

# Controlling the Perceptual Organization of Sound: Guidelines Derived from Principles of Auditory Scene Analysis (ASA)

Albert S. Bregman and Wieslaw Woszczyk

## 1 Introduction

In this article, we present a theory of the perceptual organization of sound and show how it applies to a variety of practical problems. Specifically, we explore many subtle factors that contribute to making a sound be perceived as being distinct from other sounds or as a component of another sound. We will demonstrate how this fundamental human ability is crucial to understanding our environment.

A list of guidelines is presented for designing sounds to meet specific goals. Audio examples that illustrate the sometimes surprising principles are provided on the CD-ROM accompanying this book, and an extensive annotated reference section is provided for further inquiry.

The many applications explored include musical orchestration, sonification of data, optimization of voice recognition and hearing aid design. Application of ASA principals to sound recording is presented in “Creating Mixtures: The Application of Auditory Scene Analysis (ASA) to Audio Recording,” (page ??).

## 2 What is Auditory Scene Analysis (ASA)?

In our everyday lives, there are usually many sounds occurring simultaneously at any given time. The waveform that reaches each of our two ears is actually the sum of the waves that have been created by all the individual sound-producing events. This means that the listener's eardrum undergoes a complex motion in which the contributions of the individual sound-producing events are not distinguished from one another. Yet the goal of perception in natural environments is to build representations of the individual sounds, and through them, representations of the separate events—such as impacts, dragging, tearing, blowing, ringing, and oscillation—that have created them. Such events, in turn, may come from natural sound sources, machines, human and animal movements and vocalizations, or musical instruments, to name a few.

The goal of perception is to separately represent the sonic properties of each individual event, so that we can recognize it. Therefore, the auditory system faces the problem of somehow decomposing the pressure wave that reaches our ears in order to build up a number of separate representations. Auditory Scene Analysis" (ASA) is the process by which the auditory system groups together, over frequency and over time, all the spectral components that belong to each individual sound source so that it can be recognized independently of the properties of co-occurring sounds [2].

Imagine the recognition errors that might take place if ASA were carried out incorrectly. Listeners might hear syllables spoken by different talkers as if they were parts of a single word, and the individual syllables of a single spoken word might be allocated to different talkers.

Furthermore, we might not be able to recognize a person's voice or a musical instrument by its timbre. The perception of timbre is partly the result of the relative strength of the different spectral components of a sound. However, different backgrounds have different sets of frequency components of their own. If these background components were merged with the components of the voice (or of the instrument) in the computation of its timbre, they would change it. So, without ASA, the timbre of a voice or instrument would be perceived differently with each different background.

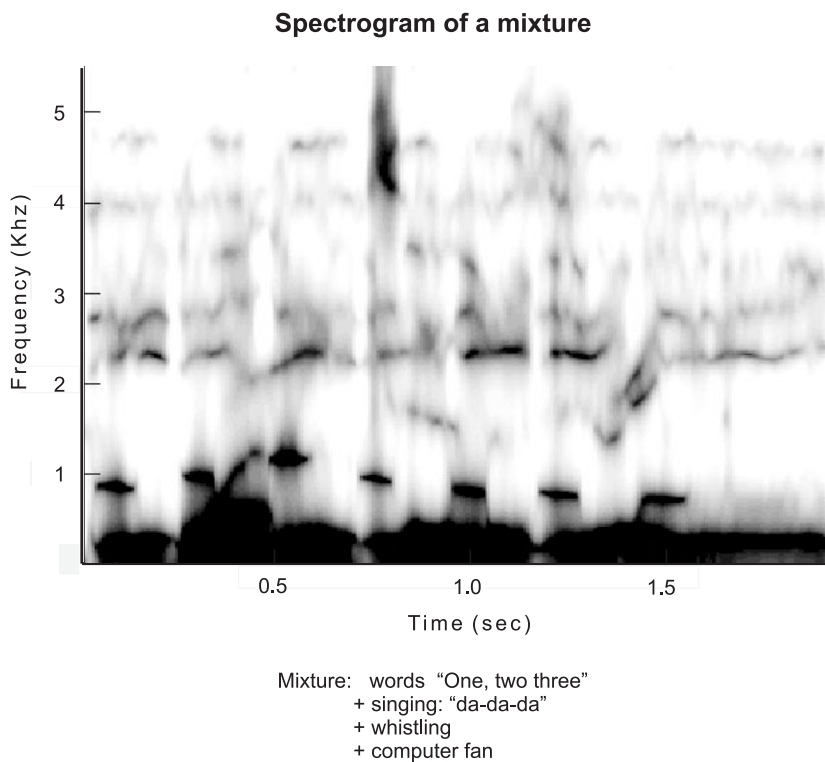
While success in everyday listening requires that the distinct environmental sounds be heard as such, the same thing is not always true in music. Music is *auditory fiction* in which the sounds of voices or instruments are combined to produce sounds that never appear in nature. The goal in music is often to temporarily lose the timbre and continuity of

individual sound sources in favor of what can be called “ensemble blend qualities,” global properties that are true of the ensemble, not of the individual sound generators. If each guitar string is viewed as a separate sound generator, then a *chord* possesses an ensemble quality. Different chords have their own holistic properties that are not the same as the qualities of the individual notes. Other ensembles are formed from separate instruments played together, or a group of voices singing at the same time (a choral effect). An “*ensemble timbre*” can emerge from a set of notes played together on different instruments, and this timbre may be different from the individual timbres of the instruments. When a blending of notes to create chords or ensemble timbres is wanted, knowledge of the principles of ASA can be used to achieve it. Because these principles don’t function in an all-or-none way, it is possible to achieve degrees of blending, from totally isolated to totally blended.

One might think that it would be simple to decompose the summed pressure wave by simply representing it in the frequency domain (as a set of frequency components). A mixture of natural sounds is shown in Figure 1 as a spectrogram, which shows time on the  $x$ -axis and frequency on the  $y$ -axis, the darkness at any point  $(x_i, y_i)$  representing the amount of energy at that time and frequency. Identifying an individual source is made very difficult by the fact that its spectral components don’t necessarily occupy a distinct part of the spectrum, but can be spread over a wide spectral range. So the components coming from separate environmental events can be interleaved in frequency. For the same reason, the total energy within each of the narrow frequency bands, whose energy is plotted in the spectrogram, can be the sum of the energy from two or more sounds.

## 2.1 Two Dimensions of Perceptual Organization: Simultaneous and Sequential

In our everyday life, the role of ASA is to group components that are either simultaneous or successive in time so that the brain’s pattern recognition processes can more easily recognize each of them. In the spectrogram of Figure 1, deciding how to group the components that are present at the same instant so that they “fuse” into a more global sound is called *simultaneous grouping*. For example, when two people are talking at once, grouping the component present at a given instant into two perceptual representations, each representing a single voice, with its own pitch and timbre, is simultaneous grouping. On the other hand, the grouping of the parts of the same voice over time is called *sequential grouping*. Of course,



**Figure 1.** Spectrogram of a mixture of sounds.

natural sounds are not pure cases of one or the other. The components in most of what we hear are partially, but not completely, overlapped in time. However, many of the laboratory studies of grouping have studied simultaneous and sequential grouping separately.

Fusion is the default. The first general observation about the segregation of sounds or their components is that the default state seems to be integration. If you squint at the spectrogram of Figure 1 until the details disappear, you will see what we mean. The image seems to become an undifferentiated whole. Only when the details of the signal give the auditory system some reason for segregating parts of it, will this happen. So any transformation of the signal that blurs the properties will favor integration of the signal.

## 3 Simultaneous Grouping (Fusion)

### 3.1 Harmonic Relations (Principle of Harmonicity)

Let us first discuss the organization of acoustic components that are present at the same time. These are discussed in Chapter 3 of [2] and examples are given on the audio CD of [3].

We start with the principle of harmonicity. One of the most important methods used by ASA applies only to sounds with periodic waveforms, such as the human voice or the sound of a violin. In such sounds, the frequency components are all harmonics (integer multiples of the fundamental frequency ( $F_0$ ) of the sound). It is this  $F_0$  that provides the pitch of the sound. Therefore, it is a useful strategy for ASA to find all the frequency components that are multiples of the same frequency and segregate this set from the rest of the sound. Doing so also fuses them together to form a single sound with its own distinct pitch. For example, when a single harmonic is gradually mistuned from its proper frequency in the series, it gradually becomes audible as a separate pure tone [15].

Audio Example 1 on the CD-ROM that accompanies Audio Anecdotes (demonstration 18 from [3]) gives an example of this. You are to listen for the third harmonic of a complex tone. First, this component is played alone as a standard. Then over a series of repetitions, it remains at a constant frequency, while the rest of the components are gradually lowered, as a group, in steps of one percent. Now, after two more presentations of the target component alone, the other partials of the tone are raised in steps of one percent until the target component loses its identity and once more forms part of the complex tone.

More than one harmonic series can be found by the ASA system at the same time and isolated from the rest of the incoming sensory information. Otherwise, we would not be able to hear two or more voices, each with its own momentary pitch, at the same time. This basis for grouping is called the principle of harmonicity. In addition to being responsible for fusion and segregation of the harmonics of the same tone, harmonic relations play a role when two tones are played at the same time.

- (1) When the fundamental frequencies are in a simple ratio, such as 2:1, 3:2, or 4:3, two results occur: The tones blend together because they share many harmonics and this seems to fool the process that segregates sounds; also, the tones sound consonant rather than dissonant.
- (2) When the fundamentals are not in a simple ratio, many harmonics of the two tones beat with one another at different rates, leading

to perceived “roughness” or dissonance, and the two tones do not blend very well.

This example shows how the principles used by ASA *utilize regularities in the environment*. For example, it is highly probable, in the environment, that a set of frequency components that are all multiples of a single fundamental has come from a single source. This harmonic relationship would be an extremely unlikely event by chance in the natural world (excluding music). Similarly, when two sets of components are detected, each related to its own fundamental, the chances are pretty good that there are two periodic sounds present in the mixture. So by grouping the components according to the most probable causes of their observed pattern, ASA takes advantage of these regularities in the world. In fact, it is likely that all of the ASA principles of grouping take advantage of relations that tend to be present when a set of acoustic components are all part of the same sound. We can call this the principle of “ecological validity” in ASA.

### 3.2 Other Factors Influencing Simultaneous Grouping

Besides harmonicity, other cues that influence the grouping or fusion of components into a single sound are the following: (see also “Perceivable Audio Latencies,” (page 63)):

- Onset and offset asynchrony. Components of the same sound tend to start and stop together; therefore asynchronies of more than a few milliseconds in the onsets and offsets of components are taken as evidence that the components belong to different sounds—the greater the asynchrony the stronger the evidence. It is likely that this asynchrony also allows the auditory system to separate the direct sound from the reflected sound of the source. Room reflections and reverberation are expected to arrive in a delayed sequence (see Section 4).
- Envelope independence. When the amplitude envelopes of components (i.e., fluctuations in intensity) are not synchronized, this is taken as evidence that the components are parts of sounds produced by different sources; when the envelopes are synchronized, the components are treated as parts of the same sound. (It is possible that the cues of “onset and offset asynchrony” and “envelope independence” reduce to the same principle, since an onset or an offset can be viewed simply as a change in the amplitude envelope.)

- **Spatial separation.** If components arrive from different spatial direction, this favors their segregation into sounds produced by separate sources.
- **Spectral separation.** The frequency separation and clustering of spectral components affect segregation. For example, when two spectrally limited bands of noise are played concurrently, if they are far apart in the spectrum, with a spectral gap (absence of frequency components) between them, they will sound like two concurrent sounds, whereas when there is no spectral gap, they will sound like a single sound.

A weighted combination (not necessarily linear) of these features forms the perceptual basis for fusion and segregation of sounds. While multiple analytical processes are at work simultaneously, not all cues have the same weight, and more robust cues dominate over less powerful ones. This is useful because no one cue can be depended on absolutely. Here are some examples:

- Spatial differences are useless when the sounds caused by several independent events originate from around a corner.
- The strategy of finding one or more fundamental frequencies in a spectrum applies only to sounds that have periodic waveforms; but many sounds, such as those of footsteps, air conditioners, scraping sounds, doors slamming, and keyboards clicking are not periodic.
- Fine-grained details about the asynchrony of onsets of different sounds and the independence of their amplitude envelopes are washed out in reverberant environments.

## 4 Sequential Grouping

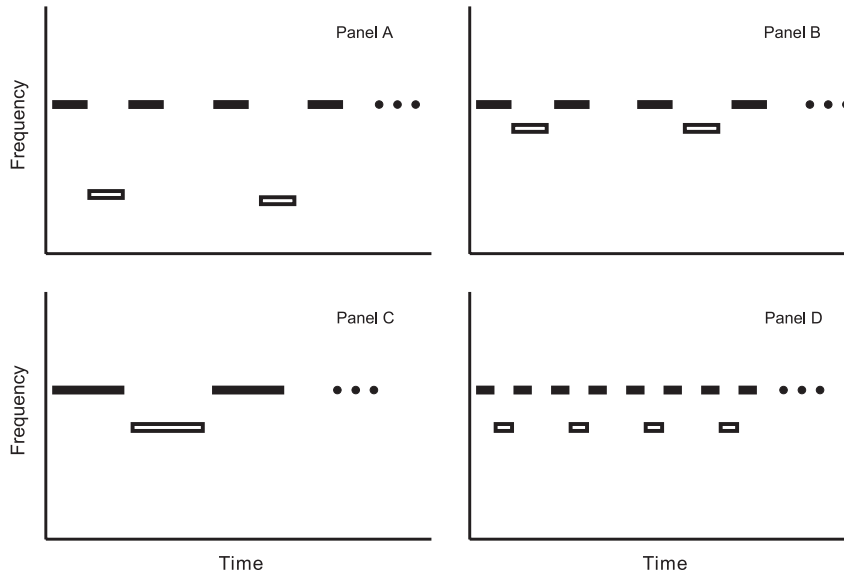
Certain ASA principles group sounds together over time, rejecting others as belonging to different sequences. The perceived sequences whose parts form a single perceptual entity are called “auditory streams.” As new sounds arrive, they are either assigned to existing streams, or form a new stream. Each stream has its own global properties, such as melody and rhythm, derived from the patterning of the sounds assigned to that stream, but not from sounds that are assigned to other streams. The principles of sequential grouping, which follow, are described in Chapter 2 of [2] and are illustrated in the audio CD of [3].

### Example 1. Sequential grouping (streaming)

A vivid demonstration of the formation of auditory streams occurs in the “streaming” phenomenon, in which the perceived order of events is not their actual physical order. A rapid sequence of tones may sound as if it were two sequences going on in parallel. This can be heard in Audio Example 2, Part A, which appears on the CD-ROM that accompanies Audio Anecdotes, (actually Demonstration 3 from [4]). The pattern is also illustrated in Figure 2, Panel A. The example consists of a repeating cycle of a high tone (H) and a low tone (L), in which high and low tones alternate in a galloping rhythm (e.g., HLH–HLH–HLH–...), —... where the dash represents a silence, equal in duration to a single tone. At a moderate frequency separation between high and low tones, if the cycle is played slowly the actual sequence is heard. However, as it is played faster and faster, it breaks apart perceptually into two different cycles, one involving the high tones (H–H–H–H–H–H–...), and the other the low ones, repeating at a slower rate (L—L—L—...). This segregation is called “streaming.” The listener’s attention switches from one cycle to the other, but cannot accurately perceive the timing relations between the two. Figure 2, Panel B, and Audio Example 2, Part B, on the CD-ROM, show what happens when H and L are close in frequency. The HLH– units remain integrated even at high speeds.

The streaming effect is explained as follows: We can define the acoustic “distance” (or difference),  $d$ , between any two nonsimultaneous tones, A and B, as the weighted combination (not necessarily linear) of their separations on a number of physical dimensions. Low values of  $d$  favor the perceptual integration of A and B into a single sequence whereas high values favor their segregation. In the example shown in Figure 2, two dimensions are involved, frequency and time, each of which contributes to  $d$ . At the slower speed (Figure 2, Panel C), when we consider adjacent high and low tones, the  $d$  between them—roughly proportional to their spatial separation on the graph—is less than the  $d$  between the nearest tones of the same frequency. Therefore, the adjacent high and low tones will be connected up in perception and we will hear all the tones as a single sequence. However, speeding up the sequence (Figure 2, Panel D) reduces the temporal separations while leaving the frequency separations unchanged. Now the nearest tone to a given one, in terms of  $d$ , is no longer the next tone in the overall sequence, but the nearest one of the same frequency. Therefore, the tones will connect up (high with high and low with low) to form two separate streams. This example illustrates two important factors in sequential grouping.





**Figure 2.** A repeating sequence of tones of higher (H) and lower (L) frequencies, in the pattern HLH—HLH—..., etc. In Panel A, the H and L tones are well separated in frequency. In Panel B, they are close in frequency. In Panel C, the sequence is slow, and in Panel D, it is fast.

- (1) Frequency separation. The further the sounds are from each other in frequency, the greater their tendency to segregate.
- (2) Temporal separation. It used to be believed that speeding up a sequence increased its segregation by reducing the onset-to-onset times of the tones in the sequence and speeding up the rhythm. However, it has been shown that speed works by shortening the time gap between the end of one tone and the beginning of the next [4]. Suppose we have a sequence of alternating high and low tones in which the onset-to-onset times (the rhythm) is held constant, but the frequency separation of high and low tones is small enough that the high and low tones don't form separate streams. If we now lengthen each tone, causing the time gap between it and the next tone to decrease, without slowing down the rhythm, there will be an increased tendency for the sequence to separate into two streams, because the lengthening lowers the offset-to-onset times of tones in the same frequency region, reducing the contribution of

temporal separation to  $d$ , leaving frequency separations to dominate  $d$ . This has important implications for understanding the effects of reverberation.

While the differences shown in Figure 2 are those of time and frequency, there are other acoustic factors that contribute to  $d$ , and hence to the grouping of the sounds in a sequence.

- (3) Differences in fundamental frequency ( $F_0$ ). Complex tones, such as the human voice or the violin, have many harmonics, all of which are multiples of a fundamental frequency. The fundamental, whether it is actually present in the sound, or absent, determines the pitch. The further apart the fundamental frequencies of two tones (and, hence, their pitches), the more they will segregate from one another in a sequence.
- (4) Differences in the shapes of the spectrum (e.g., the positions of peaks in the spectrum).
- (5) Differences in the spatial angles. from which the sounds arrive. Similar sounds will increasingly segregate from one another as the direction from which they arrive gets larger.
- (6) Differences in center frequencies; for example, when the sounds are band-passed noise bursts.

Other factors that produce weaker, but real, effects are:

- (7) Differences in intensity. This effect is not symmetrical. When sounds are of different intensities, the louder ones are easy to hear in isolation, but not the softer ones. The louder ones tend to mask the softer ones via sequential masking (the tendency for a softer sound to be masked by a louder one that comes either right before it (“forward masking”) or right after it (“backward masking”).
- (8) Differences in the rise times. (time from onset to maximum intensity) of the sounds.
- (9) Differences in the noisiness of the tones. Pure tones segregate from filtered noise bursts that have the same center frequency.

Note that Factors 4, 6, 8, and 9 influence the “timbre” of sounds. However, we prefer to relate stream segregation to the specific physical causes, one at a time, rather than to the general concept of timbre, since timbre is a poorly defined concept. In any case, timbre is a perceptual result, not a physical cause, and it may be simply a parallel result of the physical causes and not, in itself, a cause of grouping. Similarly, Factor 3 refers to the physical cause, fundamental frequency ( $F_0$ ), rather than the perceptual result, pitch, as one of the causes of grouping, even though different fundamental frequencies give different pitches. Again the perceptual result, pitch, may not be an actual cause of grouping, but just another perceptual effect of  $F_0$ .

- (10) The abruptness of the change between one tone and the next also affects their grouping. In a sequence in which high- and low-frequency tones alternate, if consecutive tones are joined by frequency glides, the sequence is more likely to hold together as a single perceptual stream. We can state this principle for the more general case in which  $v_a$  is the vector of properties of sound A, and  $v_b$  is the vector for sound B: If the transition between A and B involves a gradual transition between  $v_a$  and  $v_b$ , (i.e., interpolation of values), this will favor hearing A and B as parts of the same stream.
- (11) A competition between *ds* takes place in the formation of coherent streams. ASA compares alternative groupings and settles on those that minimize the within-stream *ds*. So it is possible that tones A and B may be grouped into the same stream when there are no better groupings available. However, when tones C and D are added to the sequence (where X is very similar to A and D very similar to B), tone A may be grouped with C into one stream and B with D into another stream. This means that it is not the absolute “distance” (or difference) between A and B in the property space that determines stream formation, but the size of the A-B distance relative to other distances between.

It appears that two stages of grouping are at work: The first is the very low-level grouping of acoustical features into timbres, pitches, event durations, and spatial locations; the second is the grouping of these perceptual attributes into separate sources acting over time, and a separate representation of their acoustic environments. This grouping may operate over some time window, because local properties may not provide enough information to create a robust image of the source and the environment.

There is an ecological basis for grouping sounds in a sequence according to their relative similarity and the smoothness of their changes: When physical objects interact to produce sounds (by impacts, scraping, oscillation, etc.), the properties of these physical objects will tend, on average, to change relatively slowly and continuously over time; so the sounds that they produce will change in the same way (by “slowly,” we mean over the course of second or so).

- (12) Grouping shows cumulative effects: It becomes stronger as the evidence accumulates over time. For example, the first few cycles of the tone sequence of Figure 2 are usually heard as a single stream. Then with more repetitions, the cycle may subdivide into high and low streams. The accumulation of evidence that there are tones in two distinct frequency regions builds up for at least four seconds, and after the sequence stops, it may take at least four seconds for the bias to disappear (these durations are rough approximations that may vary among listeners and among types of sound).

## 5 Competition between Simultaneous and Successive Grouping

The processes of simultaneous and successive grouping interact, for example, when a pure tone A rapidly alternates with a complex tone B that has several harmonics. If the frequency of A is close to, or matches, the frequency of one of the low harmonics (A') of tone B, instead of hearing a rapid alternation of A and B, the listener may hear a rapid succession of pure tones A–A'–A–A'–A–A'... accompanied by a slower sequence B'—B'—B'.... where B' is the complex tone B with A' removed. The effect becomes stronger as the silent gap between A and B is made shorter. This effect of the time gap between tones of similar frequency is the same as the one observed in sequential stream formation; so the capturing of A' by A is best regarded as a sequential grouping effect. This, then, is a case in which sequential and simultaneous integration are competing. A' is either part of a sequential grouping of pure tones, A–A'–A–A'..., or is part of tone B, contributing to its timbre. The choice of which allocation of A' actually occurs is determined by the relative strength of the factors favoring sequential and simultaneous grouping that we discussed earlier. In general, since sounds are layered rather than purely simultaneous or purely sequential, most cases of perceptual integration involve a compe-

tition for components: Is a set of components to be taken as a part of a sequence, or is it merely a subset of the simultaneous components that form an ongoing sound?

### 5.1 The Old-Plus-New Heuristic.

There is an interesting phenomenon in auditory perception that is called apparent continuity ([2]; [22]; [23]; [24]; [25]). Here is one of the many forms it can take (Audio Example 3 which appears on the CD-ROM that accompanies Audio Anecdotes, taken from [3], Demonstration 34): A short band of noise (B) with frequencies from 0 to 2000 Hz, is alternated with a longer band (A) that contains only B's lower frequencies. Noise A is perceived as present during noise B because A's frequencies group with the low ones of B to form one continuous low noise, and the high components of B are heard as an intermittent high sound. This is the residual formed when the lower frequencies of B are captured. In perception, B has been broken into two parts: Its lower part has been grouped with A to form an unbroken sound, whereas its higher part has formed a separate higher-pitched sound. Similarly, the high frequencies of B can be captured by a high noise leaving a low noise as the residual. In Audio Example 3, these two cases are presented twice in alternation. Remember that B is physically identical in the two cases. (See also demonstrations 28 and 29 in Bregman and Ahad, 1996.)

Another example involves the repeated alternation of a long, soft, steady tone with a brief, loud noise burst. The tone appears to continue through the bursts of noise, even though it is physically absent. However, this "apparent continuity" only occurs if the noise contains frequency components around the frequency of the tone. It has been posited that there is an ASA principle, named the "old-plus-new heuristic", which is responsible for these examples of apparent continuity" [2]. It can be stated as follows: "If a spectrum suddenly becomes more complex, determine whether the components of the sound that preceded the change are still there. If so, take those components and treat them as a continuation of the previous sound. Take the rest of the spectrum and treat it as a newly added sound." If the old sound appears again after the interruption, this serves to validate the decision and, therefore, to strengthen the effect. This heuristic has its ecological basis in the fact that sounds in the natural world are rarely solitary, purely synchronous, or purely successive, but are usually partially overlapped in time. The old-plus-new heuristic uses the moment of onset of a sound to detect its new properties and to add a new entity to the brain's representation of the auditory scene, while

carrying out the conservative strategy of maintaining the old entities as far as possible.

## 5.2 Combining the Information from Many Cues

ASA makes robust decisions about grouping by combining information from many features of the mixture. The contribution made by each cue to simultaneous or sequential grouping is not all-or-none but quantitative in nature. For example, the effect of frequency separation upon  $d$  is proportional to its size: Large separations produce greater effects. Because the effects of different acoustic variables are combined, if two variables—say frequency separation and spatial separation—can both be employed to achieve a certain degree of segregation, the sizes of the two variables can be traded off against one another. A reduction in frequency separation combined with a greater spatial separation may keep the segregation constant. However, not all variables are of equal strength. For example, in sequential grouping (as measured by the streaming effect), frequency separation appears to be much stronger than either loudness or abruptness of onset, so it may not be possible to obtain the effect of a large frequency difference on segregation by increasing differences in the loudness or abruptness of onsets in a sequence of sounds. Interesting situations may occur in which some cues “vote” for segregation and others for fusion. Their constantly changing balance may create the impression of hybrid sources that are “morphed” between partly fused and partly segregated, exhibiting both the features of the fused whole and those of the component sounds. The control of this process is the core of the music mixing process where sound sources are electronically reshaped to promote either blend or separation, or both.

## 6 Areas of Application

There are a number of areas of application—some already exploited, some not—for our scientific knowledge about ASA. These include the following:

Control of the layering in music by composers, instrumentalists and conductors. ASA principles can help in understanding how the blending or segregation of musical lines can be controlled by composers of either instrumental or synthesized music. It appears that traditional composers took advantage of human perceptual properties to keep voices or melodic lines distinct and to reduce or magnify the perception of dissonance [27], [?]. Modern composers, while not necessarily writing the same sort of

music, seem to be using the same ASA principles [16]. The study of ASA can furnish an understanding of how conductors can improve the blends of certain groups of instruments in a performance. A study of which instruments blend together better, and why, has been done by [18]. The use of ASA principles in music is also described in [2] and [19], and, in African music, by [26]. See also our article “Creating Mixtures: The Application of Auditory Scene Analysis to Audio Recording,” (page ??).

Design of signals for the workplace. An example might be a console at which an operator is monitoring a chemical factory or an airplane cockpit full of instruments, some of which use auditory signals. In such situations, for functions such as alarms and notifications, it is important to choose the best combination of sounds so that they won't lose their identities if several go off at the same time. Also, understanding how to link auditory and visual information in a collaborative way may depend upon principles of audio-visual grouping that resemble some of the principles of ASA (such as synchrony of changes).

Computer interfaces and data sonification. In human-computer interaction, people have questions about the best way of using sound in interfaces (see [13], especially the foreword by Bregman). How can we keep the sounds from interfering with other concurrent ones? Which sounds should be used for which functions? How can sound collaborate with pictures for maximum effectiveness? The terms *audification* and *sonification* are used to describe the process of presenting patterns of data by representing them as patterns of sound, much like a chart does by using a visual pattern. What features of the sound will make regularities in the data obvious when the sounds group in auditory perception? Are particular features best for emphasizing different aspects of the data? This gets us into the area of auditory metaphor, and to the question of whether some metaphors are better than others [10]. (See also “Designing the Auditory Narrative: A Methodology for Sound Design” by Maribeth Back (page ??); “Auditory Feedback for Computer Devices” by Linda Roberts and Cynthia Sikora (page ??); and “Auditory Masking in Audio Compression,” by Henrique S. Malvar (page 205).)

Designing better hearing aids. In hearing science, one might be concerned with why people with impaired hearing report that their major problem is in following a single voice in a mixture of voices. Why is this so, and can any device be designed that will help the listener deal with this

problem? Do people differ in their dependence on different cues for ASA (e.g., are some people more dependent on spatial cues and others on temporal asynchrony cues? ) If so, different therapeutic strategies might be used in fitting an appropriate hearing prosthesis to each individual person.

Robust speech recognition and music transcription by computer. The study of speech recognition by humans has shown that the ASA principles that apply to artificial sounds in the laboratory apply to speech sounds as well ([2], [7]), although the listener's knowledge about the speech signal also plays a role. In the area of Artificial Intelligence (AI), such questions as the following arise: How can a computer be programmed to segregate speech from background sounds for easier recognition? To what extent is it advisable to mimic how the human accomplishes this task? The attempt to develop computer systems that perform ASA on speech signals has come to be called Computational Auditory Scene Analysis (CASA). For those interested in robust speech recognition based on CASA, we can recommend [17] and [5]. Another AI question asks how a computer could be programmed to do automatic transcription of the music played by each instrument in a group [6].

Sound transmission. It is important to understand how the auditory system responds to errors and discontinuities in transmission. We can see an example in the laboratory phenomenon called "phonemic restoration." It depends on the old-plus-new heuristic that we mentioned earlier. If we splice out a short segment from a signal, leaving a gap about the length of a single syllable, the signal—as would be expected—sounds as if it has a gap in it. The listener's interpretation of the verbal material on the two sides of the gap will be altered because the offset of the sound will resemble a stop consonant, as will the onset after the gap. Therefore, if the brain tries to recognize the pieces that remain on the sides of the gap, it will make mistakes. However, if the gap is filled with loud white noise that contains sufficient energy at the frequencies present in the signal, no gap will be heard. Instead the signal will sound continuous, but "covered" in certain places by the noise. The interruption will not divide the longer sound into two perceptual pieces. Also, if the noise is not too long, our brains will supply a representation of the longer sound so that we hear it continuing behind the interruption. If the long sound is a tone, then that tone will be heard as continuing behind the noise. If the long sound is a spoken sentence, and if the missing part is predictable from the rest of the sentence, the missing part will be restored by our brains.



This process is called *phonemic restoration*. It is an example of the use of the old-plus-new heuristic, in which certain frequencies in the interrupting noise are interpreted as continuations of the long sound, the remainder being heard as an interrupting noise. This phenomenon suggests that, in speech reproduction or transmission, if short silent gaps are inevitable, comprehension of the signal can be increased by filling in the gaps with loud white noise. This may not be esthetically pleasing, but comprehension of the message will be better. It will work best when the gaps are fairly short.

Judge for yourself. Audio Example 4 on the CD-ROM first plays some speech in which half the sound has been eliminated by taking out every one-sixth-second segment and replacing it with a silence ([3], Demonstration 31). Notice how the speech sounds choppy and disconnected. Next, loud noise bursts replace the silences. Now the speech is experienced as more continuous and many more of the words are intelligible. Finally, the complete sentence is played with no interruptions. Research has shown that only the frequencies that are required to complete the speech (or other) signal need to be present in the noise. If the others are left out, the interrupting sound will be less loud, and hence less objectionable. This fact could be exploited in devices that tried to make gaps in signals less disruptive.

Perceptual coding/decoding of sound. Another specific area of application in sound transmission involves digital compression of sound to achieve lower transmission bit rates without any loss in perceived sound quality or with a loss that is acceptable within a given quality standard. The ultimate goal is to achieve a transparent codec (coder plus decoder) that would be judged as a perfectly neutral transmission medium (does not add to or take away from the original sound). Bit rate reduction is achieved by not encoding low-level sounds that are thought to be inaudible based on the established masking curves. (See also "Auditory Masking in Audio Compression" (page 205) by Henrique S. Malvar.) However, because these curves were developed using stationary sounds, the judgement of nonaudibility of low-level sounds is made without any consideration for any auditory streaming that may render some low-level sound patterns audible. It remains to be determined whether continuing sound patterns presented well below the masking level are audible (not necessarily as independent objects, but as parts contributing to stronger sounds), and whether they should be reconstructed in the coding/decoding process to render full transparency. For example, low-level overtones of an instrument may become segregated out of the noise because stronger partials

guide the auditory system in recognizing the total pattern of partials of the instrument.

Neurobiology. Questions arise for the area of neurobiology as well: How can we study perceptual grouping in animals (e.g., see [14]) so we can learn about the physiological mechanisms underlying it? Can electrical recording or imaging of the human brain ([1], [20]) provide an insight into the formation of auditory units? How do the brains of animals carry out ASA ([9], [8])?

In some applied areas, such as the composing of music, the designing of workplace signals, or the mixing of recorded music, existing practice is already guided by many of these ASA principles, although not via a formal understanding. The practitioners, whose own auditory systems are governed by these principles, need only listen to what they have created to find out whether the sounds blend and segregate in desirable ways. So the function of the principles that we describe will not be to alter how people in these fields shape sound. We hope, however, that by being aware of these principles, practitioners can understand more clearly what they are doing, and possibly shorten the process of experimentation. The contribution to the construction of computer programs capable of robust speech recognition, or the study of how ASA is implemented in the brains of animals or humans, is different. Here, the use of ASA principles did not precede their exposition by researchers in perception, but resulted from it.

## 7 Conclusions

There are many principles of grouping that have evolved through the auditory system to deal with the problem of auditory scene analysis. These have been extensively studied in the McGill Auditory Research Laboratory and elsewhere using very simple sounds, but are believed to work in all sonic environments, including musical ones. We have offered a brief sketch of some of these principles, and some ideas about how they apply in different practical contexts. For a discussion of other practical applications, see our article “Creating Mixtures: The Application of Auditory Scene Analysis to Audio Recording,” (page ??).

## Annotated Bibliography

- [1] C. Alain, S. R. Arnott, and T. W. Picton. “Bottom-Up and Top-Down Influences on Auditory Scene Analysis: Evidence from Event-

Related Brain Potentials.” *Journal of Experimental Psychology: Human Perception and Performance*. To appear.

*A recent study of perceptual organization using the method of event-related EEG evoked potentials of the brain to study the brain processes underlying the segregation of concurrent sounds. The stimuli were complex sounds containing multiple harmonics, one of which could be "mistuned" so it was no longer an integer multiple of the fundamental. Increasing the mistuning caused the mistuned partial to stand out as a separate sound, and was accompanied by negative and positive waves that peaked at 180 and 400 ms post-stimulus, respectively. The negative wave, referred to as the "object-related negativity," was present during passive listening, but the positive wave was not present. These results show separate bottom-up and top-down influences during auditory scene analysis. Distinguishing simultaneous auditory objects involved a widely distributed neural network.*

- [2] A. S. Bregman. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA.: Bradford Books, MIT Press, 1990.

*This book summarizes a huge body of empirical work done by the author and others in the field of auditory perception, and establishes their coherence via a theoretical framework (ASA) evolved by the author. It considers the theoretical questions involved in ASA, describes the major cues used by the auditory system to achieve ASA, and evolves a set of principles that describe the working of the system. The relation between top-down and bottom-up processes in ASA are examined. Working from the principles established in the earlier part of the book, the later chapters apply them to the perception of music and speech. This book has stimulated research in psychology (psychoacoustics, speech perception, psychology of music), biology, computer science, engineering, hearing science, and musicology (about 700 references).*

- [3] A. S. Bregman and P. Ahad. *Demonstrations of Auditory Scene Analysis: The Perceptual Organization of Sound*. Audio CD. Cambridge, MA: MIT Press, 1996.

*This is a set of 41 audio demonstrations of the principles of auditory scene analysis underlying the perceptual organization of sound. It illustrates many of the phenomena described in included on the CD-ROM that accompanies *Audio Anecdotes*. The 41 demonstrations illustrate the principles governing the integration of sounds over time (sequential integration), the perceptual integration or segregation of*

sounds presented at the same time (perceptual fusion), and the old-plus-new heuristic. Sequential integration is shown to depend on the time interval between successive sounds and their similarity. Similar sounds close together in time are shown to group together and exclude dissimilar sounds into separate streams. Similarities can be in the frequencies of components, in spatial locations, in fundamental frequencies, and in spectral envelopes and other aspects of timbre. Repetitions of a pattern are shown to favor its segregation into substreams, and smooth continuation between sounds to favor their integration into the same auditory stream. Examples show that in music, both melody and rhythm depend on how the sounds are organized into streams. In speech, the quality of a vowel can be altered by causing one of its harmonics to be heard as a separate sound. Also illustrated is the phenomenon of apparent continuity, in which parts of a soft sound are deleted and replaced by bursts of a louder sound, but the listener hears the soft sounds continue right through the louder one. Perceptual fusion of sounds presented at the same time is shown to be affected by the synchrony of their onsets and offsets, by their harmonic relations, and by synchronized changes in their amplitudes. Fusion, in turn, is shown to affect the perception of timbre. Most of these demonstrations can be used effectively in classrooms, but a few require listening over headphones.

- [4] A. S. Bregman, P. A. Ahad, P. A. C. Crum, and J. O'Reilly. "Effects of Time Intervals and Tone Durations on Auditory Stream Segregation." *Perception & Psychophysics* 62:3(2000), 626–636.

*Adult listeners listened to a sequence of high (H) and low (L) tones that alternated in a repetitive galloping pattern (HLH–HLH–HLH...). They could hear the gallop when the sequence was perceived as a single stream, but when it segregated into two substreams, they heard H–H–... in one stream and L—L—... in the other. The onset-to-onset time of the tones, their duration, the InterStimulus Interval (ISI) between tones of the same, and the frequency separation between H and L tones were varied. Results showed that the well-known effect of speed in increasing stream segregation is primarily due to its effect on the ISI between tones in the same frequency region. This has implications for several theories of streaming.*

- [5] G. J. Brown and M. P. Cooke. (1994) "Computational Auditory Scene Analysis." *Computer Speech & Language* 8:4(1994), 297-336.

*Presents a system that is able to segregate speech from a variety of intrusive sounds, including other speech, with some success. The segregation system consists of four stages that model (1) the auditory periphery, (2) extraction of auditory features into separate maps, and (3) construction of a symbolic description of the auditory scene. In the final stage of the system, a search strategy is employed which groups elements according to the similarity of their fundamental frequencies, onset times, and offset times. Following the search, a waveform can be resynthesized from a group of elements so that segregation performance may be assessed by informal listening tests (59 references).*

- [6] C. Chafe and D. Jaffe. "Source Separation and Note Identification in Polyphonic Music. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 1985.

*Describes experiments in automatic music recognition (analysis and transcription) by computer, of digitized sound recordings of polyphonic music with a variety of instruments and musical styles. The overall goal is to provide a tool for the study of musical performance, for applications requiring tracking of live musicians, for manuscript work, and for segmentation of digital audio recordings.*

- [7] C. J. Darwin and R. P. Carlyon. "Auditory Grouping." In Handbook of Perception and Cognition, Vol. 2, edited by B.C.J. Moore New York: Academic Press, 1985.

*Basic principles of ASA are discussed: the system's use of the property of harmonicity to "hear out" individual mistuned components; differences in fundamental frequency ( $F_0$ ) of different groups of harmonics; FM differences, onset and offset synchrony of partials; and AM phase differences. These principles are applied to the perception of speech sounds, e.g., separating voices by different  $F_0$ s, identifying speech sounds, pitch and timbre perception, and separate perceived locations in space.*

- [8] A. S. Feng and R. Ratnam. "Neural Basis of Hearing in Real-World Situations." Annual Review of Psychology 51(2000), 699–725.

*In real-world situations, animals are exposed to multiple sound sources originating from different locations. Most vertebrates have little difficulty in attending to selected sounds in the presence of distractors, even though sounds may overlap in time and frequency. This*

*chapter selectively reviews behavioral and physiological data relevant to hearing in complex auditory environments (141 references).*

- [9] Y. Fishman, D.H. Reser, J. C. Arezzo, M. Steinschneide. "Neural Correlates of Auditory Stream Segregation in Primary Auditory Cortex of the Awake Monkey." *Hearing Research* 151:1-2(2001), 167-187.

*Auditory stream segregation can be demonstrated by presenting When a sequence of high and low frequency tones in presented in an alternating pattern, ABAB, when the tone presentation rate is slow or the frequency separation between the tones is small (< 10%), a connected alternating sequence ABAB is perceived. When the rate is fast or the difference in frequency is large, however, the alternating sequence perceptually splits into two parallel auditory streams, one composed of interrupted 'A' tones, and the other of interrupted 'B' tones. The neurophysiological basis of this perceptual phenomenon is unknown. Neural correlates of auditory stream segregation were examined in area A1 of the awake monkey using neuronal ensemble techniques (multiunit activity and current source density). Responses evoked ABAB... sequences were studied as a function of rate (5, 10, 20, and 40 Hz). 'A' tones corresponded to the best frequency of the cortical site, while 'B' tones were situated away from the best frequency by varying frequency separations. At slow rates, 'A' and 'B' tones evoked responses that generated an overall pattern of activity at the stimulus rate. In contrast, at fast rates, 'B' tone responses were differentially suppressed, resulting in a pattern of activity consisting predominantly of 'A' tone responses at half the presentation rate. The magnitude of 'B' tone response suppression increased with the frequency separation. Differential suppression of best-frequency and non-best-frequency tone responses at high rates can be explained by physiological principles of forward masking. The effect of frequency separation is explained by the hypothesis that responses to tones distant from the best frequency are more susceptible to suppression by best-frequency tones than responses to tones near the best frequency. These results parallel human psychoacoustics of auditory stream segregation and suggest a cortical basis for the perceptual phenomenon.*

- [10] W. W. Gaver. "Auditory Icons: Using Sound in Computer Interfaces." *Human-Computer Interaction* 2(1986), 167-177.

*Gaver is a pioneer in the use, in computer interfaces, of sounds that are essentially "caricatures" of natural sound, such as dragging sounds, impacts, etc. For example, an incoming email could "hit"*

*the mailbox, with a realistic hitting sound, the length of the message affecting the “weight” behind the impact. Thus, the magnitude being portrayed is represented by the magnitude of some property of the cartoon world whose physics determines the property of the sound. Notice how this is different from mapping the size of the file directly onto the intensity of the sound. Gaver’s approach is contrasted with the use of sounds, such as tone sequences that have only an arbitrary relation to the class of event being portrayed. He thinks that the listener should “hear the world, not the sound.” Use of his method reduces the learning time, since the quality of natural sounds is governed by physical laws affecting the things that make those sounds. The caricature world can employ the same physical laws that people are used to.*

- [11] D. Huron. “The Avoidance of Part-Crossing in Polyphonic Music: Perceptual Evidence and Musical Practice.” *Music Perception* 9:1(1991a), 93–03.

*Huron has a huge database of musical scores in computer-readable form on which he tests hypotheses about the composer’s strategies, using a software system called “Humdrum.” The cited paper is a study of part-crossing in 105 polyphonic works by J. S. Bach, in which Huron confirms that Bach showed a marked reluctance to have parts cross. When the textural density increased beyond two concurrent voices, Bach became more vigilant to avoid part-crossing. In light of evidence from another study of Huron’s—that perceptual confusion increases with the number of concurrent voices—these results are consistent with the hypothesis that Bach tried to minimize perceptual confusion as the density of auditory images increased.*

- [12] D. Huron. “Tonal Consonance versus Tonal Fusion in Polyphonic Sonorities.” *Music Perception* 9:2(1991b), 135—154.

*An analysis of a sample of polyphonic keyboard works by J. S. Bach shows that the prevalence of different vertical intervals (pitch separation of simultaneous notes) is directly correlated with their degree of tonal consonance. A major exception to this pattern arises with respect to those intervals that contribute to tonal fusion. The prevalence of the latter intervals is negatively correlated with the degree to which each interval promotes tonal fusion. Bach’s avoidance of tonally fused intervals is consistent with the objective of maintaining the perceptual independence of the contrapuntal voices. In summary, two factors appear to account for much of Bach’s choice of vertical*

*intervals: the pursuit of tonal consonance and the avoidance of tonal fusion.*

- [13] G. Kramer, Editor. Auditory Display: Sonification, Audification, and Auditory Interfaces. Proceedings, Vol. XVII, Santa Fe Institute. Reading, MA: Addison-Wesley, 1994.

*Describes the state of the art in 1994 in the use of sound to communicate signals and quantitative information to a person working with data on a computer. Kramer initiated an annual conference called the International Conference on Auditory Display (ICAD) which meets every year, and discusses auditory exploration of data via sonification (data-controlled sound) and audification (audible playback of data samples); real-time monitoring of multivariate data; sound in immersive interfaces (virtual environments) and teleoperation; perceptual issues in auditory display; sound in generalized computer interfaces; technologies supporting auditory display creation; data handling for auditory display systems; and applications of auditory display. At the time of first publication of this book, the papers were available on the ICAD web site [www.icad.org](http://www.icad.org).*

- [14] S. A. MacDougall-Shackleton, S. H. Hulse, T. Q. Gentner, and W. White. (1998) “Auditory Scene Analysis by European Starlings (*Sturnus vulgaris*): Perceptual Segregation of Tone Sequences.” *Journal of the Acoustical Society of America* 103:6(1998), 3581–3587.

*Like humans, animals that use acoustic stimuli to perceive their world ought to be able to parse the auditory scene into functionally significant sounds. The ability to do so ought to have significant adaptive value when, for example, an animal can identify the sounds of a predator among other natural noises. In earlier work, it was shown that a species of songbird, the European starling, can identify excerpts of both its own song and songs from other avian species when the songs are mixed concurrently with other natural signals. In this experiment, it is demonstrated that starlings can segregate two synthetic pure-tone sequences when the sequences differ in frequency. Taken together, the experiments show that at least one nonhuman species is capable of auditory scene analysis both for natural and for non-natural acoustic stimuli. This suggests, in turn, that auditory scene analysis may be a general perceptual process that occurs in many species that make use of acoustic information.*

- [15] B. C. J. Moore, B. R. Glasberg, and R. W. Peters. “Thresholds for Hearing Mistuned Partial as Separate Tones in Harmonic Com-



plexes.” *Journal of the Acoustical Society of America* 80(1986), 479–483.

*This study showed that when a low harmonic in a harmonic complex tone is mistuned from its harmonic value (i.e., is no longer an exact multiple of the fundamental) by a sufficient amount, it is heard as a separate tone, standing out from the remainder of the complex tone (which is still heard as a whole). The experiment estimated the degree of mistuning required for this to occur for a number of the lower frequency components of a complex tone that had 10 or 12 equal-amplitude components (60 dB SPL per component).*

- [16] R. Mountain. “An Investigation of Periodicity in Music, with Reference to Three Twentieth-Century Compositions: Bartok’s Music for Strings, Percussion, & Celesta; Lutoslawski’s Concerto for Orchestra; Ligeti’s Chamber Concerto. Ph.D. diss., School of Music, University of Victoria, 1993.

*In her study of the role of periodicity in music, she makes reference to perceptual tendencies including Gestalt principles of grouping, the phenomena of auditory streaming and fusion, and the temporal limits of the perceptual present. Relationships between levels of periodic events are described in terms of rhythmic consonance and dissonance.*

- [17] D. F. Rosenthal and H. G. Okuno. *Computational auditory scene analysis*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 1998.

*This book arose from the first meeting of researchers in Computational Auditory Scene Analysis (CASA) and represents a variety of approaches to the subject. Many of the systems use acoustic information that is known to be used in human auditory scene analysis. Most systems are oriented towards speech recognition in complex backgrounds, but some deal with music. Researchers come from various universities, industries, and research institutes in Canada, the US, England, Japan, Israel, Germany, and Singapore. Methods include simulating the peripheral auditory system, cross-channel summation of simulations of neural firing, correlation of neural outputs, wavelet analysis, and use of “sound atoms.” There are systems that use features such as amplitude modulation, ones that group features by synchronizing “neural” oscillators: There are binaural models; blackboard systems; multiagent systems; and ones that use Bayesian probability, information-theoretic approaches, or knowledge-based analysis.*

- [18] G. J. Sandell. "Concurrent Timbres in Orchestration: A Perceptual Study of Factors Determining "Blend." Ph.D. diss., Northwestern University, 1991.

*Evaluates the "blend" of combinations of musical sounds: whether the instruments fuse into a single composite timbre, segregate into distinct timbral entities, or fall somewhere in between the two extremes. This study investigates, through perceptual experimentation, the acoustical correlates of blend for 15 natural-sounding orchestral instruments presented in concurrently sounding pairs (e.g., flute-cello, trumpet-oboe, etc.). Ratings of blend show the strongest effects for centroid (the location of the midpoint of the spectral energy distribution) and duration of the onset for the tones. Lower average values of both centroid and onset duration for a pair of tones lead to increased blends, as do closeness in value for the two factors. To confirm the importance of centroid as an independent factor determining blend, pairs of tones including instruments with artificially changed centroids are rated for blend. Judgments for several versions of the same instrument pair show that blend decreased as the altered instrument increased in centroid, corroborating the earlier experiments. A survey of orchestration manuals shows many illustrations of "blending" combinations of instruments that are consistent with the results of these experiments. This study's acoustically based guidelines for blend serve to augment instance-based methods of traditional orchestration teaching, providing underlying abstractions helpful for evaluating the blend of arbitrary combinations of instruments.*

- [19] J. A. Sloboda. *The musical mind: the cognitive psychology of music.* New York: Clarendon /Oxford University Press, 1985.

*This is an excellent book on all aspects of the psychology of music. Chapters include "Music as a Cognitive Skill"; "Music, Language, and Meaning"; "The Performance of Music"; "Composition and Improvisation"; "Listening to Music"; "Musical Learning and Development"; "The Musical Mind in Context: Culture and Biology." Chapter 5, discusses primitive grouping mechanisms, mentioning the work of Deutsch, Johanssen, Bregman, Dowling, and Kubovy. It describes musical examples of grouping by frequency, sequential capturing of tones, and attention-based grouping. Discusses the role of melodic structure in generating "implications," expectations, figure and ground, and how harmonic coherence aids attention.*

- [20] E. Sussman, W. Ritter, and H. G. Vaughan. "An Investigation of the Auditory Streaming Effect Using Event-Related Brain Potentials." *Psychophysiology*, 36(1999), 22–34.

*There is uncertainty concerning the extent to which the auditory streaming effect is a function of attentive or preattentive mechanisms. This research uses MisMatch Negativity (MMN), which indexes preattentive acoustic processing, to probe whether the segregation associated with the streaming effect occurs preattentively. Alternating high and low tones are presented at fast and slow paces while subjects ignore the stimuli. At the slow pace, tones are heard as alternating high and low pitches, and no MMN is elicited. At the fast pace, a streaming effect is induced and an MMN is observed for the low stream, indicating a preattentive locus for the streaming effect.*

- [21] H. Wallach, E. Newman, and M. Rosenzweig. "The Precedence Effect in Sound Localization." *American Journal of Psychology* 62:3(1949), 315–336.

*The authors discuss the "precedence effect" in this 1949 paper two years before the publication of the 1951 paper by Haas, "On the Influence of a Single Echo on the Intelligibility of Speech" in *Acustica* 1:49–58, after which the precedence effect became also known as the "Haas effect." The authors point out that, in 1930, von Békésy reported the existence of a precedence effect, "meaning a suppression or concealment of the sound image which arrives later" (v. Békésy, "Zur Theorie des Hörens," *Physical. Zsch.* 31(1930), 858–859). Using earphones and loudspeakers, the authors study the time interval over which fusion of two sounds takes place. The integration interval is found to be between 5 ms for clicks and up to 40 ms for more complex sustained sounds. The localization of the fused sound is largely determined by the localization of the first arriving sound, and is based on the time difference between the two ears when the two sounds have nearly the same intensity. The authors point out that the precedence effect can be overridden if the later arriving sound is 15 dB or more louder than the first sound.*

- [22] R. M. Warren. *Auditory Perception: A New Synthesis*. New York: Pergamon Press, 1982.

*This textbook summarizes the work done by Richard Warren up to 1982. One chapter describes the phenomenon of auditory induction, also known as "apparent continuity," "illusory continuity," and "perceptual restoration". It occurs when a softer sound is interrupted by a*

*louder sound and the softer one is turned off during the interruption. Despite its absence during the interruption, the softer one appears to continue behind the louder one. The phenomenon can occur with a variety of softer sounds: steady state tones, tones that glides in pitch, a speech excerpt, a musical scale, and other sounds. The louder sound must be loud enough and contain the right frequencies to have masked the softer sound had the latter actually been present during the interruption.*

- [23] R. M. Warren. "Perceptual Restoration of Obliterated Sounds." *Psychological Bulletin*, 96(1984), 371–383.

*A comprehensive survey and discussion of the phenomenon of auditory induction, laying out the requirements for obtaining it, and an ecological explanation of its utility for the listener.*

- [24] R. M. Warren and C. J. Obusek. "Speech Perception and Phonemic Restorations." *Perception and Psychophysics*, 9(1971), 358–362.

*When a stream of speech has short sections, about the length of a single phoneme, removed and replaced by loud noise, the listener hears the speech as continuing behind the interruptions, including the physically absent sound. This type of illusory continuity is known as phonemic restoration. It depends both on acoustic factors and on the listener's knowledge of the language. This article reviews the phenomenon and the circumstances under which it can be obtained.*

- [25] R. M. Warren, C. J. Obusek, and J. M. Ackroff. "Auditory Induction: Perceptual Synthesis of Absent Sounds." *Science*, 176(1972), 1149–1151.

*In auditory induction (apparent continuity, illusory continuity), the louder interrupting sound must be a potential masker for the softer one. "If there is contextual evidence that a sound may be present at a given time, and if the peripheral units stimulated by a louder sound include those which would be stimulated by the anticipated fainter sound, then the fainter sound may be heard as present." See also *The continuity of a narrow-band noise interrupted repeatedly by a louder broad-band noise may persist through such interruptions, even when latter are up to tens of seconds in duration.**

- [26] U. Wegner. "Cognitive Aspects of Amadinda Xylophone Music from Buganda: Inherent Patterns Reconsidered." *Ethnomusicology*, 37(1993), 201–241.

*Describes the role of stream segregation in the music of the amadinda, a type of xylophone, used in the music of East Africa. Two musicians play notes, with strict alternation of the notes of the two players. Although each player's part is isochronous (exactly even tempo), a complex rhythm and melody emerges because of perceptual segregation of the high and low notes, some contributed by each player. The perceptual "streams" that emerge carry melodies that the listeners know, but not the complete melodies. The listeners must imagine the missing notes, as is often the case in Western jazz variations on melodies.*

- [27] J. K. Wright and A. S. Bregman. "Auditory Stream Segregation and the Control of Dissonance in Polyphonic Music." *Contemporary Music Review*, 2(1987), 63-93.

*The theory of auditory stream segregation is introduced and used in the study of the linear and harmonic dimensions of polyphonic music. The paper focuses on how the perceptual organization can be manipulated to control the perceptual salience of harmonic dissonance, arguing that it can be made less salient if the dissonant notes are prevented from fusing perceptually. The approach demystifies certain well-established principles of traditional music theory and provides new ways to explain contrapuntal phenomena.*