.

Darwin, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset time. *Quarterly Journal of Experimental Psychology, 33A*, 185–208.

Darwin, C. J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of the Acoustical Society of America, 76*, 1636–1647.

Darwin, C. J. (1991). The relationship between speech perception and the perception of other sounds. In I. G. Mattingly & M. G. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 239–259). Hillsdale, NJ: Erlbaum.

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Ed.), *Hearing* (pp. 387–424). San Diego: Academic Press.

Darwin, C. J., & Ciocca, V. (1992). Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component. *Journal of the Acoustical Society of America, 91,* 3381–3390.

Darwin, C. J., Ciocca, V., & Sandell, G. R. (1994). Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic. *Journal of the Acoustical Society of America, 95,* 2631–2636.

Darwin, C. J., & Gardner, R. B. (1986). Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality. *Journal of the Acoustical Society of America, 79,* 838–845.

Darwin, C. J., & Hukin, R. W. (1999). Auditory objects of attention: The role of interaural time differences. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 617–629.

Darwin, C. J., & Sutherland, N. S. (1984). Grouping frequency components of vowels: When is a harmonic not a harmonic? *Quarterly Journal of Experimental Psychology, 36A,* 193–208.

de Cheveigné, A. (1993). Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing. *Journal of the Acoustical Society of America, 93,* 3271–3290.

de Cheveigné, A. (1999). Waveform interactions and the segregation of concurrent vowels. *Journal of the Acoustical Society of America, 106,* 2959–2972.

de Cheveigné, A., Kawahara, H., Tsuzaki, M., & Aikawa, K. (1997). Concurrent vowel identification: I. Effects of relative level and F0 difference. *Journal of the Acoustical Society of America, 101,* 2839–2847.

de Cheveigné, A., McAdams, S., Laroche, J., & Rosenberg, M. (1995). Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement. *Journal of the Acoustical Society of America, 97,* 3736–3748.

de Cheveigné, A., McAdams, S., & Marin, C. M. H. (1997). Concurrent vowel identification: II. Effects of phase, harmonicity, and task. *Journal of the Acoustical Society of America, 101,* 2848–2856.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception, 4,* 325–360.

Deliège, I. (1990). Mechanisms of cue extraction in musical grouping: A study of Sequenza VI for Viola Solo by L. Berio. *Psychology of Music, 18,* 18–45.

Deliège, I. (1993). Mechanisms of cue extraction in memory for musical time. *Contemporary Music Review, 9*(1&2), 191–205.

Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behavior and Development, 5,* 261–276.

Demany, L., & Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *Journal of the Acoustical Society of America, 1,* 57–66.

Demany, L., & Armand, F. (1985). A propos de la perception de la hauteur et du chroma des sons purs. *Bulletin d'Audiophonologie, 1–2,* 123–132.

Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature, 266,* 718–719.

Desain, P. (1992). A (de)composable theory of rhythm perception. *Music Perception, 9*(4), 439–454.

Deutsch, D. (1975). Two-channel listening to musical scales. *Journal of the Acoustical Society of America, 57,* 1156–1160.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics, 28,* 381–389.

Deutsch, D. (1995). *Musical illusions and paradoxes*. La Jolla, CA: Philomel Records. Compact disc available at http://www.philomel.com.

Deutsch, D. (Ed.). (1999). *The psychology of music* (2nd ed.). San Diego: Academic Press.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88*(6), 503–522.

Divenyi, P. L., & Danner, W. F. (1977). Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Perception & Psychophysics, 21*(2), 125–142.

Dowling, W. J. (1973a). The perception of interleaved melodies. *Cognitive Psychology, 5,* 322–327.

Dowling, W. J. (1973b). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics, 14,* 37–40.

Dowling, W. J. (1984). Development of musical schemata in children's spontaneous singing. In W. R. Crozier & A. J. Chapman (Eds.), *Cognitive processes in the perception of art* (pp. 145–163). Amsterdam: North-Holland.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition.* New York: Academic Press.

Drake, C. (1993a). Perceptual and performed accents in musical sequences. *Bulletin of the Psychonomic Society, 31*(2), 107–110.

Drake, C. (1993b). Reproduction of musical rhythms by children, adult musicians and adult nonmusicians. *Perception & Psychophysics, 53*(1), 25–33.

Drake, C. (1997). Motor and perceptually preferred synchronisation by children and adults: Binary and ternary ratios. *Polish Journal of Developmental Psychology, 3*(1), 41–59.

Drake, C. (1998). Psychological processes involved in the temporal organization of complex auditory sequences: universal and acquired processes. *Music Perception, 16*(1), 11–26.

Drake, C., & Bertrand, D. (2001). The quest for universals in temporal processing in music. *Annals of the New York Academy of Sciences, 930,* 17–27.

Drake, C., & Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics, 54,* 277–286.

Drake, C., Dowling, W. J. & Palmer, C. (1991). Accent structures in the reproduction of simple tunes by children and adult pianists. *Music Perception, 8,* 315–334.

Drake, C., & Gérard, C. (1989). A psychological pulse train: How young children use this cognitive framework to structure simple rhythms. *Psychological Research, 51,* 16–22.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition, 77,* 251–288.

Drake, C., & McAdams, S. (1999). The continuity illusion: Role of temporal sequence structure. *Journal of the Acoustical Society of America, 106,* 3529–3538.

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception, 10*(3), 343–378.

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition, 74*(1), 1–32.

Drake, C., Penel, A., & Bigand, E. (2000). Tapping in time with mechanically and expressively performed music. *Music Perception. 18,* 1–23.

Duifhuis, H., Willems, L. F., & Sluyter, R. J. (1982). Measurement of pitch in speech: An implementation of Goldsteins's theory of pitch perception. *Journal of the Acoustical Society of America, 83,* 687–695.

Ehresman, D., & Wessel, D. L. (1978). *Perception of timbral analogies,* IRCAM Report no. 13. Paris: IRCAM.

Essens, P. J. (1995). Structuring temporal sequences: Comparison of models and factors of complexity. *Perception & Psychophysics, 57*(4), 519–532.

Essens, P. J., & Povel, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics, 37*(1), 1–7.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* New York: Wiley.

Ferland, M. B., & Mendelson, M. J. (1989). Infant categorization of melodic contour. *Infant Behavior and Development, 12,* 341–355.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2000). Complex tone processing in primary auditory cortex of the awake monkey: I. Neural ensemble correlates of roughness.

Journal of the Acoustical Society of America, 108, 235–246.

Fitzgibbons, P. J., & Gordon Salant, S. (1998). Auditory temporal order perception in younger and older adults. *Journal of Speech, Language, and Hearing Research, 41*(5), 1052–1060.

Fitzgibbons, P. J., Pollatsek, A., & Thomas, I. B. (1974). Detection of temporal gaps within and between perceptual tonal groups. *Perception & Psychophysics, 16,* 522–528.

Fraisse, P. (1956). *Les structures rythmiques*. Louvain, Belgium: Publications Universitaires de Louvain.

Fraisse, P. (1963). *The psychology of time*. New York: Harper & Row.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York: Academic Press.

Fraisse, P., Pichot, P., & Clairouin, G. (1969). Les aptitudes rythmiques: Etude comparée des oligophrènes et des enfants normaux. *Journal de Psychologie Normale et Pathologique, 42,* 309–330.

Gardner, R. B., Gaskill, S. A., & Darwin, C. J. (1989). Perceptual grouping of formants with static and dynamic differences in fundamental frequency. *Journal of the Acoustical Society of America, 85,* 1329–1337.

Getty, D. J. (1975). Discrimination of short temporal intervals: A comparison of two models. *Perception & Psychophysics, 18,* 1–8.

Giard, M.-H., Collet, L., Bouchet, P., & Pernier, J. (1993). Modulation of human cochlear activity during auditory selective attention. *Brain Research, 633,* 353–356.

Giard, M.-H., Fort, A., Mouchetant-Rostaing, Y., & Pernier, J. (2000). Neurophysiological mechanisms of auditory selective attention in humans. *Frontiers in Bioscience, 5,* 84–94.

Glucksberg, S., & Cowen, G. N. (1970). Memory for non-attended auditory material. *Cognitive Psychology, 1,* 149–156.

Green, D. M. (1988). Auditory profile analysis: Some experiments on spectral shape discrimination. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function: Neurobiological bases of hearing* (pp. 609–622). New York: Wiley.

Green, D. M., & Mason, C. R. (1985). Auditory profile analysis: Frequency, phase, and Weber's law. *Journal of the Acoustical Society of America, 77,* 1155–1161.

Green, D. M., Mason, C. R., & Kidd, G. (1984). Profile analysis: Critical bands and duration. *Journal of the Acoustical Society of America, 74,* 1163–1167.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America, 61,* 1270–1277.

Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America, 63,* 1493–1500.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., & Collet, L. (2000). Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. *Journal of the Acoustical Society of America, 108,* 263–271.

Guski, R. (2000). Studies in auditive kinetics. In A. Schick, M. Meis, & C. Reckhardt (Eds.), *Contributions to psychological acoustics: Results of the 8th Oldenburg Symposium on Psychological Acoustics* (pp. 383–401). Oldenburg: Bis.

Hafter, E. R., & Buell, T. N. (1985). The importance of transients for maintaining the separation of signals in space. In M. Posner & O. Marin (Eds.), *Attention and performance XI* (pp. 337–354). Hillsdale, NJ: Erlbaum.

Hafter, E. R., Buell, T. N., & Richards, V. M. (1988). Onset-coding in lateralization: Its form, site, and function. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function: Neurobiological bases of hearing* (pp. 647–676). New York: Wiley.

Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harshberger, M. L. (1997). Methodological issues in timbre research. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 253–306). Hove, England: Psychology Press.

Hall, J. W., Grose, J. H., & Mendoza, L. (1995). Across-channel processes in masking. In B. C. J.

Moore (Ed.), *Hearing* (pp. 243–266). San Diego: Academic Press.

Halpern, A. R., & Darwin, C. J. (1982). Duration discrimination in a series of rhythmic events. *Perception & Psychophysics, 31*(1), 86–89.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

Hartmann, W. M. (1988). Pitch perception and the segregation and integration of auditory entities. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function* (pp. 623–645). New York: Wiley.

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception, 9,* 155–184.

Hartmann, W. M., McAdams, S., & Smith, B. K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *Journal of the Acoustical Society of America, 88,* 1712–1724.

Heise, G. A., & Miller, G. A. (1951). An experimental study of auditory patterns. *American Journal of Psychology, 64,* 68–77.

Helmholtz, H. L. F. von. (1885). *On the sensations of tone as a physiological basis for the theory of music*. New York, from 1877 trans by A. J. Ellis of 4th German ed.: republ.1954, New York: Dover.

Hill, N. I., & Darwin, C. J. (1996). Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning. *Journal of the Acoustical Society of America, 100,* 2352–2364.

Houtsma, A. J. M., Rossing, T. D., & Wagenaars, W. M. (1987). *Auditory demonstrations on compact disc*. Melville, NY: Acoustical Society of America. Compact disc available at http://asa.aip.org/discs.html.

Hukin, R. W., & Darwin, C. J. (1995a). Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification. *Perception and Psychophysics, 57,* 191–196.

Hukin, R. W., & Darwin, C. J. (1995b). Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel. *Journal of the Acoustical Society of America, 98,* 1380–1387.

Huron, D., & Fantini, D. (1989). The avoidance of inner-voice entries: Perceptual evidence and musical practice. *Music Perception, 7,* 43–48.

Iverson, P. (1995). Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 751–763.

Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America, 94,* 2595–2603.

Jensen, K. J., & Neff, D. L. (1993). Development of basic auditory discrimination in preschool children. *Psychological Science, 4,* 104–107.

Jones, M. R. (1976). Time, our last dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83*(5), 323–355.

Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception & Psychophysics, 41*(6), 631–634.

Jones, M. R. (1990). Learning and development of expectancies: An interactionist approach. *Psychomusicology, 2*(9), 193–228.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96,* 459–491.

Jones, M. R., & Yee, W. (1993). Attending to auditory events: The role of temporal organization. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 69–112). Oxford: Oxford University Press.

Jones, M. R., & Yee, W. (1997). Sensitivity to time change: the role of context and skill. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 693–709.

Kameoka, A., & Kuriyagawa, M. (1969). Consonance theory: Part II. Consonance of complex tones and its calculation method. *Journal of the Acoustical Society of America, 45,* 1460–1469.

Killeen, P. R., & Weisse, N. A. (1987). Optimal timing and the Weber function. *Psychological Review, 94*(4), 445–468.

Koehler, W. (1929). *Gestalt psychology*. New York: Liveright.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes: II. Analyses acoustiques et quantification psychophysique [Characterization of the timbre of complex sounds: II. Acoustic analyses and psychophysical quantification]. *Journal de Physique, 4*(C5), 625–628.

Kristofferson, A. B. (1980). A quantal step function in duration discrimination. *Perception & Psychophysics, 27,* 300–306.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzén & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43–53). Amsterdam: Excerpta Medica.

Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1,* 70–73.

Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Jaervinen, T., & Eerola, T. (1999). Melodic expectation in Finnish spiritual folk hymns: Convergence of statistical, behavioral, and computational approaches. *Music Perception, 17*(2), 151–195.

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Jaervinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition, 76*(1), 13–58.

Kubovy, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 55–98). Hillsdale, NJ: Erlbaum.

Kubovy, M., Cutting, J. E., & McGuire, R. M. (1974). Hearing with the third ear: Dichotic perception of a melody without monaural familiarity cues. *Science, 186,* 272–274.

Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 279–294.

Lakatos, S. (2000). A common perceptual space for harmonic and percussive timbres. *Perception and Psychophysics, 62,* 1426–1439.

Lakatos, S., McAdams, S., & Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric form. *Perception & Psychophysics, 59,* 1180–1190.

Lambourg, C., Chaigne, A., & Matignon, D. (2001). Time-domain simulation of damped impacted plates: II. Numerical model and results. *Journal of the Acoustical Society of America, 109,* 1433–1447.

Large, E. W., Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119–159.

Lea, A. (1992). *Auditory models of vowel perception.* Unpublished doctoral dissertation, University of Nottingham, Nottingham, England.

Lee, C. S. (1991). The perception of metrical structure: Experimental evidence and a model. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure* (pp. 59–127). London: Academic Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge: MIT Press.

Little, V. M., Thomas, D. G., & Letterman, M. R. (1999). Single-trial analysis of developmental trends in infant auditory event-related potentials. *Developmental Neuropsychology, 16*(3), 455–478.

Longuet-Higgins, H. C., & Lee, C. S. (1982). The perception of musical rhythms. *Perception, 11,* 115–128.

Longuet-Higgins, H. C., & Lee, C. S. (1984). The rhythmic interpretation of monophonic music. *Music Perception, 1,* 424–440.

Luce, G. G. (1972). *Body time.* London: Temple Smith.

McAdams, S. (1989). Segregation of concurrent sounds: I. Effects of frequency modulation coherence. *Journal of the Acoustical Society of America, 86,* 2148–2159.

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146–198). Oxford: Oxford University Press.

McAdams, S., & Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *Journal of the Acoustical Society of America, 102,* 2945–2953.

McAdams, S., & Bigand, E. (Eds.). (1993). *Thinking in sound: The cognitive psychology of human audition*. Oxford: Oxford University Press.

McAdams, S., Botte, M.-C., & Drake, C. (1998). Auditory continuity and loudness computation. *Journal of the Acoustical Society of America, 103,* 1580–1591.

McAdams, S., & Bregman, A. S. (1979). Hearing musical streams. *Computer Music Journal, 3*(4), 26–43.

McAdams, S., & Cunibile, J. C. (1992). Perception of timbral analogies. *Philosophical Transactions of the Royal Society, London, Series B, 336,* 383–389.

McAdams, S., & Marin, C. M. H. (1990). *Auditory processing of frequency modulation coherence.* Paper presented at the Fechner Day '90, 6th Annual Meeting of the International Society for Psychophysics, Würzburg, Germany.

McAdams, S., & Winsberg, S. (2000). Psychophysical quantification of individual differences in timbre perception. In A. Schick, M. Meis, & C. Reckhardt (Eds.), *Contributions to psychological acoustics: Results of the 8th Oldenburg Symposium on Psychological Acoustics* (pp. 165–182). Oldenburg, Germany: Bis.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research, 58,* 177–192.

McFadden, D., & Wright, B. A. (1987). Comodulation masking release in a forward-masking paradigm. *Journal of the Acoustical Society of America, 82,* 1615–1630.

Meric, C. & Collet, L. (1994). Attention and otoacoustic emissions: A review. *Neuroscience and Behavioral Review, 18,* 215–222.

Michon, J. A. (1975). Time experience and memory processes. In J. T. Fraser & N. Lawrence (Eds.), *The study of time* (pp. 2–22). Berlin: Springer.

Michon, J. A. (1978). The making of the present: A tutorial review. In J. Requin (Ed.), *Attention and performance VII* (pp. 89–111). Hillsdale, NJ: Erlbaum.

Miller, G. A. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Monahan, C. B., Kendall, R. A., & Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception & Psychophysics, 41*(6), 576–600.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America, 77,* 1853–1860.

Moore, B. C. J., Peters, R. W., & Glasberg, B. R. (1985). Thresholds for the detection of inharmonicity in complex tones. *Journal of the Acoustical Society of America, 77,* 1861–1867.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11,* 56–60.

Morrongiello, B. A., & Trehub, S. E. (1987). Age-related changes in auditory temporal perception. *Journal of Experimental Child Psychology, 44,* 413–426.

Näätänen, R. (1992). *Attention and brain function.* Hillsdale, NJ: Erlbaum.

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear & Hearing, 16,* 6–18.

Nozza, R. J., & Wilson, W. R. (1984). Masked and unmasked pure-tone thresholds of infants and adults: Development of auditory frequency selectivity and sensitivity. *Journal of Speech and Hearing Research, 27,* 613–622.

Olsho, L. W. (1985). Infant auditory perception: Tonal masking. *Infant Behavior and Development, 8,* 371–384.

Olsho, L. W., Koch, E. G., Carter, E. A., Halpin, C. F., & Spetner, N. B. (1988). Pure-tone sensitivity of human infants. *Journal of the Acoustical Society of America, 84*(4), 1316–1324.

Ornstein, R. E. (1969). *On the experience of time.* Harmondsworth, England: Penguin.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 331–346.

Palmer, C., & Drake, C. (1997). Monitoring and planning capacities in the acquisition of music performance skills. *Canadian Journal of Experimental Psychology, 51*(4), 369–384.

Penel, A., & Drake, C. (1998). Sources of timing variations in music performance: A psychological segmentation model. *Psychological Research, 61,* 12–32.

Peretz, I., & Morais, J. (1989). Music and modularity. *Contemporary Music Review, 4,* 279–294.

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 397–414). Leiden: Sijthoff.

Plomp, R. (1976). *Aspects of tone sensation: A psychophysical study*. London: Academic Press.

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America, 38,* 548–560.

Plomp, R., & Steeneken, J. M. (1971). Pitch versus timbre. *Proceedings of the 7th International Congress of Acoustics, Budapest, 3,* 377–380.

Pollard-Gott, L. (1983). Emergence of thematic concepts in repeated listening to music. *Cognitive Psychology, 15,* 66–94.

Povel, D. J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 3–18.

Povel, D. J. (1985). Perception of temporal patterns. *Music Perception,* 2(4), 411–440.

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2,* 411–440.

Pressnitzer, D., & McAdams, S. (1999a). An effect of the coherence between envelopes across frequency regions on the perception of roughness. In T. Dau, V. Hohmann, & B. Kollmeier (Eds.), *Psychophysics, physiology and models of hearing* (pp. 105–108). London: World Scientific.

Pressnitzer, D., & McAdams, S. (1999b). Two phase effects in roughness perception. *Journal of the Acoustical Society of America, 105,* 2773–2782.

Pressnitzer, D., McAdams, S., Winsberg, S., & Fineberg, J. (2000). Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception and Psychophysics, 62,* 66–80.

Preusser, D. (1972). The effect of structure and rate on the recognition and description of auditory temporal patterns. *Perception & Psychophysics, 11*(3), 233–240.

Rand, T. C. (1974). Dichotic release from masking for speech. *Journal of the Acoustical Society of America, 55,* 678–680 (Letter to the editor).

Rasch, R. (1978). The perception of simultaneous notes as in polyphonic music. *Acustica, 40,* 21–33.

Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica, 43*(2), 121–131.

Repp, B. H. (1987). The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America, 81,* 1100–1109.

Repp, B. H. (1992). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition, 44,* 241–281.

Risset, J. C., & Wessel, D. L. (1999). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 113–168). San Diego: Academic Press.

Roberts, B., & Bailey, P. J. (1993). Spectral pattern and the perceptual fusion of harmonics: I. The role of temporal factors. *Journal of the Acoustical Society of America, 94,* 3153–3164.

Roberts, B., & Bailey, P. J. (1996). Regularity of spectral pattern and its effects on the perceptual fusion of harmonics. *Perception and Psychophysics, 58,* 289–299.

Roberts, B., & Bregman, A. S. (1991). Effects of the pattern of spectral spacing on the perceptual fusion of harmonics. *Journal of the Acoustical Society of America, 90,* 3050–3060.

Rogers, W. L., & Bregman, A. S. (1993a). An experimental evaluation of three theories of auditory stream segregation. *Perception and Psychophysics, 53,* 179–189.

Rogers, W. L., & Bregman, A. S. (1993b). An experimental study of three theories of auditory stream segregation. *Perception & Psychophysics, 53,* 179–189.

Rogers, W. L., & Bregman, A. S. (1998). Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness. *Perception and Psychophysics, 60,* 1216–1227.

Roussarie, V. (1999). *Analyse perceptive des structures vibrantes [Perceptual analysis of vibrating structures].* Unpublished doctoral dissertation, Université du Maine, Le Mans, France.

Roussarie, V., McAdams, S., & Chaigne, A. (1998). Perceptual analysis of vibrating bars synthesized with a physical model. *Proceedings of the 16th International Congress on Acoustics, Seattle,* 2227–2228.

Saldanha, E. L., & Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America, 36,* 2021–2126.

Scharf, B. (1998). Auditory attention: The psychoacoustical approach. In H. Pashler (Ed.), *Attention* (pp. 75–117). Hove, England: Psychology Press.

Scharf, B., Quigley, S., Aoki, C., & Peachey, N. (1987). Focused auditory attention and frequency selectivity. *Perception and Psychophysics, 42*(3), 215–223.

Scheffers, M. T. M. (1983). *Sifting vowels: Auditory pitch analysis and sound integration.* Unpublished doctoral dissertation, University of Groningen, Groningen, Netherlands.

Schneider, B. A., & Trehub, S. E. (1985). Behavioral assessment of basic auditory abilities. In S. E. Trehub & B. A. Schneider (Eds.), *Auditory development in infancy* (pp. 104–114). New York: Plenum.

Schneider, B. A., Trehub, S. E., & Bull, D. (1980). High frequency sensitivity in infants. *Science, 207,* 1003–1004.

Schneider, B. A., Trehub, S. E., Morrongiello, B. A., & Thorpe, L. A. (1986). Auditory sensitivity in preschool children. *Journal of the Acoustical Society of America, 79*(2), 447–452.

Schulze, H. H. (1978). The detectability of local and global displacements in regular rhythmic patterns. *Psychological Research, 40*(2), 173–181.

Schulze, H. H. (1989). The perception of temporal deviations in isochronic patterns. *Perception & Psychophysics, 45,* 291–296.

Singh, P. G. (1987). Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre? *Journal of the Acoustical Society of America, 82,* 886–899.

Singh, P. G., & Bregman, A. S. (1997). The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *Journal of the Acoustical Society of America, 120,* 1943–1952.

Sinnot, J. M., & Aslin, R. N. (1985). Frequency and intensity discrimination in human infants and adults. *Journal of the Acoustical Society of America, 78,* 1986–1992.

Sinnot, J. M., Pisoni, D. B., & Aslin, R. M. (1983). A comparison of pure tone auditory thresholds in human infants and adults. *Infant Behavior and Development, 6,* 3–17.

Slawson, A. W. (1968). Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *Journal of the Acoustical Society of America, 43,* 97–101.

Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music*. Oxford: Oxford University Press.

Sokolov, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual Review of Physiology, 25,* 545–580.

Sorkin, R. D., Boggs, G. J., & Brady, S. L. (1982). Discrimination of temporal jitter in patterned sequences of tones. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 46–57.

Spiegel, M. F., & Green, D. M. (1982). Signal and masker uncertainty with noise maskers of varying duration, bandwidth, and center frequency. *Journal of the Acoustical Society of America, 71,* 1204–1211.

Stoffer, T. H. (1985). Representation of phrase structure in the perception of music. *Music Perception, 3,* 191–220.

Strong, W., & Clark, M. (1967a). Perturbations of synthetic orchestral wind-instrument tones. *Journal of the Acoustical Society of America, 41,* 277–285.

Strong, W., & Clark, M. (1967b). Synthesis of wind–instrument tones. *Journal of the Acoustical Society of America, 41,* 39–52.

Summerfield, Q., & Culling, J. F. (1992). Auditory segregation of competing voices: Absence of effects of FM or AM coherence. *Philosophical Transactions of the Royal Society, London, series B, 336,* 357–366.

Susini, P., McAdams, S., & Winsberg, S. (1999). A multidimensional technique for sound quality assessment. *Acustica/Acta Acustica, 85,* 650–656.

Sussman, E., Ritter, W., & Vaughan, J. H. G. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research, 789,* 130–138.

Sussman, E., Ritter, W., & Vaughan, J. H. G. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology, 36*(1), 22–34.

Teas, D. C., Klein, A. J., & Kramer, S. J. (1982). An analysis of auditory brainstem responses in infants. *Hearing Research, 7,* 19–54. *Attention* . New York: Academic

ten Hoopen, G., Boelaarts, L., Gruisen, A., Apon, I., Donders, K., Mul, N., & Akerboon, S. (1994). The detection of anisochrony in monaural and interaural sound sequences. *Perception & Psychophysics, 56*(1), 110–120.

Terhardt, E. (1974). On the perception of periodic sound fluctuations (roughness). *Acustica, 30,* 201–213.

Thorpe, L. A., & Trehub, S. E. (1989). Duration illusion and auditory grouping in infancy. *Developmental Psychology, 25,* 122–127.

Trainor, L. J. (1997). Effect of frequency ratio on infants' and adults' discrimination of simultaneous intervals. *Journal of Experimental Psychology: Human Perception and Performance, 23*(5), 1427–1438.

Trainor, L. J., & Adams, B. (2000). Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns. *Perception and Psychophysics, 62*(2), 333–340.

Trainor, L. J., & Trehub, S. E. (1989). Aging and auditory temporal sequencing: Ordering the elements of repeating tone patterns. *Perception and Psychophysics, 45*(5), 417–426.

Trehub, S. E., Bull, D., & Thorpe, L. (1984). Infant's perception of melodies: The role of melodic contour. *Child Development, 55,* 821–830.

Trehub, S. E., Endman, M. W., & Thorpe, L. A. (1990). Infant's perception of timbre: Classification of complex tones by spectral structure. *Journal of Experimental Child Psychology, 49,* 300.

Trehub, S. E., Schneider, B. A., & Endman, M. (1980). Developmental changes in infant's sensitivity to octave-band noise. *Journal of Experimental Child Psychology, 29,* 283–293.

Trehub, S. E., Schneider, B. A., Morrongiello, B. A., & Thorpe, L. A. (1988). Auditory sensitivity in school-age children. *Journal of Experimental Child Psychology, 46*(2), 273–285.

Trehub, S. E., Thorpe, L., & Morrongiello, B. A. (1987). Organizational processes in infant's perception of auditory patterns. *Child Development, 58,* 741–749.

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock." *Psychological Monographs, 77*(13, Whole No. 576).

Treisman, M., & Rostran, A. B. (1972). Brief auditory storage: A modification of Sperling's paradigm applied to audition. *Acta Psychologica, 36,* 161–170.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Eindhoven, Netherlands: Eindhoven University of Technology.

van Noorden, L. P. A. S. (1977). Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. *Journal of the Acoustical Society of America, 61,* 1041–1045.

Vliegen, J., Moore, B. C. J., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *Journal of the Acoustical Society of America, 106,* 938–945.

Vliegen, J., & Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *Journal of the Acoustical Society of America, 105,* 339–346.

Vos, J., & Rasch, R. (1981). The perceptual onset of musical tones. *Perception & Psychophysics, 29*(4), 323–335.

Vos, P. G. (1977). Temporal duration factors in the perception of auditory rhythmic patterns. *Scientific Aesthetics, 1,* 183–199.

Vos, P. G., Mates, J., & van Kruysbergen, N. W. (1995). The perceptual centre of a stimulus as the cue for synchronization to a metronome: Evidence from asynchronies. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48A*(4), 1024–1040.

Vos, P. G., van Assen, M., & Franek, M. (1997). Perceived tempo change is dependent on base tempo and direction of change: Evidence for a generalized version of Schulze's (1978) internal beat model. *Psychological Research, 59*(4), 240–247.

Warren, R. M. (1999). *Auditory perception: A new analysis and synthesis.* Cambridge: Cambridge University Press.

Warren, R. M., Bashford, J. A., Healy, E. W., & Brubaker, B. S. (1994). Auditory induction: Reciprocal changes in alternating sounds. *Perception and Psychophysics, 55,* 313–322.

Warren, R. M., Hainsworth, K. R., Brubaker, B. S., Bashford, J. A., & Healy, E. W. (1997). Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps. *Perception and Psychophysics, 59,* 275–283.

Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science, 176,* 1149–1151.

Warren, W. H., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 704–712.

Werner, L. A., Folsom, R. C., & Mancl, L. R. (1993). The relationship between auditory brainstem response and behavioral thresholds in normal hearing infants and adults. *Hearing Research, 68,* 131–141.

Werner, L. A., & Marean, G. C. (1991). Method for estimating infant thresholds. *Journal of the Acoustical Society of America, 90,* 1867–1875.

Werner, L. A., & Rubel, E. W. (Eds.). (1992). *Developmental psychoacoustics.* Washington, D.C.: American Psychological Association.

Wertheimer, M. (1925). *Der Gestalttheorie.* Erlangen: Weltkreis-Verlag.

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal, 3*(2), 45–52.

Winsberg, S., & Carroll, J. D. (1988). A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. *Psychometrika, 53,* 217–229.

Winsberg, S., & De Soete, G. (1993). A latent class approach to fitting the weighted euclidean model: CLASCAL. *Psychometrika, 58,* 315–330.

Winsberg, S., & De Soete, G. (1997). Multidimensional scaling with constrained dimensions: CONSCAL. *British Journal of Mathematical and Statistical Psychology, 50,* 55–72.

Wood, N. L., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 255–260.

Woods, D. L., Alho, K., & Algazi, A. (1994). Stages of auditory feature conjunction: An event-related brain potential study. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 81–94.

Zera, J., & Green, D. M. (1993). Detecting temporal onset and offset asynchrony in multicomponent complexes. *Journal of the Acoustical Society of America, 93,* 1038–1052.