Demonstrations

to accompany Bregman's

Auditory Scene Analysis

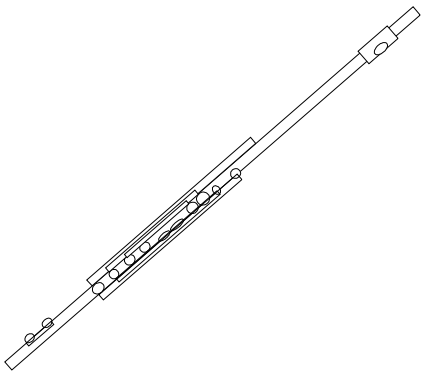
The perceptual organization of sound

MIT Press, 1990

Albert S. Bregman
Pierre A. Ahad

Department of Psychology
Auditory Research Laboratory
McGill University

To the memory of John Macnamara,
scholar and friend.



Contents

Introduction

The demonstrations on this disk illustrate principles that lie behind the perceptual organization of sound. The need for such principles is shown by the following argument: Sound is a pattern of pressure waves moving through the air, each sound-producing event creating its own wave pattern. The human brain recognizes these patterns as indicative of the events that give rise to them: a car going by, a violin playing, a woman speaking, and so on. Unfortunately, by the time the sound has reached the ear, the wave patterns arising from the individual events have been added together in the air so that the pressure wave that reaches the eardrum is the sum of the pressure patterns coming from the individual events. This summed pressure wave need not resemble the wave patterns of the individual sounds.

As listeners, we are not interested in this summed pattern, but in the individual wave patterns arising from the separate events. Therefore our brains have to solve the problem of creating separate descriptions of the individual happenings, but it doesn't even know, at the outset, how many sounds there are, never mind what their wave patterns are; so the discovery of the number and nature of the sound sources is analogous to the following mathematical problem: “The number 837 is the sum of an unknown number of other numbers; what are they? There is a unique answer.”

To deal with this scene analysis problem, the first thing the brain does is to analyze the incoming array of sound into a large number of frequency components. But this does not solve the problem; it only changes it. Now the problem is this: how much energy from each of the frequency components, present at a given moment, has arisen from a particular source of sound, such as the voice of a particular person continuing over time? Only by solving this problem can the identity of the signals be recognized.

For example, particular talkers can be recognized, in part, by the frequency composition of their voices. However, there are many more frequencies arriving at the ear than just the ones coming from a single voice. Unless the spectrum of the voice can be isolated from the rest of the spectrum, the voice cannot be recognized. Furthermore, the recognition of what it is saying – its linguistic message – depends on the sequence of sounds coming from that voice over time. But when two people are talking in the same room, a large set of acoustic components will be generated. These have to be stitched together in the right way. Otherwise illusory syllables could be perceived by grouping components derived from both voices into a single stream of sound.

The name given to the set of methods employed by the auditory system to solve this problem is "auditory scene analysis", abbreviated ASA. This name emphasizes the analogy with "scene analysis", a term used by researchers in machine vision to refer to the computational process that decides which regions of a picture to treat as parts of the same object. It has been argued by Bregman (1990) that there exists a body of methods for accomplishing auditory scene analysis that are not specific to particular domains of sound such as speech, music, machinery, traffic, animal sounds, and so on, but cut across all domains.

These methods take advantage of certain regularities that are likely to be present in the total spectrum whenever it has been created by multiple events. The regularities include such things as harmonicity, the tendency of many important types of acoustic event to generate a set of frequency components that are all multiples of the same fundamental frequency. Here is an example of how the auditory system uses this environmental regularity: If it detects two different sets of harmonics (related to different fundamentals) it will decide that each set represents a different sound. There are many other kinds of regularities in the world that the brain can exploit as it tries to undo the mixture of sounds and decide which frequency components to fit together. They include the fact that all the acoustic components from any single sonic event (such as a voice saying a word) tend to rise and fall together in frequency and in amplitude, to come from the same spatial location, and that the spectrum of the particular event does not change in its frequency profile (spectrum) too rapidly.

Illustrations of some of these regularities and how they affect grouping are given by the demonstrations on this disk. They are meant to illustrate the principles of perceptual organization described in the book, ***Auditory Scene Analysis: The Perceptual Organization of Sound*** (Bregman, 1990), published by the MIT Press. This book will be mentioned fairly often; so its title will be abbreviated as "***ASA-90".*** The phenomenon of auditory scene analysis, itself, will be abbreviated simply as "***ASA".***

ASA-90 attempts to integrate the phenomena of the perceptual organization of sound by interpreting them as parts of ASA. It also applies the same framework to the study of music and speech, and connects the problem of auditory grouping to the "scene analysis" problem encountered in machine vision.

The research described in ***ASA-90*** has shown that the well-known Gestalt principles of grouping, conceived in the early part of this century to describe the perceptual organization of visual stimuli, can also be found, in a modified form, in auditory perception, where they facilitate the grouping together of the auditory components that have been created by the same sound source. While the Gestalt principles have been shown to be useful, it is the contention of ***ASA-90*** that they are merely a subset of a larger set of scene analysis principles, some of which are unique to particular sense modalities.

Choice of demonstrations.
For the present disk, we tried to choose demonstrations that a listener should be able to hear without special training or conditions of reproduction. For this reason, they do not always correspond directly to the stimulus patterns used in the research discussed in ***ASA-90***, many of which require training of the listener, presentation against a quiet background, and statistical evaluation before regularities can be seen. However, the present examples illustrate the same principles.

References.
In each description in the booklet, there is section entitled "Reading". This pertains to chapters and page numbers in the ***ASA-90*** book, and to other publications. The articles cited in the description of each demonstration are collected at the end of the booklet. Many of these are discussed in ***ASA-90***.

The use of cycles as stimuli.
Many of the demonstrations use a repeating cycle of sounds to illustrate principles of perceptual organization. While not typical of our acoustic environment, cycles have a number of advantages. One is that a short sequence of sounds can be repeated to make sequences of any desired length. Although they vary in length, they are still subject to simple descriptions, and can be generated simply. When we explain the demonstrations, we use the *ellipsis* symbol, (**...**) to mean "repeated over and over", as in "ABAB**...**".

A second reason for using cycles is that segregation increases over time. Cyclic presentation allows us to drive, to unnaturally high levels, the segregative effects of the stimulus properties that we are examining. Furthermore, the use of cycles allows the listener to have repeated chances to observe the ensuing perceptual effects, allowing stable judgments to be made. By using sequences composed of a large number of repetitions, we can also minimize the special effects that occur at the beginning and ends of sequences (e.g., echoic memory) so that purer effects can be observed.

In many of the demonstrations, we present high (H) and low (L) tones in a galloping sequence, a pattern first used by van Noorden (1975) to study the segregation of auditory streams. When the sequence HLH-HLH-HLH-**...** segregates into a high and a low stream, the galloping rhythm seems to disappear. Instead we hear a regular rhythm of the high tones H-H-H-H-H-H-**...** and a slower regular rhythm of the low tones -L---L---L--. This change in rhythm and melodic pattern makes it easy for listeners to recognize that stream segregation has taken place.

Standard and comparison patterns.
In order to clarify how you are organizing the sequence of sounds, many of the demonstrations ask you to listen for a particular pattern, A, inside a larger pattern, B.

The pattern that you are to listen for, A, is presented first, in the form of a "standard". Then, right afterwards, the larger pattern, B, is played as a "comparison sequence", which is always more complex than the standard. It may have more sequential components or more simultaneous ones. If some standard (A1) can more easily heard than other

standards (A2, A3, etc.) in a given comparison pattern, this implies that sub-pattern A1 is more strongly isolated from the rest of B, by principles of grouping, than the other sub-patterns are.

If you concentrate very hard, you may be able to hear the standard whether or not perceptual organization favors its isolation. So try to listen to the standards in all conditions with the same degree of attention. Then you should be able to tell whether the isolation of the standard has been helped by the grouping cues whose effects are being examined in that demonstration.

Figures.
There are a set of conventions that govern the format of the figures. We will list them now. Should this format be altered for a particular figure, it will be explained in the text.

1. In most of the figures there are two or more panels which are referred to, in the text, as Panel 1, Panel 2, etc. The panel numbers are not included in the figure itself, but the order of numbering is consistent, going from left to right and then from top to bottom; i.e., as in normal reading.

2. The format of most of the displays are schematic spectrograms, with time on the horizontal axis and frequency on the vertical. Tones are usually represented as horizontal black bars. A noise burst appears as a rectangle with a gray pattern filling it; its horizontal extent indicates duration and its vertical extent, the range of included frequencies.

Monophonic versus stereophonic presentation.
Most of the demonstrations are monophonic (same signal on both channels). This is because spatial location is only one of many cues used by ASA. The mono demonstrations, 1 to 37, will work when listened to over loudspeakers as long as there is little reverberation in the room ("dry" listening conditions). If the room is too reverberant, headphones should be used. Only Demonstrations 38 to 41 are in stereo. They are grouped at the end of the disk for convenience. Although listening to these stereo examples over loudspeakers may reproduce some of the effects, headphones are strongly recommended.

Track numbers for demonstrations and calibration signals.
For simplicity, the first 41 track numbers on the disk correspond with the 41 demonstration numbers in this booklet. However, there are two extra tracks at the end of the disk. The first one, track 42, contains the loudness calibration signal described later in this section. The second one, track 43, is a signal for calibrating the stereo balance of your playback equipment. It is described on page 81 in this booklet.

Shaping onsets and offsets.
When a sound is turned on instantly, the listener hears a click. To prevent this, we turn sounds on and off gradually. Whenever the description of a signal mentions a rise (onset) time or a decay (offset) time, the amplitude of the signal is shaped, over time, by a quarter-sine-wave function, the first quarter of the sine wave for onsets and the second

quarter for offsets. These functions seem to minimize the perceived onset and offset clicks, as compared, in our laboratory, with other functions typically used for this purpose.

Laboratory facilities.
The demonstrations were created in the Auditory Research Laboratory of the Department of Psychology at McGill University. The computers were IBM-compatible PC's using Data Translation 16-bit converters (DT-2823) for acquisition and playback. The sampling rate was 22255 samples/sec for all synthesized and sampled signals, except for the sine-wave speech which was synthesized and played back at 20,000 samples/sec. Signals were recorded and played back using 8-kHz low-pass filters with a passive Tchebychev design having 60 dB attenuation at 11.2 kHz, THD < 0.1%. As a result of this filtering, the "white noise" referred to in various examples is actually 0-8 kHz flat-spectrum noise.

Software.
The signal processing software was version 8.1 of the MITSYN system of William Henke (1990). The playback program that controlled the sequencing of the sounds to form patterns, demonstrations and the overall program of demonstrations was written in MAPLE, a language specified by Albert Bregman and designed and implemented by André Achim and Pierre A. Ahad as a superset of the ASYST (1982) programming language.

Recording of announcements.
The spoken announcements were taped at the Recording Studio of the Faculty of Music of McGill University, using a Neumann U87 cardioid microphone placed about a foot from the announcer, and a Sony model DT-90 digital tape recorder, set to 48 kHz. The announcer was Albert Bregman.

N.B. Procedure for setting the playback volume.

Track 42 presents a sound pattern for calibrating the volume of your playback equipment. It contains both the loudest and softest sounds that are present in the demonstrations. You will hear a soft tone interrupted by a loud noise, played repeatedly for twenty seconds. Start with the volume low and turn it up gradually. Stop at the point at which the soft tone is heard clearly but the noise is not too loud.

Make a note of this setting and ***never*** set the volume higher than this in listening to the demonstrations. In some cases, you may want to turn it up or down for a particular demonstration, but in resetting it afterwards, ***never exceed the setting derived from the calibration procedure.*** If you play the sound too loud, you run the risk of damaging the playback equipment or your ears. Caution about volume is particularly important when listening over headphones.

Here is a hint: if you are listening over loudspeakers and the high frequencies seem weak, instead of compensating by increasing the volume, aim the speakers directly at you.

Overview

Sequential and simultaneous integration.
There are at least two dimensions of perceptual grouping. The first is sequential, in which bits of auditory data are connected across time. An example would be connecting the parts of the same melody together. The second is simultaneous, in which pieces of data arriving at the same time are either integrated or segregated from one another. An example would be the awareness of three notes as separate entities in a chord, or the segregation of one talker from another. We begin by illustrating sequential integration (Demonstrations 1 to 17). In a second set (18 to 24) we illustrate the integration or segregation of simultaneous components.

The old-plus-new heuristic.
The third group of demonstrations (25 to 37) illustrate the "old-plus-new heuristic" which helps in the decomposition of a mixed spectrum by comparing a complex spectrum with an immediately preceding simpler one. This heuristic can also be seen as a case of competition between sequential and spectral organization.

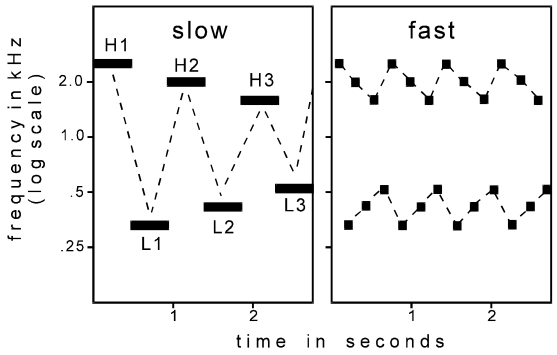
Most of the already mentioned demonstrations will work quite well without headphones, i.e., over loudspeakers, as long as the room is not too reverberant. If it is, try positioning yourself close to the speakers. However, in the fourth set, Demonstrations 38 to 41, we present a number of dichotic demonstrations *which do* require headphones. They illustrate each of the previously listed three categories of perceptual organization, but use spatial cues to control the organization.

Sequential integration

When parts of a spectrum are connected over time, this is known as sequential integration. An example is the connecting of the notes of the same instrument together to create the melody carried by that instrument. Another example is our ability to continue to hear a sound when other sounds join it to form a mixture.

Sequential integration is favored by the absence of any sharp discontinuities when changes occur in the frequency content, timbre, fundamental frequency, amplitude, or spatial location of a spectrum of sound.
Sequential grouping leads to the formation of auditory streams, which represent distinct environmental events and serve as psychological entities that bear the properties of these events. For example, when a stream is formed, it can have a melody and rhythm of its own, distinct from those of other concurrent streams. Also, fine judgements of temporal order are made much more easily when they are among sounds in the same stream.

1. Stream segregation in a cycle of six tones.



This demonstration begins the set that illustrates sequential organization. It is based on an experiment by Bregman and Campbell (1971), one of the early examinations of stream segregation that used a cycle of alternating high and low tones. The sequence used in the present demonstration consists of three high and three low tones in a six-tone repeating cycle. Each of the high tones (H1, H2, H3) and the low tones (L1, L2, L3) has a slightly different pitch. The order is H1, L1, H2, L2, H3, L3, ...

First we hear the cycle played slowly, as shown in Panel 1, and can clearly hear the alternation of high and low tones. This is indicated by the dashed line joining consecutive tones. Then, after a silence, it is played fast, as shown in Panel 2. We no longer hear the alternation. Instead, we experience two streams of sound, one formed of high tones and the other of low ones, each with its own melody, as if two instruments, a high and a low one, were playing along together. This is indicated by the two dashed lines, – one joining the high tones and the other joining the lower ones.

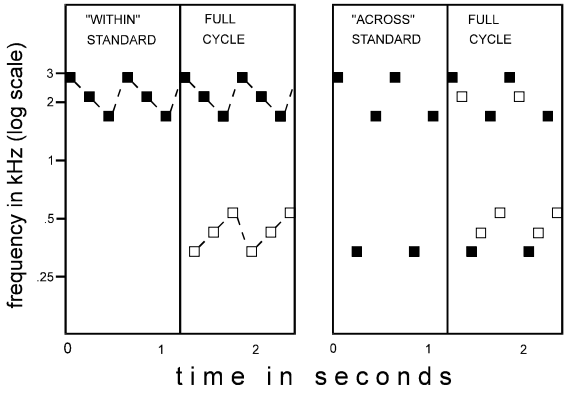
Not only does this demonstration show that stream segregation becomes stronger at higher tone rates but also that segregation affects the perceived melodies. At the slow speed, a full six-note melody is heard, but at the high one, we hear two distinct three-note melodies. The ability to correctly perceive the order of the tones is also affected. For example, in the experiment by Bregman and Campbell, when inexperienced listeners were asked to write down the order of the tones in the fast sequence, about half wrote either that a set of three high ones preceded a set of low ones, or vice-versa (i.e., HHHLLL or LLLHHH).

Technical details: The frequencies of the three high tones, taken in order, are 2500, 2000, and 1600 Hz, and the low ones are 350, 430, and 550 Hz. There are no silences between them. Each tone has 10-msec onsets and offsets. The onset-to-onset time of successive tones is 400 msec in the slow sequence and 100 msec in the fast one. To

equate for salience, the intensity of the low tones is made 6 dB higher than that of the high ones.

Reading: This pattern is discussed in ***ASA-90***, pp. 50, 139-40, 147, 153, and other parts of Ch.2.

2. Pattern recognition, within and across perceptual streams.



Perceptual organization normally assists pattern recognition by grouping those acoustic components that are likely to be part of the same meaningful pattern. In the laboratory, however, we can arrange it so that the perceptual organization of auditory material does not correspond with the units that the subject is trying to hear. This breaking up of meaningful patterns, so as to cause inappropriate groupings, is the principle behind camouflage. The present demonstration can be viewed as an example of auditory camouflage.

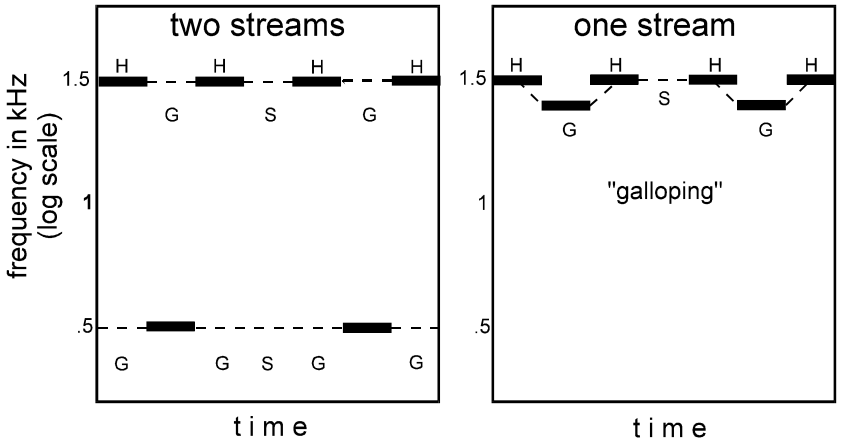
It uses stream segregation to affect pattern recognition. A six-tone cycle, like the one in Demonstration 1, is the test stimulus. We show that when it splits into two perceptual streams, it is very difficult to pay attention to members of the high and low streams at the same time. It seems that the two-stream structure limits what attention can do. The difficulty of considering both high and low tones in the same act of attention is demonstrated by asking you to listen for a subset of three tones from among the six (Panels 1 and 2). Detecting the subset within the full pattern can be done much more easily when the three tones are in the same stream than when they cross streams.
In this demonstration, the subset that you are to listen for is presented first as a standard, a repeating three-tone cycle (left sides of Panels 1 and 2). The full six-tone cycle is played right afterward (right sides of the panels). The figure shows the last part of the standard followed by the beginning of the full cycle.

In the first part of the demonstration, the standard is chosen so that when the full cycle breaks up into streams, the three tones of the standard are left in a single stream (Panel 1, black squares). Therefore this type, called a “within-stream” standard, is easy to hear as a distinct part within the full cycle. In the second part, the standard is chosen so that when the cycle breaks up into streams, two of the notes of the standard would be left in one stream and the third note in the other stream (see Panel 2, black squares). This type, called an “across-stream” standard, is very hard to detect in the six-tone cycle.

Technical details. The three-tone standards are derived from the six-tone sequence by replacing three of the latter's tones with silences. In Part 1, the full six-tone sequence is exactly the same as the fast sequence in Demonstration 1 (a strict alternation of high and low tones) except that the low frequencies are only 1.5 dB louder than the high ones. In Part 2, the across-stream standard, two high tones and a low one, was made isochronous (equally spaced in time), just as it was in the within-stream standard of Panel 1. To accomplish this, the full pattern could no longer involve a strict alternation of high and low tones.

Reading. This pattern was used by Bregman & Campbell (1971) and is discussed in ***ASA-90***, pp. 50, 138-140, 147, 153, and other parts of Ch.2.

3. Loss of rhythmic information as a result of stream segregation.



When a repeating cycle breaks into two streams, the rhythm of the full sequence is lost and replaced by those of the component streams (Panel 1). This change can be heard clearly if the rhythm of the whole sequence is quite different from those of the component streams. In the present example, we use triplets of tones separated by silences, HLH-HLH-HLH-**...** (where H represents a high tone, L a low one, and the hyphen corresponds to a silence equal in duration to a single tone). We perceive this pattern as having a galloping rhythm.

An interesting fact about this pattern is that when it breaks up into high and low streams, neither the high nor the low one has a galloping rhythm. We hear two concurrent streams of sound in each of which the tones are isochronous (equally spaced in time). One of these streams includes only the high tones (i.e., H-H-H-H-H-**...**), joined by dotted lines in Panel 1.

The apparent silences between H tones arise from two sources: Half of them are supplied by the actual silence, labeled S in the figure, that follows the second H tone in the HLH-sequence. The other apparent silences derive from the fact that perceptual organization has removed the low tones from the high-tone stream leaving behind gaps that are experienced as silences. These are labeled by the letter “G” in the figure.

Similarly, the low stream (Panel 1, bottom) is heard as containing only repetitions of the low tone, with three-unit silences between them (i.e., ---L---L---L--- L---...). Again, one of these silences (labeled S) is supplied by the inter-triplet silence of the HLH-HLH-sequence, but the other two (labeled G) are created in our perception to account for the two missing H tones, which have disappeared into their own stream. So stream segregation causes the original triplet rhythm, illustrated in Panel 2, to disappear and be replaced by the two isochronous rhythms of Panel 1, a more rapid high-frequency one

and a slower low-frequency one. The change also affects the perceived melody. When we hear the sequence as a single stream, its melodic pattern is HLH-. This disappears when segregation occurs. The demonstration shows that both rhythms and melodies occur mainly within streams, and when the stream organization changes, so do the perceived melodies and rhythms.

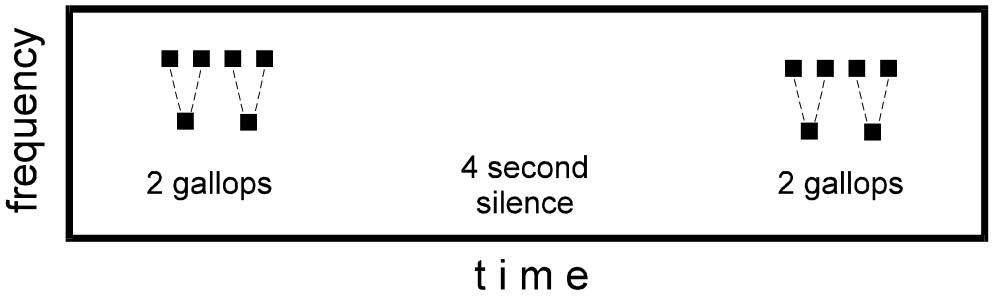
It also shows the importance of speed and frequency separation of sounds in the formation of sub-streams. Segregation is favored both by faster sequences and by larger separations between the frequencies of high and low tones. The role of speed is seen as the sequence gradually speeds up. At slow speeds there is no segregation, but at high speeds there may be, depending on the frequency separation. In the first example, the H and L tones are far apart in frequency (about 18 semitones), as in Panel 1. At the slowest speed, people hear the one-stream percept, but as the sequence accelerates, a point is reached (which may vary from person to person) where stream segregation based on frequency differences inevitably occurs. In the second example (Panel 2), the H and L tones differ by only one semitone and the sequence usually fails to break into two separate streams even at the highest speed.

One can view speed as decreasing the *temporal* separation of the tones that lie in the same frequency range. Tones are believed to group according to their separations on frequency-by-time coordinates. When the speed is slow and the frequencies are close together, each tone's nearest neighbor is the tone of the other frequency; so a single-stream percept, of the type shown in Panel 2, is favored, as indicated by the dotted lines that connect the tones. When a high speed is coupled with a large frequency separation, the combination of these factors places the tones of the same frequency nearer together than they are to the tones of the other frequency. This causes a grouping by frequency to occur, creating the two stream percept of Panel 1.

Technical details. All tones are sinusoidal with rise times of 10 msec and decay times of 20 msec. In the first example, the H and L frequencies are 17.8 semitones apart (1400 and 500 Hz). Eventually, the sequence speeds up from a rate of 287 msec per unit to one of 88 msec per unit, where a unit is either a tone or the silence between successive HLH triplets. The second example is also an accelerating HLH-HLH-**...** rhythm. However, this time the H and L frequencies, 1400 and 1320 Hz, are only a semitone apart. The amplitudes of all tones are equal.

Reading. A galloping pattern was first used by van Noorden (1975), but a later paper (van Noorden, 1977) is more accessible. Effects of frequency separation and speed on grouping are discussed in ***ASA-90***, pp. 17-21, 48-73. For a description of the effects of grouping on perception see ***ASA-90***, pp. 131-172.

4. Cumulative effects of repetition on streaming.



If the auditory system were too responsive to short-term properties of a sequence of sounds, its interpretations would oscillate widely. Therefore it is moderately conservative. Rather than immediately breaking a sequence into streams (i.e., deciding that there is more than one source of sound in the environment) as soon as a few tones have fallen in different frequency ranges, the auditory system waits and only gradually increases the tendency for the streams to segregate as more and more evidence builds up to favor a two-source interpretation.

In the present demonstration, we present cycles of a high-low-high galloping rhythm. We assume that a tendency towards segregation builds up whenever the cycle is playing, and dissipates during silences. Therefore we present the tones in groups of cycles, each separated from the next by a 4-second silence. First we present 2 cycles bracketed by silences, then 4 cycles, then 8, then 16, then 32. The figure shows the 2-cycle (2 gallops) groups separated by 4-second silences. We expect the longer unbroken sequences to segregate more than the shorter ones.

Technical details. The cycle is formed from a 2000-Hz high tone (H) and a 700-Hz low tone (L) presented in a galloping rhythm (HLH-HLH-**...**). Each tone has a 12.5-msec rise in amplitude at the onset, an 88-msec steady state, and a 12.5-msec decay, followed by a 12-msec silence. If the lower tone is the same intensity as the higher one, it tends to be less noticeable. Therefore the lower tone has been made 6 dB more intense than the higher one. The silences (-) between gallops in the HLH-HLH- **...** rhythm are 125 msec long. The groups of cycles are separated by 4-sec silences.

Reading. Cumulative segregation effects are discussed in ***ASA-90***, pp. 128-130, 648.

5. Segregation of a melody from interfering tones.

So far, on this disk, stream segregation has been demonstrated only with repeating cycles of tones. However, one can obtain it without using cycles. In this demonstration, randomly selected distractor tones are interleaved between successive notes of a simple familiar melody. On the first presentation, the distractors are in the same frequency range as the melody's notes. This tends to camouflage the melody.

In the demonstration we present the tone sequence several times. On each repetition we raise the pitch of the melody's notes, but the distractors always remain in the range in which the melody was first played. As the melody moves out of that range, it becomes progressively easy to hear, and its notes are less likely to group with distractors.

This change in grouping can be interpreted in terms of stream formation. In the untransposed presentation, the ASA system puts both the notes and distractors into the same perceptual stream. Therefore they unite to form melodic patterns that disguise the melody. At the highest separation between melody and distractors, they form separate streams. At the intermediate separations, familiarity with the melody can sometimes allow listeners to hear it out, but there is a still a tendency to hear runs of tones that group melodic and distractor notes.

All notes are of the same duration (nominally quarter notes). Those of the melody alternate with distractors. To prepare the melody for this demonstration, any of its notes that were longer than a quarter note were broken into a series of discrete quarter notes. This is done for two reasons: the first purpose was to not give away the melody by its pattern of note durations; the second was to allow for a strict alternation of quarter notes from the melody and the distractor with no variations in the pattern of alternation.

With extended listening, it becomes progressively easier to hear the notes of the melody. This illustrates the power of top-down processes of pattern recognition. The present demonstration, which focuses on bottom-up processes, will work best in the listener's earliest exposures to the demonstration.

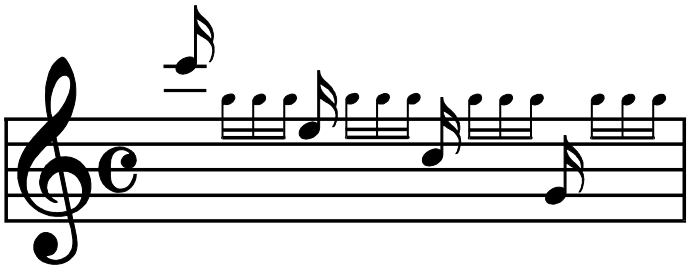
Technical details. The melody's notes begin in the range running from C4 (C in the fourth octave: 256 Hz) up to G4 (G in the fourth octave: 384 Hz).
On each of five repetitions, all the notes of the melody are transposed upward by two semitones relative to the previous one. On the first presentation, each random distractor

note is drawn from a range of plus or minus four semitones relative to the previous note in the original untransposed melody; so the interfering tones track the untransposed melody roughly. However, the range from which they are selected is not transposed when the melody is; so the melody, over repetitions, gradually pulls away in frequency from the distractors. The rate of tones is 7 per sec, or 143 msec/tone. The duration of the steady state of each tone is 97 msec, its rise time is 3 msec and its decay time is 20 msec; a 23-msec silence follows each tone. All tones are of equal intensity.

The identity of the melody is given at the end of the booklet under "Answers to listening tests."

Reading. Dowling (1973) studied the perception of interleaved melodies. This topic is also discussed in ***ASA-90***, pp. 61-64, 140, 462-466.

6. Segregation of high notes from low ones in a sonata by Telemann.



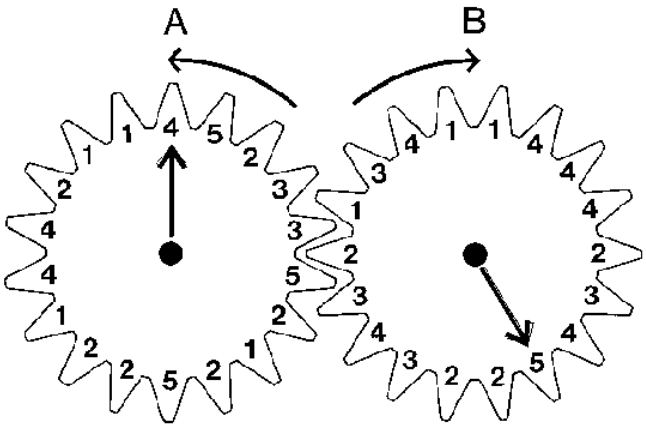
The formation of perceptual streams has played an important role in Western music. Even the “streaming” illusion, in which an alternating sequence of high and low tones gives rise to separate high and low streams, has its counterpart in music, as we show in Demonstrations 6 and 7.
Composers in the Baroque period (approximately 1600-1750) frequently wrote music in which individual instruments rapidly alternated between a high and a low register, giving the effect of two intertwined melodic lines (“virtual polyphony” or “compound melodic line”). While the performances were typically not fast enough to fully segregate the high and low melodic lines, it still produced a certain degree of segregation.

Gert ten Hoopen found and sent us this performance of Telemann's Sonata in C Major, from Der Getreue Musikmeister. It is played more quickly than usual and its rendition on a recorder, rather than on a transverse flute, seems to produce the notes with a greater dryness, allowing us to hear each one distinctly. The style is such that the splitting into sub-streams occurs only rarely and for only a few bars. So the net effect is that certain tonal sequences seem to stand out from their background. The section with the clearest instance begins 27 seconds into track 6. The notes that are printed larger in the excerpt shown above are the ones that we hear as standing out in the music. There is an alternation between these notes and an unvarying three-note sequence of G’s. The fact that the pitch of the latter does not change also helps in the perceptual isolation of the ones that do, since the auditory system tends to focus on changes.
Technical details. This is a short extract from a compact disk entitled “Telemann Recorder Sonatas” (Pavane Records, No. ADW 7135) reproduced with permission from Pavane Records, rue Ravenstein, 17, B-100 Bruxelles, Belgium. Tel: (322) 5130965, Fax: (322) 5142194. The performers are Frédéric de Roos, recorder, Jacques Willemyns, harpsichord, and Philippe Pierlot, bass viol. We found that channel 1 of the original,

which contained less energy from the other instruments and less reverberation than channel 2, provided a crisper presentation, allowing the listener to hear individual notes more clearly; so rather than reproducing the original stereo version, we have recorded channel 1 onto both channels.

Reading. The role of perceptual organization in music is discussed in ***ASA-90***, Ch.5. For the use of compound melodic lines by Baroque composers, see ***ASA-90***, p.464.

7. Streaming in African xylophone music.



An even stronger example of streaming in music than the Telemann excerpt of Demonstration 6 can be found in the traditional music of East Africa. One style makes use of repeating cycles of notes. An example is the piece, “SSematimba ne Kikwabanga”, from Buganda (a region of Uganda). In the style exemplified by this piece, each of two players plays a repeating cycle of notes, the notes of each player interleaved with those of the other.

The cycle played by each player is isochronous (notes equally spaced in time), but the interleaving of the two cycles creates high and low perceptual streams with irregular rhythms. The figure, kindly provided by Dr. Ulrich Wegner, represents the two cycles of notes as interlocking cogwheels, moving in contrary directions, labeled with numbers representing the pitches of notes (in steps on a pentatonic scale) in the cycle. The sequence of resulting notes is represented by the series of numbers appearing at the point of intersection of the wheels (e.g., **...** 1 3 2 5 3 2 4 1 3 **...**). The players do not start at the same time; the arrows in the figure point to the starting note of each player. This instrumental style is typical of music for the amadinda (a twelve-tone xylophone). Similar interlocking techniques are used in East and Central African music for harp, lyre, lamellophone, and other instruments.

Dr. Wegner wrote this comment on the music:

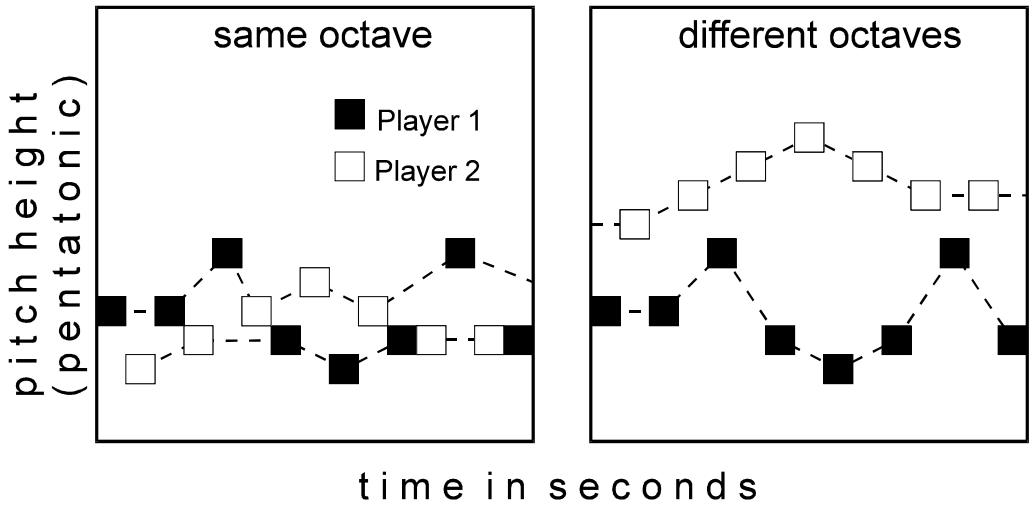
> “What is striking about most of these instrumental traditions is the difference between what is played and what is heard. While the musicians, of course, know about the constituent parts of a composition, the listener's perception is guided to a great extent by the auditory streaming effect. What aids the emergence of auditory streams is the fact that with the basic playing technique, an absolute equilibrium [equalization] of all musical parameters except pitch is intended. Synthesized amadinda music played by a computer, with machine-like precision, received highest ratings by musicians from Uganda.” (Personal communication, April, 1991)

In addition, it seems that the Ugandan culture dictates that the listener be able to hear a sequence, the "nuclear theme" that is only partly contained in one of the emerging streams. Therefore, the listener must supplement the cues given by the stream organization with knowledge of the nuclear theme. This is similar to how westerners listen to the types of jazz in which there is a basic melody that is ornamented and varied to the point that it can be recognized only by listeners who are very familiar with it.
In this demonstration, we have a chance to hear the individual parts as well as their combination. First we hear one player's part (the first cogwheel in the figure). Then it is joined by the second part (the second cogwheel), causing the melody and isochronous rhythm of the first part to be lost perceptually. Then the *second* part is played alone. Finally when both are played together, the melody and isochronous rhythm of the second part are lost.

Technical details. This example was originally synthesized by Dr. Ulrich Wegner from digital samples of the sound of the amadinda. It is an extract from a digital audio tape (DAT) generously provided to us, with an accompanying booklet, by Dr. Wegner. The tape was then sampled into a computer in our laboratory at 22255 samples per second. Each part is played at a rate of 4.2 tones per second; so the combined rate, when both are present, is 8.4 tones per second, which is in the range in which stream segregation is obtainable. The sounds are very percussive, having a roughly exponential decay with a half-life ranging between 15 and 30 msec. Their peak intensities are all within a range of 5 dB.

Reading. The involvement of principles of perceptual grouping in this musical style is described by Wegner (1993). Also, a cassette tape, with an explanatory booklet, was published by Wegner (1990). ***ASA-90***, Ch.5, presents a discussion of the effects of perceptual organization in music.

8. Effects of a difference between pitch range of the two parts in African xylophone music.



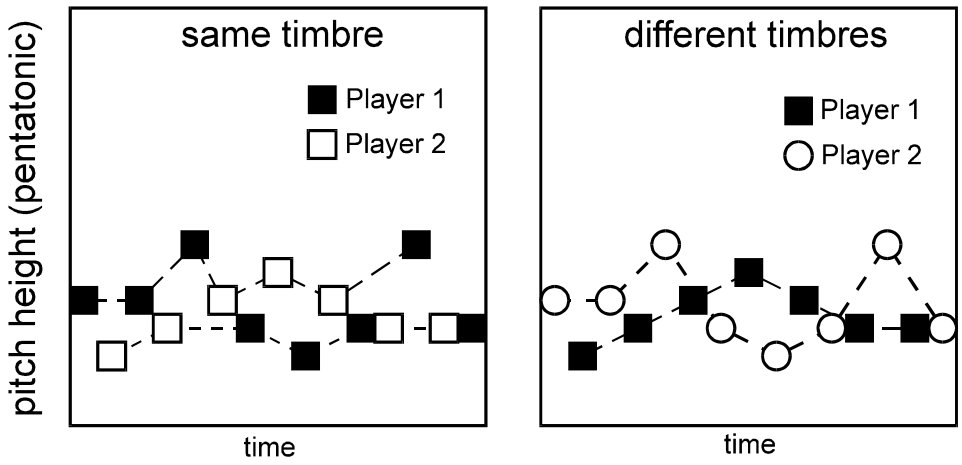
The phenomenon of segregation due to difference in pitch range has been illustrated by Ulrich Wegner, who made a synthesized variation of the same Bugandan piece "SSematimba ne Kikwabanga", that was presented in Demonstration 7. Again the synthesis was done with recorded amadinda sounds. The notes of one of the two interlocking parts were shifted up by one octave to illustrate the effects of pitch differences on stream segregation.

We can appreciate how strongly pitch differences affect the formation of melodic lines in music by contrasting this demonstration with Demonstration 7. In both cases, we can hear what happens when a second part joins a first, but in 7, it becomes impossible to hear the isochronous rhythm of the first player (Panel 1). In the present demonstration, however, ASA segregates the notes of the two players because their parts are an octave apart (Panel 2); so the isochronous parts of the individual players continue to be heard when they play together.

Technical details. Each part is played at a rate of 4.1 tones per second, so that the combined rate, when both are present, is 8.2 tones per second, which is in the range where stream segregation is obtainable. The range of peak amplitudes is 6 dB. The other technical details are the same as for Demonstration 7.

Reading. See Demonstration 7. The effects of frequency differences in the formation of musical lines are also discussed in ***ASA-90***, pp. 461-471, 474-478.

9. Effects of a timbre difference between the two parts in African xylophone music.



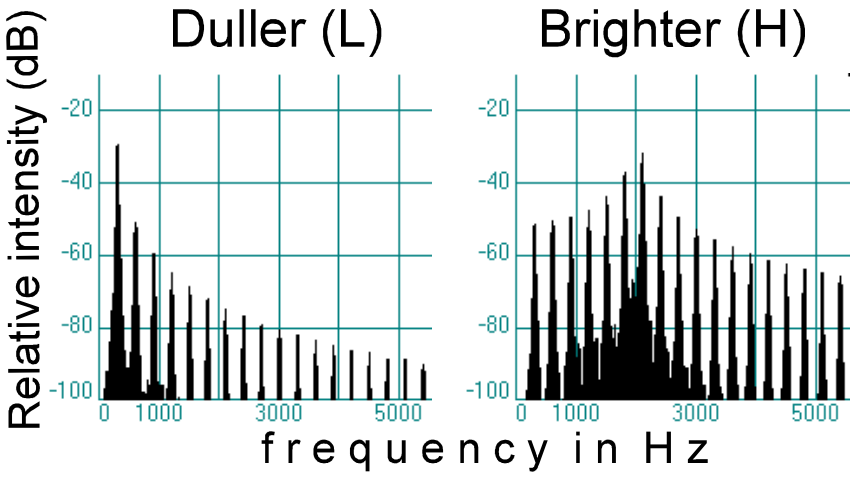
Segregation by timbre is illustrated in another synthesized variation of the Bugandan piece, "SSematimba ne Kikwabanga", of Demonstration 7. First we hear one part played with a muffled timbre, containing a preponderance of energy in the first and second harmonics. Then this is joined by the second part played with a "brittle" timbre that contains a lot of high-frequency inharmonic energy and has a weaker pitch than the first part. The difference in timbre causes the two parts to segregate.

Again contrast the case in Demonstration 7, where each stream contains notes from both players (shown here in Panel 1), with the present case in which timbre differences create two streams, each of which has the notes of only one player (Panel 2).

Technical details. Each part is played at a rate of 4.2 tones per second, so the combined rate, when both are present, is 8.4 tones per second, which is in the range in which stream segregation is obtainable. The range of peak amplitudes is 11 dB.

Reading. See Demonstration 7. ***ASA-90*** discusses the role of timbre differences in perceptual organization in general on pp. 92-126, 646, and its role in the perception of music on pp. 478-490.

10. Stream segregation based on spectral peak position.

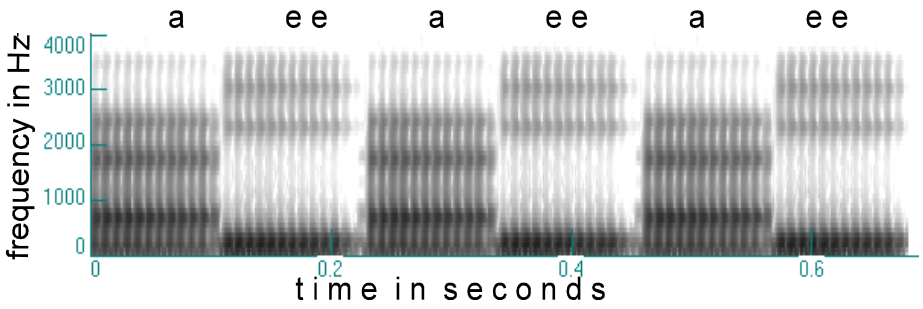


One way to study how timbre differences promote segregation is to manipulate the positions of peaks in the spectra of complex tones. The present demonstration uses two tones with the same fundamental (300 Hz) but different positions of spectral peaks. Panel 1 shows the spectrum of the duller tone (L) with a single spectral peak at 300 Hz, i.e., at its fundamental. Panel 2 shows the brighter tone (H) with its spectral peak at 2000 Hz. The two tones are alternated in an galloping pattern (LHL-LHL-**...**) which gradually speeds up. Even though both tones have the same fundamental frequency, as the sequence speeds up the brighter and duller tones segregate into separate streams.

Technical details. The digital synthesis of each tone begins with a 300-Hz tone with 30 harmonics of equal intensity. This is passed through a formant filter with a bandwidth of 80 Hz, and a peak frequency of either 2000 Hz (H) or 300 Hz (L). The duration of each tone is 65 msec, including rise and decay times of 20 msec. These H and L tones, as well as a 65 msec silence (-), are presented as a HLH-HLH-**...** galloping pattern. The gallop starts off slowly with 140 msec silences between successive sounds (not counting the 65-msec silence that replaces every second L tone to create the galloping sequence). This gives a tone onset-to-onset time of 205 msec, or a rate of about five tones per second. Gradually the extra silence is decreased to 10 msec, giving an onset-to-onset time of 75 msec, or a rate of about 13 tones per second. To equate for salience, the peak amplitudes of the dull tones are 3 dB greater than those of the bright tones.

Reading. The role of timbre in stream segregation is discussed in ***ASA-90***, pp. 478-490.

11. Stream segregation of vowels and diphthongs.



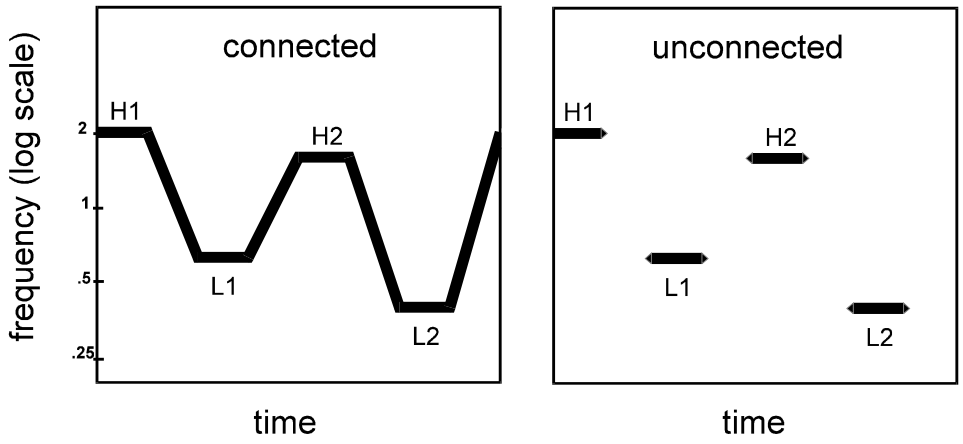
This demonstration shows that when pairs of isolated short vowels and diphthongs are presented in a repeating cycle, stream segregation will often occur because of differences in their spectra. We use the diphthongs “ay” as in the word “hay” and “o” as in “hoe”. The pure vowels are “a” as in “had”, “ee” as in “he”, and “aw” as in “haw”. You will hear 32 rapid alternations of pairs of these. Cycles of each pair of vowels are separated from those of the next pair by a silence. The pairs (in turn) are: ay-o, a-ee, and aw-ee. The spectrogram of a cycle in which “a” and “ee” are alternating is shown in the figure.

First the vowels are all presented with steady pitches; then we hear them with a more natural pitch contour. Stream segregation can be obtained with both types of signals. The ones with the pitch change sound more natural (as natural as such fast speech sounds can), but this does not prevent them from streaming.
The fact that speech need not sound unspeechlike before it can be made to form streams suggests that auditory scene analysis applies as much to speech as to other types of sounds.

Technical details. The syntheses use the formant frequencies measured by Peterson and Barney (1952). Natural-sounding amplitude envelopes were applied to all the signals (a 20-msec rise, an 80-msec decline, and finally an exponential decay with a time constant of 10 msec). In those with steady pitches, the fundamental is 112 Hz. In those with “natural” pitch contours, the fundamental starts at 118 Hz, then drops to 92 Hz. All intensities are within a 3-dB range.

Reading. The role of ASA in speech perception is discussed in ***ASA-90***, Ch.6. Vowel alternation is discussed on pp. 536-537.

12. Effects of connectedness on segregation.



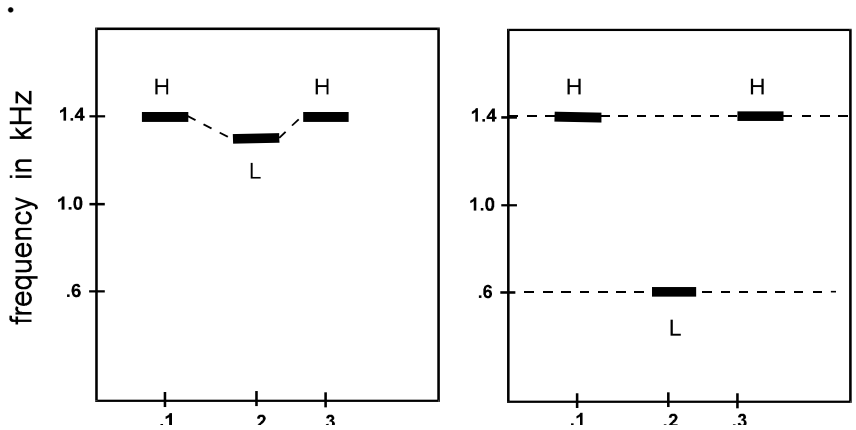
A smooth continuous change helps the ear track a rapid frequency transition. The present demonstration, based on an experiment by Bregman and Dannenbring (1973), uses two high frequency tones, (H1 and H2), and two low ones (L1 and L2). They are presented in a cycle in which high and low tones alternate (i.e., H1 L1 H2 L2 H1 L1 H2 L2,**...**). In the figure, Panel 1 shows successive tones connected by frequency transitions and the right-hand panel shows the same tones not connected. When the transitions are present, they tend to prevent the sequence from segregating into frequency-based streams.
We can more easily demonstrate the effects of continuity by using a frequency separation and speed at which the influences that favor the one- and two-stream percepts are almost equal. This allows continuity to tip the balance in favor of integration. Both the connected and unconnected sequences are presented many times. In both cases, the tendency to hear streaming builds up over repetitions, but this tendency is stronger in the unconnected sequence.

In the first sequence, consecutive tones are connected by frequency transitions. In the second, the frequency transitions are omitted. We perceive the first sequence, with the transitions, as more coherent. This demonstration shows that continuity helps hold auditory sequences together. A related phenomenon was demonstrated by Cole and Scott (1973), who either left in, or spliced out, the normal formant transitions in a syllable such as “sa”, then made a tape loop out of the syllable and played it repeatedly with no pauses. The intact syllable tended to be heard as a unit even after many repetitions, but the one without the formant transitions quickly segregated into consonant and vowel parts which were heard as separate streams. Notice that there is a difference in the acoustics of the Cole-Scott and the Bregman-Dannenbring signals. In the former case, the transitions were formants, while in the latter, they were pure-tone glides.

Technical details. The frequencies, in order, are 2000, 614, 1600, and 400 Hz. In the connected condition, the steady states and the transitions are each 100 msec long. In the unconnected condition, the frequencies of the high and low 100-msec tones are steady, but 10-msec linear amplitude rises and decays are added, keeping frequency constant. The remaining 80-msec part of the intertone interval is silent. The overall tone rate is the same in both conditions. The frequency transitions are exponential (linear in log frequency). The intensities are the same for the steady-state parts and the pitch glides, and are the same for all frequencies. In each condition, the four-tone sequence repeats 20 times.

Reading. The Bregman-Dannenbring and Cole-Scott research, as well as the general issue of continuity in sequential integration, are discussed in ***ASA-90***, Ch.2, pp. 133-136, and Ch.4, pp. 416-441.

12 The effects of stream segregation on the judgment of timing



When segregation occurs it becomes difficult to make accurate timing judgments that relate sounds from different streams. We saw, in Demonstration 3, that the galloping rhythm was lost when the high and low tones formed separate streams. Now we present a further example of this ungluing of temporal relations.

A high tone (H), at 1400 Hz, is played repetitively throughout the demonstration. Its continuous presence induces a tendency to form a stream that excludes frequencies that differ too much from 1400 Hz. When a lower tone (L) is brought in, the H and L tones may give rise to a galloping rhythm (HLH-HLH-**...**) or they may not, depending on the frequency of L.
In the present demonstration, the L tone may or may not be exactly halfway in time between successive occurrences of H. The figure shows a single galloping pattern from the sequence. In Panel 1, H and L are close in frequency and L is delayed relative to the temporal midpoint. In Panel 2, L is delayed by the same amount, but is much lower than H in frequency. It is interesting to note that even in the visual diagram, the delay of L is not as noticeable in Panel 2 as in Panel 1.

The first two times that L is brought in (first two sets of cycles on the disk), its frequency is 1350 Hz, only 4 percent below that of H. At such a small separation, there is only a single stream; so any judgment of the temporal relation between H and L is a within-stream judgment. In the first set of presentations, L is brought in for ten ABA cycles. It is placed exactly halfway between the repetitions of H. In this case, because the comparison is within-stream, it is easy to hear that the placement of L is symmetrical. In the subsequent set of ten cycles, L is delayed relative to the midway point between H's. Again, because H and L are in the same stream, the asymmetrical placement of L is easily detected.

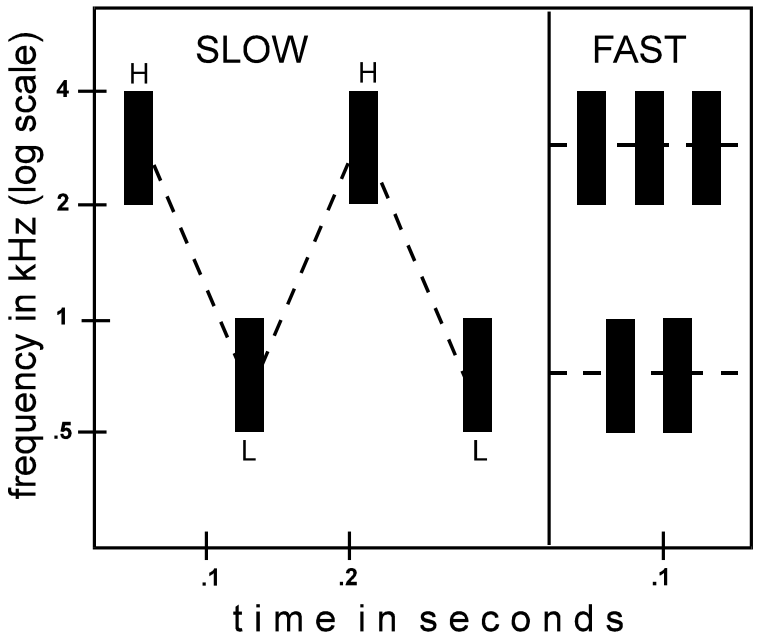
The next two times that L is brought in, its frequency is 600 Hz, more than an octave below that of H. This large separation causes two perceptual streams to emerge, H-H-H-**...**, and -L---L--**...**. Because H and L are in separate streams, it is harder to hear whether L is symmetrically placed between pairs of H's (in the third set of presentations, it is advanced relative to the midway point and in fourth set, it is at the midway point).

This demonstration, though it uses a different stimulus pattern, is based on experiments by van Noorden, in which he alternated a pair of sounds. In one experiment he used a strict HLHL alternation and in another, a galloping rhythm. In both cases, he gradually delayed or advanced the L stimulus until the listener began to hear the asymmetry. He also varied the speed of the sequence and the difference in frequency between H and L. Both speed and frequency affected the listeners' sensitivity to the asymmetry. The more strongly they favored the segregation of H and L tones, the harder is was to detect the temporal asymmetry of L's placement.

Technical details. H and L are pure tones of the same intensity, 60 msec in duration including 20 msec rise and decay times. There are 180-msec gaps between the end of one H and the beginning of the next. The 60 msec L tone is either placed in the center of this gap (60 msec on each side) or is displaced relative to the midpoint (either HL = 30 msec and LH = 90 msec or the reverse). All tones are of the same intensity.

Reading. See van Noorden 1975, pp. 65-67, and ***ASA-90***, pp. 157-163.

14. Stream segregation of high and low bands of noise.

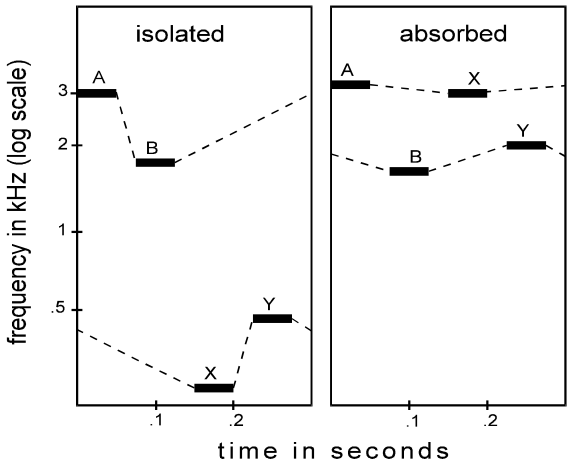


This demonstration shows that narrow bands of noise, centered at different frequencies, will segregate in the same way that tones do. The examples are related to an experiment by Dannenbring and Bregman (1976), but here we use short noise bursts with sharper band edges. The noises can be considered as high (H) and low (L), and are presented in a galloping pattern, HLH-HLH-.... The gallop is speeded up until the noise bands segregate into separate streams.

Technical details. The noise bursts are 50 msec in duration with 5-msec rise and decay times. They were synthesized by adding equal-intensity pure tones in random phases, spaced from each other by a fixed frequency-step size, covering the range from the lower band edge to the upper band edge. The result is a very flat noisy spectrum with very sharp band edges. The lower-frequency noise burst extends from 500 to 1000 Hz., and the higher one from 2000 to 4000 Hz. Therefore each band spans an octave and there is an octave separation between them. The spectral step size, used in the synthesis, was 2 Hz for the lower band and 3 Hz for the higher band. The spectrum level of both bursts is the same. The gallop starts with silences of 140 msec between successive elements, and accelerates over 52 cycles until there are only 10 msec between them. Here the word "elements" refers either to a noise burst or to the 50-msec silent gap between HLH triplets.

Reading. See ***ASA-90***, pp. 91-92.

15. Competition of frequency separations in the control of grouping.



Frequency separation is known to affect the segregation of sounds into streams. The present demonstration shows that relative frequency separation, as opposed to absolute separation, can play an important role in controlling the grouping. The figure shows one cycle of a repeating 4-tone pattern, ABXY**...**. In the case diagrammed in Panel 1, the first two tones, A and B, are high in frequency, whereas X and Y are much lower. This pattern breaks up into two streams, a high one, AB--AB--**...**, and a low one, --XY--XY**...** (where the dashes represent within-stream silences). This placement of X and Y is called "isolating", because it isolates A and B from X and Y. In Panel 2, A and B have the same frequencies as in Panel 1, but X is close in frequency to A and Y to B. This causes different streams to emerge than in the previous case: A-X-A-X-**...**, and -B-Y-B-Y**...**. This placement of X and Y is called "absorbing", because A and B are absorbed into separate streams.

Notice that A and B are in the same time slots (first two of the cycle) and have the same frequencies in the two cases. If grouping were determined by the raw separation in frequency or in time, A and B should have been either in the same stream in both examples, or in different streams in both. The fact that they can either be in the same stream or in different streams, depending on the context, exposes an inadequacy in theories that propose that the segregation of streams is caused by some limit on the rate at which a stream-forming process can shift to a new frequency (e.g., from A to B), since the grouping of A and B can vary even when the separation between them, in frequency and time, is held constant.

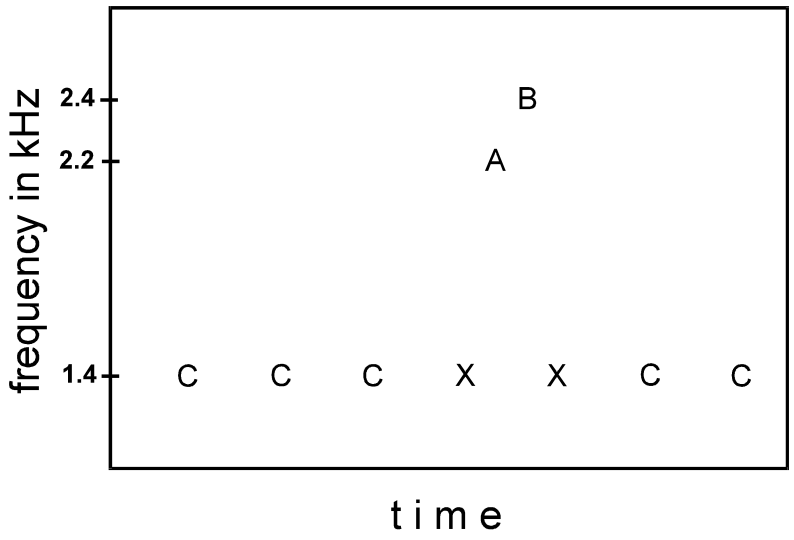
In the demonstration, several cycles of a two-tone "standard" AB--AB--**...**, (A, B, and two silences) are presented before cycles of the full four-tone pattern, A,B,X,Y,A,B,X,Y,**...**. The task of the listeners is to listen for the standard within the 4-tone test pattern, which can be either the pattern shown in Panel 1 or the one in Panel 2. If listeners can easily

detect the standard, it means that ASA has grouped the elements of the four-tone pattern so as to put A and B into the same stream. This is true in the first example (Panel 1) but not in the second (Panel 2).

Technical details. All tones are pure sinusoids, 45 msec in length, including 10 msec rise and decay times. There are 30 msec silences between successive tones. Hence each four-tone cycle takes 300 msec. The frequencies of A and B are always at 3400 and 1525 Hz. In the isolating pattern of Panel 1, X and Y are at 246 and 455 Hz, and in the absorbing pattern of Panel 2, they are at 2760 and 1860 Hz. Each presentation is 30 cycles long, including a three-cycle fade-in at the beginning and a ten-cycle fade-out at the end. The tones are approximately equal in intensity.

Reading. This demonstration is patterned on an experiment by Bregman (1978) which is discussed in ***ASA-90***, pp. 167-168. The effects of competition on perceptual grouping are discussed in ***ASA-90***, pp. 165-171, 434, 218, 296, 303-311, 335, 651.

16. The release of a two-tone target by the capturing of interfering tones.



An experiment by Bregman and Rudnicky (1975) asked whether streams are created by attention or by a pre-attentive mechanism. The strategy of the experiment was to cause a grouping mechanism to capture material away from attention, thereby showing that the grouping mechanism was not part of attention.

This demonstration is very similar. The figure shows a number of tones, symbolized as letters. Those called A and B are to be judged for their order – ascending or descending in pitch. The order shown in the figure is ascending (AB), but in the Bregman-Rudnicky experiment, it could be either ascending (AB) or descending (BA).

When the AB or BA pair is played in isolation, it is easy to judge the order of A and B. This is illustrated in the first part of the demonstration which plays the pair first in the order BA, as a standard, then in the order AB (repeating this comparison twice).
In the second part of the demonstration, we make this order discrimination very difficult by surrounding the AB or BA pair by two instances of tone X (1460 Hz), to generate the four-tone sequences XABX, or XBAX. Despite the fact that we are being asked to discriminate the same AB pair in the two- and four-tone sequences, the addition of the two bracketing tones makes the task very difficult. This is probably because the four tones form a higher-order unit in which the A and B tones lose their prominence, in favor of the X tones, which fall at the perceptually prominent beginning and end positions. In the second part of the demonstration, a two-tone standard (AB) is followed by a four-tone

test sequence XABX. Although AB is in the same order in the two- and four-tone sequences, it is very hard to judge that this is so.

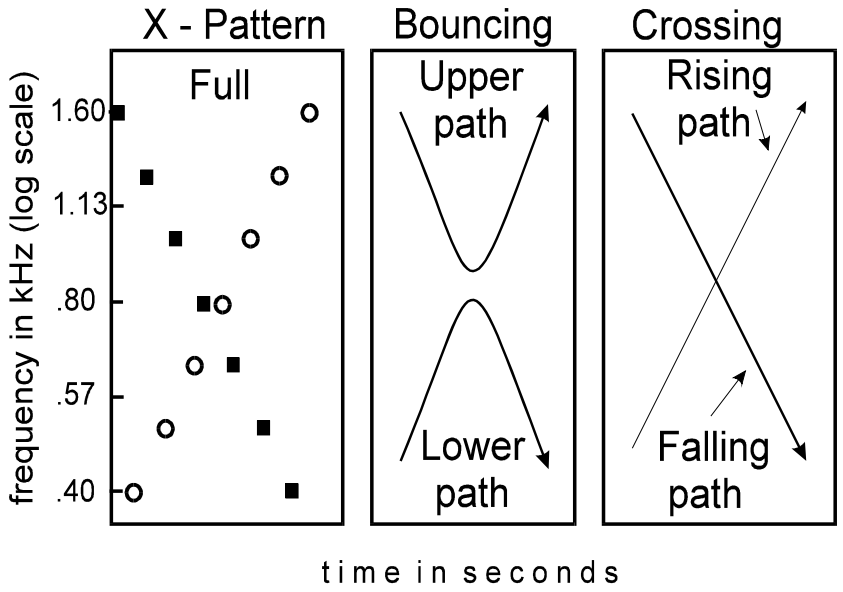
In the third part, we show that, paradoxically, we can restore some of the original salience of A and B by embedding the four-tone XABX (or XBAX) sequence in an even longer one, in which there are a sequence of C (captor) tones, which fall at the same frequency as the X tones, and have the same intertone spacing as the X tones, C--C--C--C--XABX--C--C (in which each "-" represents a silence of the same duration as A or B). The X's are captured into an isochronous single-frequency stream with the C's, rejecting A and B from this stream. This partially releases A and B from interference by the X's. Accordingly, in the final part of the demonstration, when a two-tone standard, BA, is followed by the full ten-tone sequence that contains the C's, it is fairly easy to tell that the AB order is different in the long sequence.

Like Demonstration 15, this one illustrates the fact that there is a competition in the grouping of tones, so that adding more tones to a sequence can change the perceptual organization. In the present demonstration, as in 15, this change in grouping changes the isolation of some of the tones, and hence their perceptibility.

Technical details. The tones were sinusoidal, 57 msec in duration, including 7-msec rise and 5-msec decay times, with 9-msec silences between tones. To make the tones appear equally salient, the amplitude of the 1460-Hz tone was multiplied by .4, and that of the 2400-Hz tone by .5, relative to that of the 2200-Hz tone.

Reading. The experiment by Bregman & Rudnicky (1975) is discussed in ***ASA-90***, pp. 14, 132-3, 140, 165, 169, 192, 444-5, 450, 475. The more general effects of competition on perceptual grouping are discussed in ***ASA-90***, pp. 165-171, 434, 218, 296, 303-311, 335, 651.

17. Failure of crossing trajectories to cross perceptually.



The stimulus in this example is shown in Panel 1 of the figure. It consists of a falling sequence of tones, shown as squares, interleaved with a rising sequence, shown as circles. Tones from the rising and falling trajectories are strictly alternated in time. The full pattern is referred to as an "X pattern", because of the appearance of the diagram. In crossing patterns, such as this one, listeners can rarely follow the entire rising or falling sequence. Instead, they hear the tones that lie on one of the paths shown in Panel 2, either the ones falling on the "upright V" path shown in the upper half in the panel, or on the "inverted V" path shown in the lower half. For example, when listeners follow the upright V pattern, they track the descending sequence to the crossing point and then shift over to the ascending sequence and follow it upward in frequency. This is called a "bouncing" percept, because the subjective stream seems to bounce away from the crossing point. It is contrasted with a "crossing" percept in which one sequence (the rising or falling one) is followed right through the crossover point. The two possible full-trajectory patterns, rising and falling, are shown in Panel 3.

We can determine how listeners organize the pattern by asking them how easily they can hear four different types of standards, each consisting of a subset of tones, as a distinct part of the full pattern. Each type of standard consists of a sequence of tones that follows one of the four possible paths shown in Panels 2 and 3. We assume that whenever a standard corresponds to one of the subjective streams formed by the listeners as they listen to the full "X" pattern, they will judge that this standard is easier to hear in the full pattern than the other types of standard are.

The demonstration begins with three cycles of the whole X pattern, presented without standards so that you can get a general impression of the grouping. Next you are asked whether you can hear a standard pattern as a part of the full one. First, three cycles of the "upper V" pattern are presented as a standard, followed by 3 cycles of the full X pattern, so that you can listen for that standard inside it. Then a second test is given using the lower, "inverted V" pattern. These two standards are easy to hear as part of the X. In a third test, the full rising trajectory is used as a standard, and finally, in a fourth, the full falling standard is used. These two standards are very hard to discern in X.

Apparently the tendency for bottom-up ASA to follow a trajectory, and treat its components as a distinct stream, is weak or non-existent, and the tendency to perceive streams that remain in a given frequency region is much stronger. The situation can be altered by enriching the timbre of the tones on the rising trajectory (circles in Panel 1). Now the tones of the two full trajectories are differentiated by timbre; so segregation by timbre starts to play a role. Accordingly when the comparisons of the four standards with the full X are presented again, the complete rising and falling trajectories are heard much more easily in the X pattern.

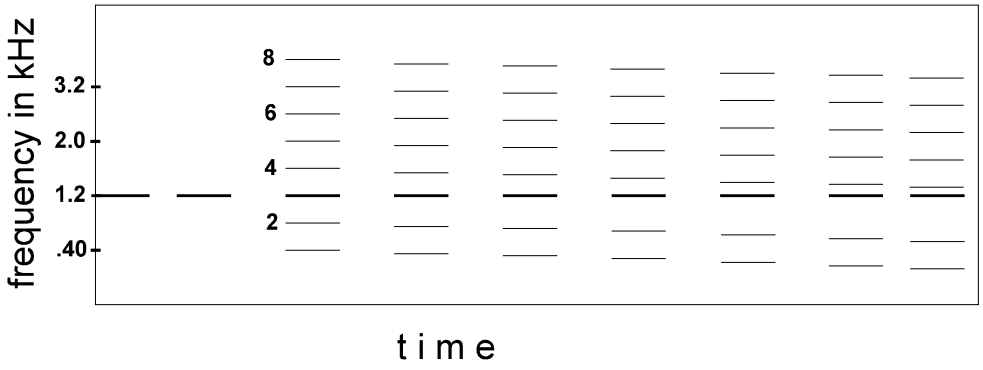
Technical details. The rising tone sequence starts at 400 and rises to 1600 Hz in equal log-frequency steps (7 tones). The falling sequence does the reverse. Tones are 100 msec in duration including 10-msec rise and decay times. There are no silences between them. The "pure" tones are sinusoidal; the "rich" ones contain harmonics 1, 2, 4, and 8, each of them 20 dB softer than the single component of the pure tone. The resulting waveform of the rich tones has an amplitude about a quarter that of the pure tones. Accordingly they are softer; so differences in loudness as well as in timbre may be contributing to the segregation of the streams.

Reading. Crossing tonal patterns were studied by Tougas & Bregman (1985). For a discussion of the perception of crossing tone patterns and, more generally, of how the auditory system deals with trajectories, see ***ASA-90***, pp. 417-442.

Spectral integration

Demonstrations 18 to 24 show how components that occur at the same time are either fused or segregated. When fused, they are heard as a single sound, and when segregated as two or more sounds. Different subsets of components can be fused into separate sounds. The acoustic relations that favor fusion are those that are likely to exist between components that have been caused by the same sound-producing event. The relations that we illustrate in demonstrations 18 to 24 are the sharing of a common fundamental frequency, or having parallel changes in either pitch or amplitude.

18. Isolation of a frequency component based on mistuning.



A test used by the auditory system to decide whether a set of frequency components come from the same source is whether they are all multiples of a common fundamental frequency. If they are, and assuming that none is a great deal more intense than the others, the system integrates them and hears them as a single rich sound. We can call this "grouping by harmonicity". On the other hand, if one partial is not related to the otherwise shared fundamental, i.e., is mistuned, it will be heard as a separate sound.

In this demonstration, the third harmonic is the component that we mistune. It begins at the "correct" or "tuned" frequency, three times that of the fundamental (3f). We could have gradually shifted its frequency away from 3f (say upwards) in small steps, and you would have eventually heard it as an isolated tone. However, this would not have been the right way to demonstrate grouping by harmonicity because the shifting tone would have been isolated not only by the fact that it was being gradually mistuned, but also by the mere fact that its frequency was changing on successive repetitions. The auditory system is extremely sensitive to any change in a spectrum and focuses on the changed components. We eliminate this problem by holding the to-be-isolated component (the "target") at a constant frequency while shifting all the other components (the "harmonic frame"). Because the target remains constant in frequency, it does not engage the "difference-noticing" mechanism of the auditory system. Only the harmonic frame does so, and this attracts attention away from the partial that is to be "heard out". Therefore any increase in audibility of this partial, as the harmonic frame changes in frequency, is due to the mistuning itself.

A schematic of the first series of sounds is shown in the figure; the amounts of change of the harmonic frame are exaggerated in order to be easily visible, and the durations of silences are not shown literally. Harmonics 2, 4, 6, and 8 are labeled. The target component (third harmonic) is drawn heavier for visibility, but, in the audio signal, it is

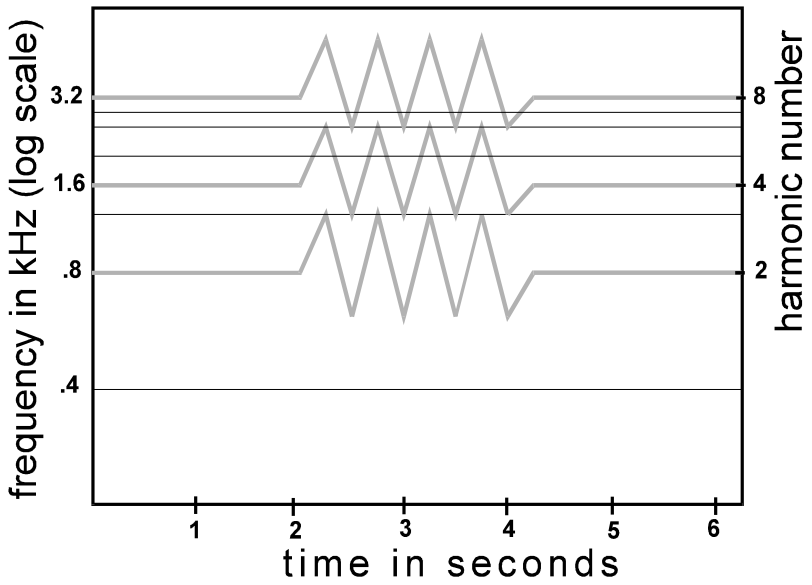
not louder than the other components. In the first series (“descending”), the target component is first played alone twice to familiarize you with it. Then the rich tone, containing the entire set of harmonics, is played. On the first presentation, all its components, including the third harmonic, are correct multiples of the fundamental. Then, on each subsequent presentation, each component, except the target, drops by one percent, but the target maintains the same frequency. After a few presentations, you can start to hear a thin pure tone, with a constant frequency, accompanying the complex tone. There are 15 steps, each with a one percent drop. By counting the number needed to achieve the segregation, you can estimate the amount of mistuning (in percent) required for perceptual isolation of the target.

Next, a second series begins with the target played alone twice. Then, beginning at the previous ending point, 15 percent below the untransposed value, the frame rises by one percent in frequency on each successive presentation until the target and the frame are back in tune. The separate perception of the target disappears during the last few steps.

Technical details. The complex tone's spectrum includes the first eight harmonics at equal amplitude, in sine phase. Its duration is 250 msec, including 10 msec rise and 9 msec decay times. There is one second of silence between successive presentations of the complex tone.

Reading. This demonstration is related to an experiment by Moore, Glasberg, & Peters (1986). The effect of harmonic relations on the integration of simultaneous components is discussed in ***ASA-90***, pp. 223, 232-248, 508, 570, 624, 656.

19. Fusion by common frequency change: Illustration 1.



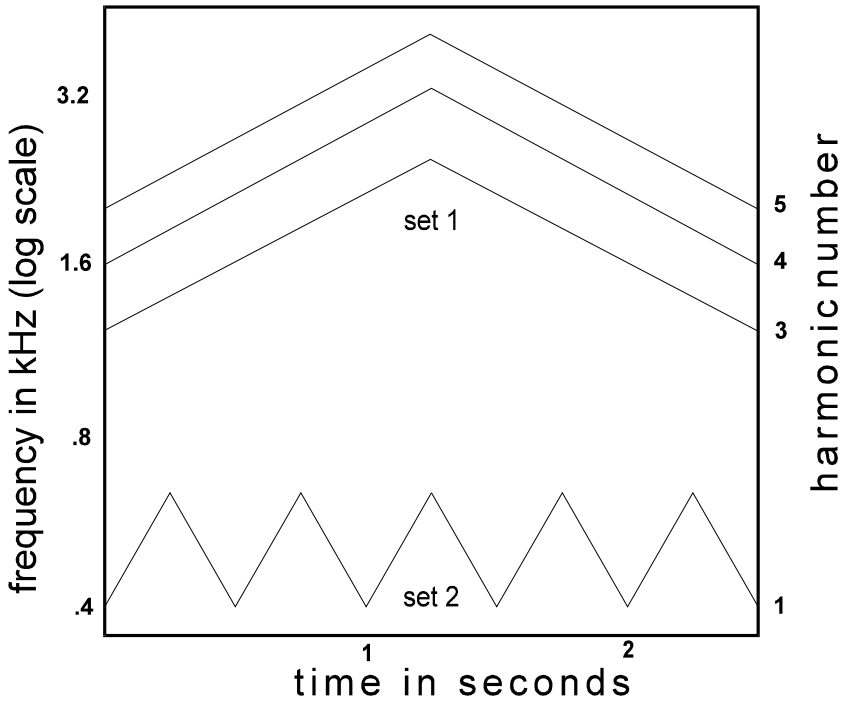
This is an example of the grouping of the frequency components of a complex tone as a result of parallel frequency changes. The components are arbitrarily divided into two subsets. Then the same modulation is applied to all the members of one subset, while the other subset remains steady, as shown in the figure. At first all the harmonics are played with steady frequencies and we hear a unified tone. Then, while harmonics 1, 3, 5, 6, and 7 remain steady, harmonics 2, 4, and 8 rise and fall four times. While this happens, the two sets are heard as separate sounds. Finally, when the partials come together to form a single steady harmonic series, they are heard again as a single tone. This pattern is played twice with a brief pause between repetitions. You may notice a timbre change in the steady tone when it loses some of its harmonics, which have become part of the rising and falling tone.

During the rise and fall of harmonics 2, 4, and 8, they maintain harmonic relations among themselves (ratio 2:4:8), because the frequency change is proportional for all of them. Similarly, the steady components stay in the ratio 1:3:5:6:7. However, when the frequency change begins, the harmonic relationship between the steady and changing sets of partials is broken. This is one reason why they segregate. There is also reason to believe that the fact that the two sets are undergoing independent (non-parallel) frequency changes contributes to their segregation. If this is correct, this example provides an instance of "grouping by common fate", a concept of the Gestalt psychologists.
Technical details. All harmonics are of equal intensity and start off in sine phase. The fundamental is 400 Hz for all harmonics in the steady-state stages, and for the unchanging harmonics throughout. The rise and decay times are 40 msec for the entire sound (which lasts for 6.2 sec). The initial unchanging period lasts 2 sec. The subsequent

rises and falls in frequency each last for 0.25 sec, and consist of linear changes in the nominal fundamental up to 600 and down to 300 Hz, four times, with a final return to 400 Hz.

Reading. The readings for the issue of grouping by harmonicity are given in Demonstration 18. There is a discussion of grouping by common fate in McAdams (1984), Chalikia & Bregman (1993), and in ***ASA-90***, pp. 248-289, 657-658.

20. Fusion by common frequency change: Illustration 2.



Like Demonstration 19, this one also illustrates segregation by "common fate", but we show that it is not necessary that one set of partials remain steady. Both sets of partials in the present demonstration are in motion. In the first two presentations, all partials follow the same up-and-down pattern of frequency change, first a faster change, then a slower one. In each case we hear a single tone changing in pitch, either quickly or slowly. Finally, in the third presentation, shown in the figure, the two subsets undergo different patterns of change: a faster and a slower one. This is played twice. We hear two distinct tones, each formed from one of the subsets. The slow-moving one sounds pure and the fast-moving one, rich.

Technical details. The rise and decay times are 40 msec for the entire sound (which lasts for 2.5 sec). There are two sets of partials. Set 1 consists of three components that can be thought of as harmonics 3, 4, and 5 of a fundamental, in that they always maintain the ratio 3:4:5. Set 2 consists of a single partial that can be thought of as the first harmonic. All partials are of equal intensity and start off in sine phase.
There are two patterns of frequency change, as shown in the figure. In the slow-moving one, the fundamental (i.e., the tone itself) rises from 400 to 800 Hz linearly in 1.25 sec, then falls back to 400 Hz linearly for 1.25 sec. In the fast-changing pattern, the (missing)

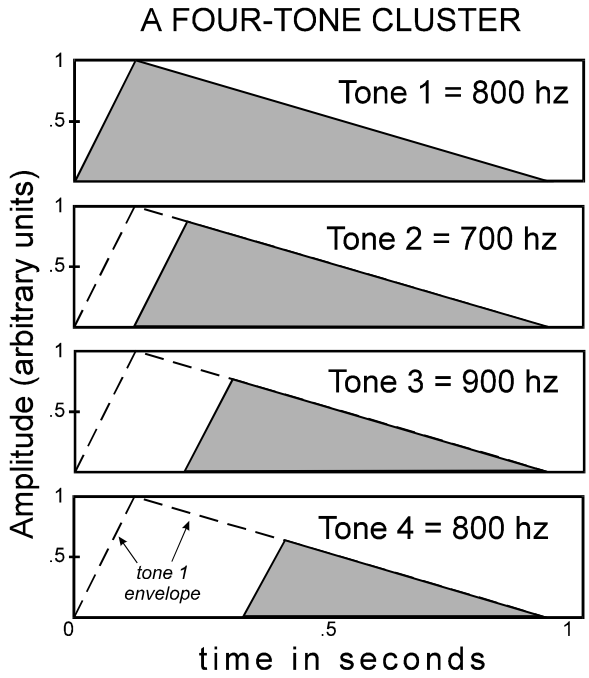
fundamental rises and falls five times, each rise and each fall lasting for 0.25 sec, and consisting of a linear change in frequency between 400 and 660 Hz. When the two sets follow the same modulation pattern, either slow or fast, they can be thought of as harmonics 1, 3, 4, and 5 of a changing fundamental.

When the sets are moving at different modulation rates, the (missing) fundamental of Set 1 follows the slow-changing pattern, whereas the fundamental of Set 2 (i.e., the tone itself) follows the fast-changing pattern.

Note: It is not necessary to choose any particular subsets in order to cause them to segregate by imposing different patterns of frequency change on them. However, if the two segregated subsets have similar timbres and cross in frequency, the ear is less able to track one of them at the points where it crosses the other. The goal of having the two subsets be easily distinguishable was what motivated the choice of partials both in this and the previous demonstration.

Reading. The readings for the issue of grouping by harmonicity are given in Demonstration 18. There is a discussion of grouping by common fate in McAdams (1984), Chalikia & Bregman (1993), and in ***ASA-90***, pp. 248-289, 657-658.

21. Effects of rate of onset on segregation.



The auditory system seems to be particularly interested in sounds with abrupt onsets. Such sounds stand out better from a background of other sounds than do slow-rising sounds. They also seem louder and better defined. An example of rapid-onset sounds in musical instruments are the notes from plucked or struck instruments, such as the piano, the guitar, or the xylophone, which reach their maximum loudness immediately upon being struck or plucked, and then decay in loudness. Examples of slower-rising sounds are the glass harmonica, or the notes of gently bowed string instruments such as the baroque viola da gamba.

The salience of a sound that has a suddenly increasing amplitude is illustrated in this demonstration. The signal is a cluster of four overlapping pure tones as shown in the figure. The order of onset of the tones in the cluster corresponds to the position of the panels. For example, the top one shows Tone 1 starting first, and the bottom one shows Tone 4, which starts last.
The components of the cluster begin at different times, but end together. In a series of presentations of clusters, we make the onsets more abrupt (rise times shorter); however, the asynchrony of onset of the four components is left unchanged. The tones sound clearer, more “bell-like”, and more distinct when their onsets are more percussive. The figure shows the case in which the onset (rise) time of Tone 1 is 125 milliseconds.

As the tones become more percussive, the order of the four becomes easier to judge. To show this, we present the clusters in pairs. While every cluster uses the same four notes, they are presented either in the order MHLM or MLHM, where H is high, M is medium, and L is low (the order MLHM is shown in the figure). You are asked to decide whether the two clusters, presented as a pair, have their components in the same or in different orders. This discrimination is made more difficult by the fact that both clusters have the same note, M, at both their beginnings and ends, which are the most salient positions. As the rise times become shorter across presentations, it becomes easier to hear the order of the components.

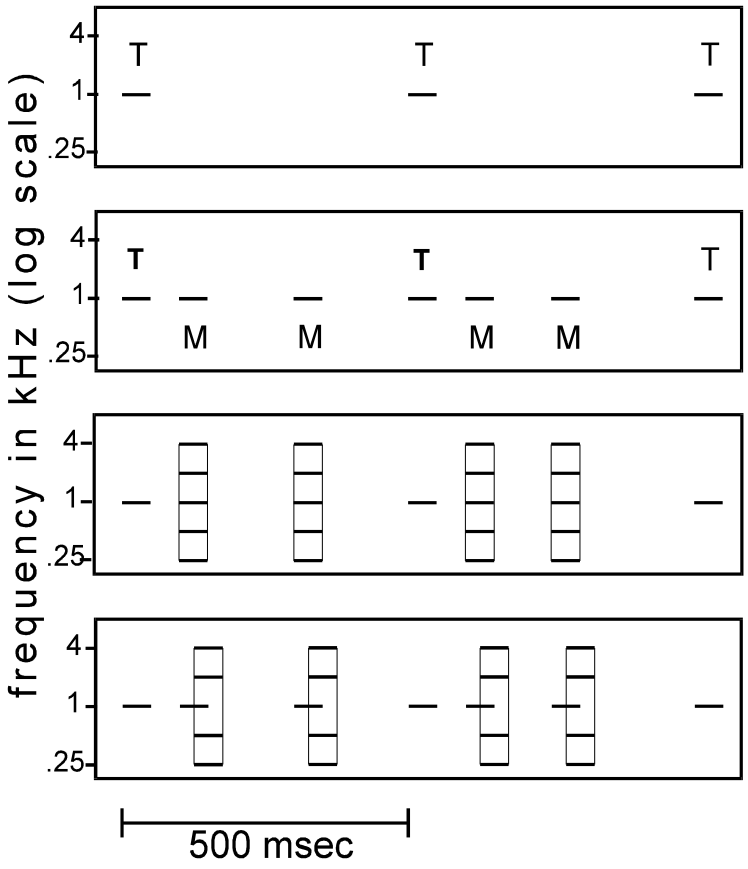
Technical details. Each note is a pure sinusoidal tone with a triangular amplitude envelope; that is, its amplitude rises linearly from zero to a maximum and then, with no pause, falls linearly to zero again. Each component begins 120 msec after the previous one does. The amplitudes can only be stated in relative terms because the actual amplitudes depend on how high the listeners set the volume on their playback equipment.

The amplitude envelope of each tone in a cluster rises at the same rate (in amplitude units per second), but not all of them reach the same maximum amplitude (designated arbitrarily as one amplitude unit). Instead, as seen in the figure, Tone 1 is the only one to reach it. The second and later tones are constrained to never exceed any previous tone in amplitude. Therefore as a tone (other than the first) rises, at some point it will reach the (decaying) amplitude envelope of the first tone, shown as a dotted line in the figure. At that point the rising leg of its amplitude envelope will stop and it will start to follow the decaying envelope of Tone 1. This guarantees that the critical L and H tones will never stand out by being of higher amplitude than the current value of any earlier-starting tone. The peak amplitudes of the second, third and fourth tones, relative to the first one, are .88, .76, and .64, respectively.

The frequencies of the components, in hertz, are as follows: L=700, M=800, H=900. The rise time (measured on the first tone of the cluster) takes the following values in successive pairs: 250, 125, 50, and 10 msec. Each full rise and fall takes one second. Therefore, the faster the onset, the slower the decay. However, the differences in decay time, between 750 and 990 msec have a negligible effect on the segregation of the components.

Reading. Tone clusters similar to these were used by Bregman, Ahad, & Kim (1994) and similar ones by Bregman, Ahad, Kim, & Melnerich (1994). Abruptness of onset was studied by Pastore, Harris, & Kaplan (1982). The role of onsets in segregating concurrent sounds is reviewed by Kubovy (1981) and in ***ASA-90***, pp. 261-264.

22. Rhythmic masking release.



The phenomenon of rhythmic masking release, presented in this demonstration, is related to one called “comodulation masking release” (CMR). In the latter, a pure-tone target is to be detected despite the presence of a masking sound formed of a narrow band of noise centered on the frequency of the target. This masker, called the “on-target” band, fluctuates in intensity.

The target tone can be made easier to hear by adding, simultaneous to the target and the on-target noise band, a third sound, the “flanking” band, consisting of another narrow band of noise, far enough removed in frequency from the target to be outside its “critical band” (the frequency range within which sounds interfere with one another). We seem to have added more masking noise and yet somehow made the target more audible. The trick is that the flanking and masker bands must be “comodulated” (i.e., the amplitude fluctuations of the two bands must be synchronized and in phase). Otherwise the release from masking does not occur.

It has been proposed that the correlation of amplitude changes in the masker and flanking bands tells the auditory system that the two bands should be treated as a single sound.

Somehow, this allows the target, which does not fluctuate in amplitude, to be extracted as a separate sound. Regardless of the explanation, CMR shows that frequencies outside the critical band of a target can influence one's ability to hear it.

The “rhythmic masking release” (RMR) of the present demonstration is a variant of CMR in which it is the rhythm of a tone, rather than its mere presence, that is first masked, then released from masking. The figure shows short sequences of tones extracted from the longer sequences used in the demonstration. First we start with a regular ten-tone sequence of pure tones (the “target” tones, labeled T in Panel 1). Then we present the same tones again, but now with a masking sequence (labeled M in Panel 2) of irregularly spaced tones of the same frequency as the T's and interleaved with them. This pattern is presented twice.

The presence of the irregular maskers causes the regularity of the target tones to be camouflaged. However, by adding four flanking tones, as in Panel 3, that are synchronized with the masking tones, but outside their critical bands, we can make it easier to hear the regular (isochronous) rhythm of the target tones, even though the presence of the set of flankers (F) creates loud M+F tones. In Panel 3, the vertical rectangles represent the fusion of the tones (horizontal lines) inside them. The demonstration is presented twice.

The explanation of the release from masking is that the synchrony of the M and F tones causes them to fuse, forming a tone whose timbre depends on the masker and flanker tones together and is therefore richer than that of the unaccompanied target tones. As a result, we hear soft pure tones in an isochronous rhythm accompanied by irregularly repeating rich ones, which the ear segregates from the targets, allowing us to pick out the regular pure-tone series.

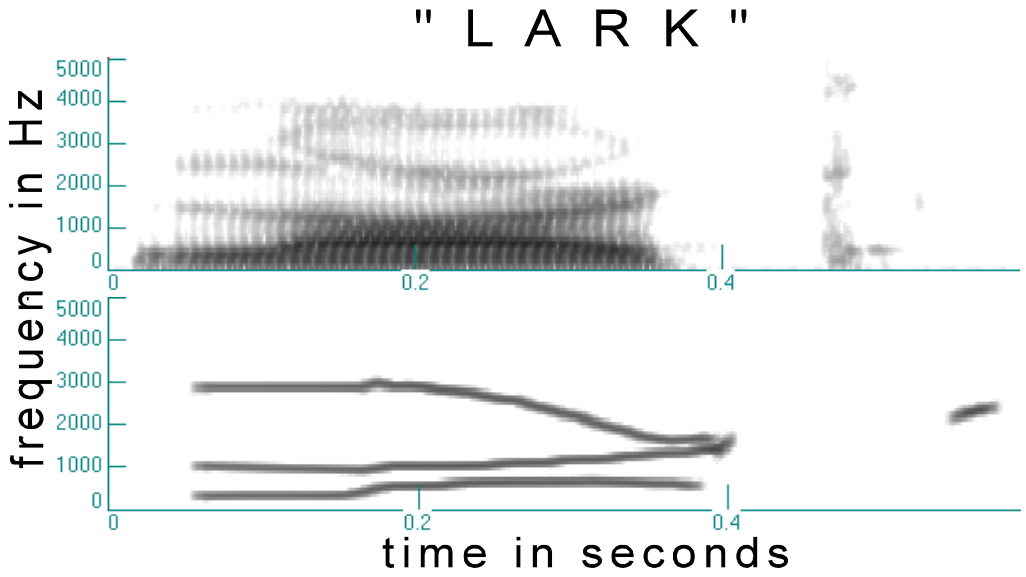
Rhythmic masking release depends on the tendency of the irregular masker and flanker tones to fuse and form a complex tone with a rich timbre. We would expect, therefore, that if the fusion failed, the maskers would again be heard as pure tones that mask the rhythm of the targets. Because asynchrony of onset is known to reduce the fusion of tones (see Demonstration 26), we use it in the final part of the demonstration, slightly delaying the onsets and offsets of the flankers relative to those of the maskers, as shown in Panel 4. The maskers are again heard as pure tones and interfere with our ability to hear the regular rhythm of the targets. This shows how essential the masker-flanker fusion is.
It should be noted that the sequential pattern of this demonstration made it necessary to arrange for a strong level of fusion. Otherwise the presence of the *target* tones would have sequentially captured the M tones, disrupting their fusion with the flanking tones. Strong fusion was created by the use of flanking tones related by octaves to the M tones.

Technical details. The target and maskers are 1000-Hz sinusoidal tones with 5-msec rise and decay times and equal amplitudes. The flanking tone has four frequency components, one and two octaves below and above 1000 Hz (i.e., 250, 500, 2000, and 4000 Hz), whose amplitudes are each one-quarter that of the masking tones. Therefore at the time of the masker tones, there are 5 components: the masker tone itself (at full amplitude) and

the four flanking components whose intensities are each 12 dB less than the masker's. The duration of the targets is 50 msec, but the flankers vary in duration between 30 and 60 msec. The targets have a regular 0.5-sec onset-to-onset time.

Reading. Comodulation masking release was first reported by Hall, Haggard, & Fernandes (1984) and has been reviewed by Moore (1990). It is discussed in ***ASA-90***, pp. 320-323. Rhythmic masking release was recently produced in our laboratory and has never before been described.

23. Sine-wave speech.



The perception of speech is very robust. People are able to use very degraded cues to recognize it. In normal speech there are many acoustic features that tell the auditory system that the acoustic components of a single voice should be integrated. However, the listener can often integrate the components of a synthetic speech sound even when many of these cues are missing. One example is duplex perception of speech. If most of the components of a syllable are played to one ear and a small bit played to the other one, listeners will, paradoxically, both integrate and segregate them. The segregation takes the form of hearing separate sounds at the two ears. Integration can be observed when listeners take the material in both ears into account in recognizing the syllable. One interpretation of this phenomenon is that when bottom-up processes of ASA integrate the acoustic evidence into “packages”, these packages are not airtight; top-down processes of speech recognition can take information from more than one at a time.
Sine-wave speech , described below, is even more degraded than speech divided between the ears, and presents even fewer of the cues for perceptual integration. In the spectrogram of a normal speech sound, shown in Panel 1, we can see peaks in the spectrum (dark regions in Panel 1) known as formants, which change in frequency over time. Vertical striations, where visible, represent the repeated pulses of air coming from the vocal cords, as they open and close. The positions and movements of the three lowest formants specify many of the voiced sounds of a language - sounds such as vowels (“ee”), diphthongs (“I”), glides (“r”), and semivowels (“w”). Formants are not pure tones, but merely the changing peaks in a dense spectrum formed of the harmonics of a low

tone. The frequency changes of formants are not caused primarily by changes in pitch but by changes in the sizes and shapes of resonances in the vocal tract of the talker.

We can make a record of the paths of the formants over the course of a speech sample. Then, amazingly enough, we can use this information to synthesize recognizable speech by using only three or four sine-wave tones and making them glide along the same paths as the formants do in the original speech. Noisy portions (e.g., “s”, “t”, “k”) of the original speech are replaced by steady sine tones positioned at the centers of their frequency bands. In effect, a rich spectrum has been reduced to a few sine tones (Panel 2). This is called “sine wave speech”.

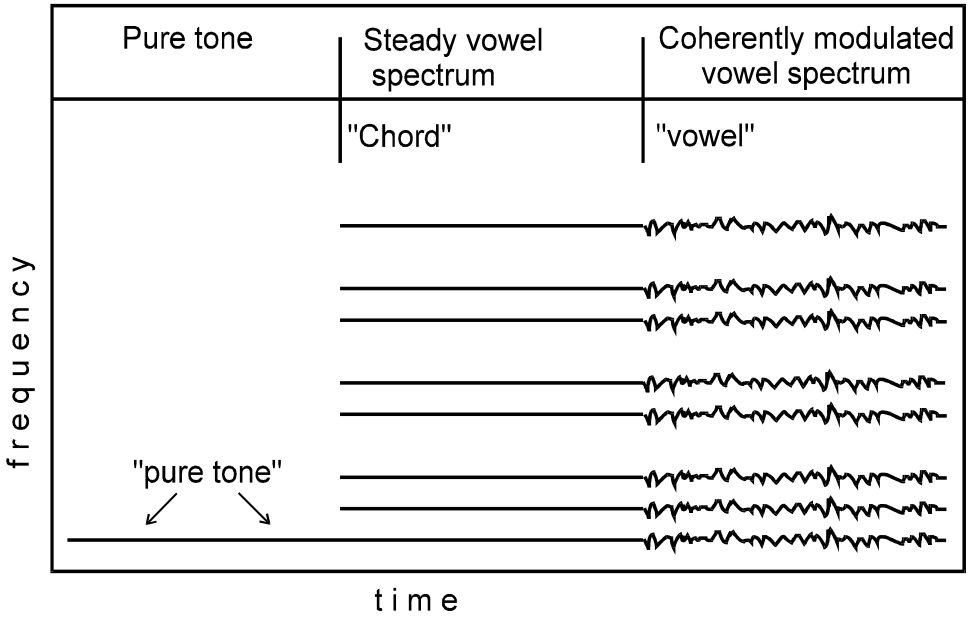
Note that Panels 1 and 2 do not correspond exactly. This is because they are based on the pronunciations of two different talkers. Despite the acoustic differences between sine-wave and normal speech, listeners, after a bit of practice, can recognize the speech. Sentences can be recognized even more easily than isolated words because the context constrains the possible interpretations more strongly in the sentences.

The demonstration first presents a sentence, and then nine single words. **Their identities are given at the end of the booklet, under the heading, “Answers to listening tests.”**

Technical details. The numerical parameters for the formant tracks for these words, previously used in an experiment by Remez, Pardo, and Rubin (1992), were kindly provided by Robert Remez of Barnard College and Philip Rubin of Haskins Laboratories. Each word consists of three or four sine-wave tones whose frequencies and amplitudes vary over time. We used the supplied parameters to synthesize the signals with software sine-wave oscillators, attenuators, and mixers, in the MITSYN signal-processing system (Henke, 1990). The speech in Panel 1 was spoken by Pierre Ahad.

Reading. See Remez (1994), Remez, Rubin, Berns, Pardo, & Lang (1994), and Remez, Rubin, Pisoni, & Carrell (1981).

24. Role of frequency micro-modulation in voice perception.



Normal speech is never steady in pitch, but contains small random fluctuations of about one percent, called micromodulation. These fluctuations affect all the harmonics of a vowel, causing them to move in parallel. This correlated change may be an instance of the Gestalt principle of common fate, or shared change, and may encourage the harmonics to coalesce into a coherent speech sound. In singing, a wider and more sinusoidal fluctuation known as vibrato is produced. The present demonstration uses a mixture of micromodulation and vibrato to generate the pattern of frequency fluctuations. We present three examples using different vowels. Each example has three sections, as shown in the figure. The first contains a pure tone. Then, in the second section, the tone is joined by some of its harmonics, at the intensities that they would have in a particular vowel. This sequence is experienced as the pure tone continuing on, joined by a steady rich tone, which contains a weak vowel quality to some listeners. The perceived continuation of the tone into the steady spectrum is an illustration of the old-plus-new principle, in which one of the frequencies in the complex spectrum - the one that has the same frequency as the pure tone - is classified as “old” and is heard as a continuation of the tone (see readings). In the third section, an identical pattern of vibrato and micromodulation is added to all the harmonics, causing them to fuse. The pure tone can no longer be heard as a separate sound. Instead a singing voice emerges.

This three-part sequence is presented for three different vowels, in turn. The fourth sequence is a simple additive mixture of the first three examples. In it, you first hear a chord of three pure tones, then a chord of three rich tones with different pitches, and finally, a mixture of voices singing three different vowels on different pitches. You will probably be successful in distinguishing at least two vowels.

Technical details. In each part, the timing is as follows: 5 sec of pure tone, 5 sec of a steady rich spectrum, and 10 sec of the spectrum being modulated. In parts 1 to 3, the pure tones are at 400, 500, and 600 Hz, respectively, and the added steady-state harmonics specify, respectively, the vowels “oh” as in “more”, “ah” as in “bomb”, and “e” as in “bed”. In part 4 the previous three parts are superimposed. In it, we first hear three pure tones, 400, 500, and 600 Hz, and then harmonics are added to each one, with the intensities they would have in the three vowels. Finally an independent pattern of vibrato and micromodulation is given to each vowel spectrum. The modulating pattern is a mixture of a sinusoidal component, to make the vibrato, and a band of low-frequency random noise, to make the vibrato irregular. The frequencies of the harmonics are modulated by the sum of these. These examples were created by John Chowning, using FM synthesis, at the Center for Computer Research in Music and Acoustics at Stanford University, and were kindly contributed to this disk by Professor Chowning in the form of a DAT tape, recorded by Glen Diener.

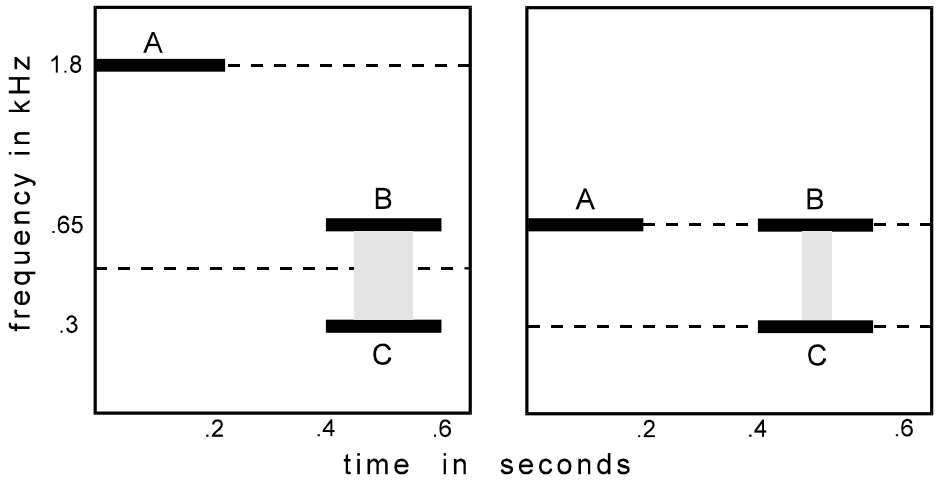
Reading. The effects of frequency modulation on perception are described in ***ASA-90***, pp. 250-260, also in Carlyon (1991), Carlyon, Darwin, & Russell (1993), Chalikia & Bregman (1993). Micromodulation is described in ***ASA-90***, pp. 252-257 and in Marin & McAdams (1991). The old-plus-new heuristic is described in ***ASA-90***, pp. 222-224, 261, 655, and will be illustrated in Demonstrations 25 to 37).

Old-plus-new heuristic

The old-plus-new heuristic is a method for decomposing mixtures of sounds. If a spectrum becomes suddenly more complex or more intense, the auditory system tries to interpret this as a continuing old sound joined by a new one that supplies the additional acoustic stimulation. Therefore the continuing parts of the spectrum are heard as a continuation of the old sound, and the spectral components that are added are used in the formation of a new sound. Since the complex spectrum is taken apart into an old and a new part, the two parts must share the energy of the complex spectrum. So, for example, a wide-band sound can be decomposed into two narrower-band sounds; and an intense spectrum can be decomposed into two softer spectra so that we hear two soft sounds rather than a single loud one.

Various “capturing” effects, in which a preceding simple sound captures a corresponding sound from a mixture are examples of this heuristic in action. Illustrations are presented in the following section.

25. Capturing a tonal component out of a mixture: Part 1.



Up to this point, it has been convenient to distinguish two types of grouping of auditory components. The first, sequential, is illustrated in Demonstrations 1 to 17. It involves connecting components over time to form streams. The second, simultaneous, involves grouping the components present at a given moment to create one or more perceived sounds. Demonstrations 18 to 24 illustrate this type of grouping.

We now go on to demonstrate how these two types of grouping compete with one another in the formation of auditory streams. The figure shows a cycle of tones labeled A, B, and C. The simultaneous process can integrate B and C into a single sound, BC, repeating only once per cycle (- BC - BC **...**), as shown in Panel 1. Similarly the sequential process can integrate tones A and B into a stream of pure tones (A B A B**...**) as shown in Panel 2. Notice that B is grouped differently in the two cases: as part of a complex tone in the first case, and as one of a sequence of pure tones in the second.

These two different groupings represent, in effect, two different theories held by the auditory system about the world. When the A's group only with each other to form one stream (A - A - **...**) and the fused sound BC forms another (- BC - BC **...**), the system is betting that a pure-tone source, A, is producing a series of sounds, accompanied by a complex-tone source, BC, doing the same. On the other hand, when the A's and B's are grouped into one stream and the C's in another, the auditory system is betting that the A's and B's are coming from the same pure-tone source, and that the C's are coming from a second one.

An important determinant of which of these groupings will emerge in perception is the frequency separation between tones A and B. If they are far apart, A segregates from B, and a sequence of A's is heard to be coming from a pure-tone source. B's segregation

from A allows it to fuse with C. B is heard only as a frequency component of the BC sound (Panel 1). However, if A and B are close in frequency, they group together to form one pure-tone stream, breaking the tendency of B to fuse with C, thereby excluding C into a second stream (Panel 2).

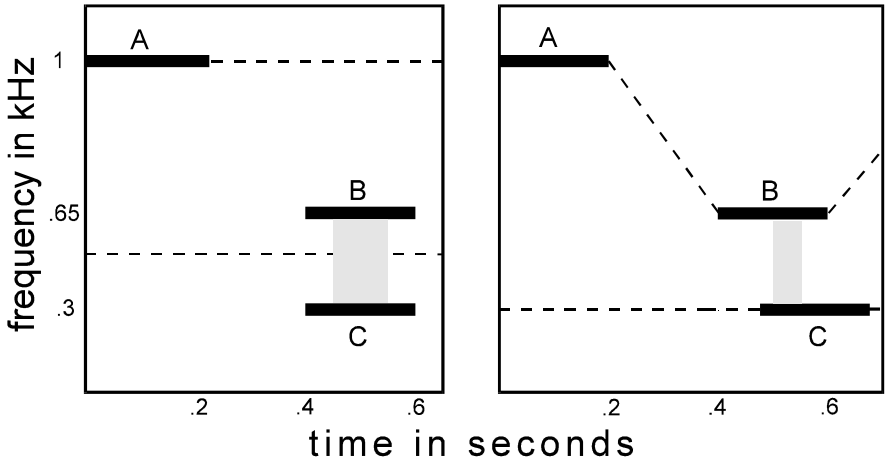
In this demonstration, three different conditions are constructed by placing A at three different frequency separations from B: far (1800 Hz), intermediate (1200 Hz), and no separation (650 Hz). In this last condition, when A groups with B, we hear repetitions of a pure tone with a single frequency, since the frequencies of A and B are identical. The capturing of B by A means that the "force" of sequential integration in this condition is stronger than that of simultaneous integration. Panels 1 and 2 show the conditions with A at 1800 and 650 Hz, respectively. The dotted line shows sequential integration and the width of the shading that joins B with C represents how strongly they fuse.
Each of the cycles of A, B, and C is preceded by a few cycles of a standard with only A and B in it (with frequencies appropriate to that condition). If A and B are sequentially integrated within the full ABC pattern, you should be able to continue to hear the sequence A B A B **...** inside the full pattern, and this should be accompanied by a series of low C tones. If, on the other hand, B and C are undergoing simultaneous integration, you should hear a complex BC tone alternating with a pure A tone.

Each example first presents the A-B standard four times, followed by 12 repetitions of the full ABC pattern. You are to judge how easy it is to hear the A-B high-tone stream in the full pattern.

Technical details. A, B, and C, are pure tones of equal amplitude. The frequencies of B and C are always 650 and 300 Hz, respectively. The three frequencies of A are, in order, 1800, 1200, and 650 Hz. The durations of the tones and intertone silences are 200 msec. For the tones, this includes 10-msec rise and decay times.

Reading. For references to this type of sound pattern, see ***ASA-90***, pp. 29-31, 171, 187-188, 216-219, 230-232, 337, 615, 630. For a discussion of the competition between simultaneous and successive forces of integration, see ***ASA-90***, pp. 218, 335.

26. Capturing a tonal component out of a mixture: Part 2.



This example uses a variation of the A-BC pattern used in the previous demonstration. We show that tones that are played concurrently are less likely to fuse into a single sound when their onsets and offsets are asynchronous. It also shows that we can strengthen the A-B grouping by weakening the BC fusion. Throughout this demonstration, we hold the frequency separation between A and B constant. The asynchrony of B and C is varied across presentations from complete synchrony to 60-msec asynchrony. The degree of fusion of B and C is represented by the width of the shaded region that connects them. When B and C are synchronous, as in Panel 1, they fuse into a single sound, rejecting A into a separate stream. When they are asynchronous by 60 milliseconds, as in Panel 2, B no longer fuses strongly with C. Instead B groups with A to form a pure-tone stream. It is as if A and C were competing for the ownership of B.

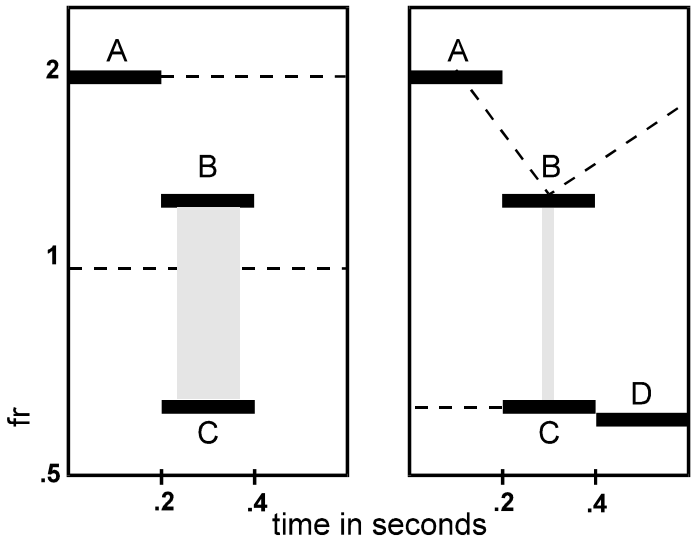
The use of an intermediate A-B frequency difference (350 Hz) ensures that the tendency to form an A-B

stream is at an intermediate strength at which it is balanced with the tendency to fuse B with C. This allows the asynchrony of B and C to display its effect more powerfully by unbalancing the grouping forces. As a result, B groups with A. Listen to how the B-C asynchrony affects your ability to hear a pure-tone stream containing A and B.

Technical details. The frequencies of A, B, and C in all conditions are 1000, 650, and 300 Hz, respectively. They are all sinusoidal tones of equal amplitude. The durations and timings of tones are the same as in Demonstration 26, except that both the onset and offset of tone C are delayed by 0, 30, or 60 msec relative to those of B in the three successive examples. For each A-B frequency separation, there are four cycles of A-B alone, then twelve cycles of the full pattern.

Reading. See Demonstration 25. See also ***ASA-90***

27. Competition of sequential and simultaneous grouping.



This demonstration shows that a change in the strength of grouping, at one time-by-frequency position, can alter the "forces" of grouping at other positions. Panel 1 shows an A-BC cycle that is like those of Demonstrations 25 and 26, except that there is a one-tone silence after the BC complex. Because of the parameters chosen for A, B, and C, the A-B pure-tone sequence is hard to hear in the A-BC pattern. However, when D, a tone that is close in frequency to C, is added to the cycle, in place of the silence, to form the sequence A-BC-D-A-BC-D ..., the sequential grouping of C with D competes with, and weakens, the fusion of C with B. This happens because when component C groups with D, C takes on the role of an independent tone, forming part of a C-D pure-tone stream, rather than being heard as just a component of the complex tone BC. This weakening of the BC fusion has the effect of releasing B to group more strongly with A in an A-B pure-tone stream. In this way, the C-D grouping has affected the A-B grouping, even though neither A nor B has been changed. Therefore, we can view the addition of D to the pattern as having introduced a constraint that has propagated to A and B, remote elements in the field. So the grouping of A and B is not only controlled by the properties of A and B themselves but by a connected set of relations that work their way across the auditory field.

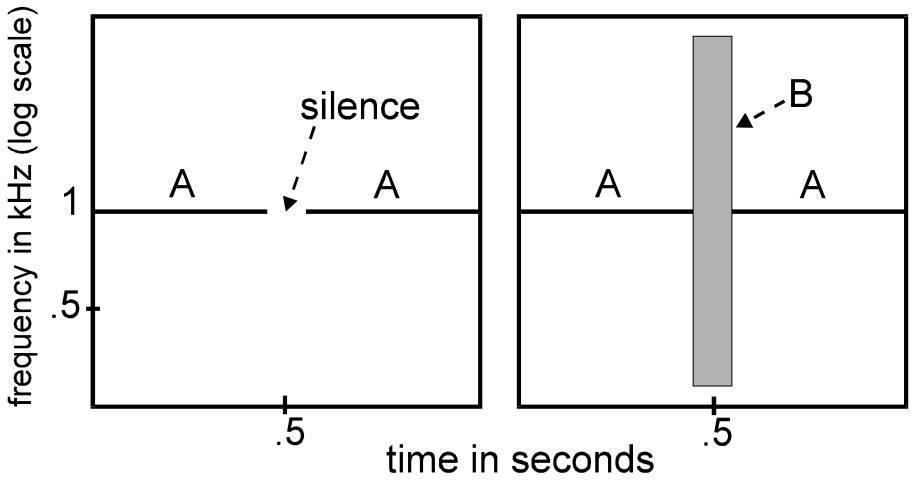
Each A-BC or A-BC-D "comparison" cycle is preceded by a standard in which only A and B are present. In the first comparison cycle, D is absent. In the second, D is present. You are to judge in which case the A-B standard can be heard more easily in the comparison cycle.

Technical details. A, B, C, and D are equal-amplitude sinusoidal tones, 200 msec in duration, including 10-msec rise and decay times. Their frequencies are 1900, 1200, 650, and 613 Hz, respectively. There are no silences between A, BC, and D.

Reading. This demonstration is based on an experiment by Bregman & Tougas (1989), and is discussed in ***ASA-90***, pp. 339-345. The competition between sequential and simultaneous grouping is discussed in ***ASA-90***, pp. 303-311, 332, 335.

28. Apparent continuity.

Loudness warning: To avoid damage to your ears or playback equipment, you should not exceed the volume settings established by the initial calibration procedure given on page9. If you are listening to this over loudspeakers, position them to point straight at you so the high frequencies will not be lost.



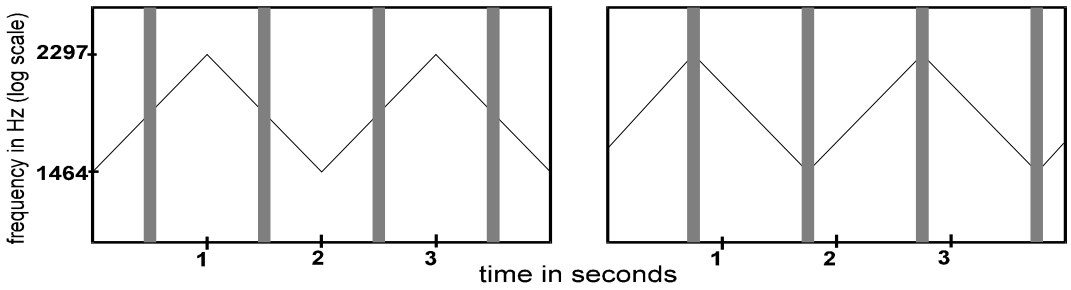
This is one of several demonstrations in which part of a signal, A, is removed and replaced by a much louder one, B (see Panel 2). If B is loud enough, A appears to continue behind it. The effect seems only to occur when the neural activity during B includes activity very similar to what would normally occur during A, so that there is evidence that A may still be present during B. For example, in the present demonstration, continuity is best when the spectrum level of B is at least equal to that of A (i.e., when the frequency of A is also present in B, with at least the same intensity as in A). So an argument can be made that the apparent continuity of A is not an illusion. It is just a normal application of the "old-plus-new" heuristic, in which, whenever a spectrum, A, is replaced by a more complex or intense one, B, the auditory system tries to detect a continuation of A's energy inside B, and if successful, groups it with A.

The noise that fills the gap in A begins at 0 amplitude (Panel 1). After each group of cycles, the noise is increased in intensity. Its amplitudes are 0, .025, .05, .1, .2, .5 and 1 (relative to the final level). In the final group of cycles (Panel 2), the spectrum level of the noise is the same as that of the tone (although the overall amplitude of the noise is about 20 times that of the tone.) At this setting, the tone seems to remain on continuously behind the noise. You can judge whether this happens at any of the earlier amplitudes.

Technical details. A is a 1000-Hz pure tone. B is a burst of white noise. The cycle alternates 900 msec of tone with 100 msec of noise.

Reading. Much of the research was performed by Richard Warren and his associates, and is described by Warren (1982, 1984). Apparent continuity is also discussed in ***ASA-90***, pp. 344-383.

29. Perceptual continuation of a gliding tone through a noise burst.



Loudness warning: To avoid damage to your ears or playback equipment, do not exceed the volume settings established by the initial calibration procedure given on page 9. If you are listening over loudspeakers, position them to point straight at you so the high frequencies will not be lost.

This demonstration shows that the apparent continuity illustrated in Demonstration 28 is not restricted to steady-state signals, i.e., those that are constant in their properties. We present the repeatedly rising and falling glide pattern shown in the figure. It has interruptions either in the middle of the ascending and descending legs (Panel 1), or at the peaks and valleys (Panel 2). These interruptions can be either silences or loud noise bursts. Whe hear a glide continue to ascend, descend, or change directions behind an interrupting noise. The auditory system is not simply restoring the missing material by duplicating whatever went on before or after the noise; rather, it is offering a guess as to what is happening behind the noise.

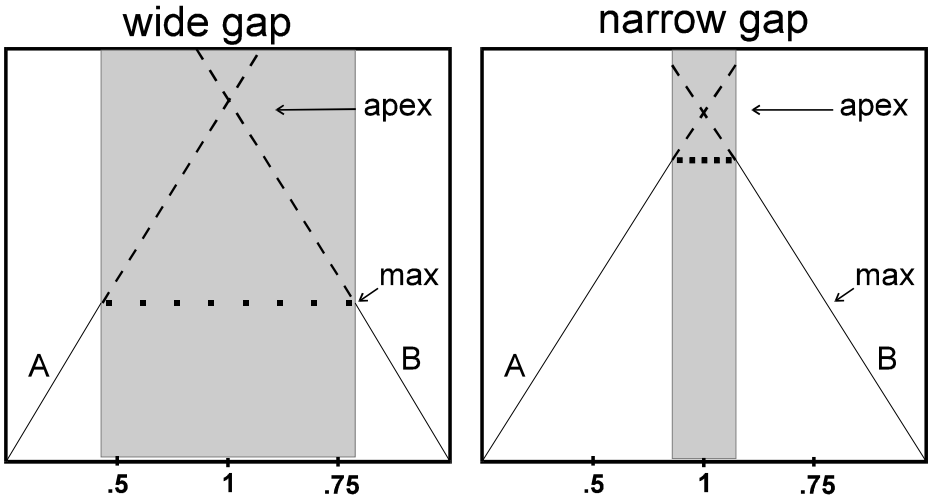
In the first example, the middle points of the rising and falling sections are deleted to create silent gaps in the signal (see Panel 1). Next, the gaps are filled with loud noise bursts. The gliding tone is heard now as being complete and continuing behind the noise.

In the next two examples (Panel 2), the gaps appear at the high and low points of the glide pattern. First they contain silence, then noise. Again the gliding tone seems to continue behind the noise.

Technical details. The gliding tone is sinusoidal and rises and falls in a straight line on log coordinates (equal pitch change per unit of time). Each full cycle, including the gap or noise, takes 2 seconds, and the duration of the gap (or noise) is 0.15 sec. The lowest and highest frequencies are 1464 and 2297 Hz. The noise is white and its peak amplitude is 21 times that of the tone.

Reading. The research on which this is based was carried out by Dannenbring (1976). Apparent continuity is discussed in ***ASA-90***, pp. 344-383.

30. Absence of pitch extrapolation in the restoration of the peaks in a rising and falling tone glide.



Loudness warning: To avoid damage to your ears or playback equipment, you should not exceed the volume settings established by the initial calibration procedure given on page 9.

Demonstration 29 showed that, in certain circumstances, when a short segment of a soft rising-and-falling glide pattern is replaced by a loud sound, it will be heard as continuing behind the noise. The present demonstration uses a similar pattern but introduces a gap only at the peak. We ask whether the bottom-up ASA process extrapolates the trajectory of the remaining parts of the glides to fill in the missing peak. The higher frequency part of one cycle of each of two glide patterns is shown in the figure. The ascending and descending parts of the glide pattern are portrayed as unbroken lines in the figure and the noise burst as a shaded region.

What will the auditory system restore behind the noise? One possibility, shown by broken lines, is that it extrapolates the ascending and descending trajectories to find the apex at which they would have met behind the noise had they been present; accordingly it would hear the glide as rising to this apex and then descending. A second option, shown as lines of dots, is that the ASA system does no extrapolation, but instead, interpolates a glide that joins the highest parts actually heard of the glides (max) by a minimum-length path.

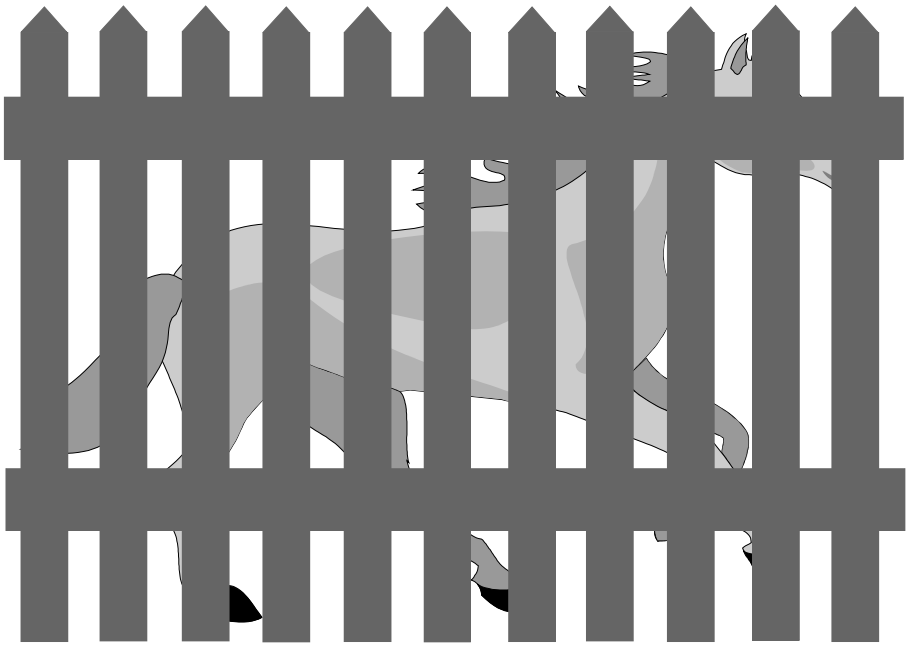
We can determine which of these is true by comparing two cases in which the gap differs in duration, but all other aspects of the glide are identical. In the first case (Panel 1), the gap is wider, and in the second case (Panel 2), narrower. If it is extrapolation that is used for finding the frequency of the missing apex, the same apex (shown in Panels 1 and 2)

should be found in the two cases and the restored glides should reach the same highest pitch. On the other hand, if only interpolation takes place, the highest pitch that is heard should depend on the highest pitch that remains after the peak is removed (marked “max” in Panels 1 and 2). This means that when the peak is perceptually restored, it should sound higher in the pattern with the narrower gap (Panel 2). In the demonstration, you are to listen for the highest pitch you can hear at the peak of the restored glide pattern. First you will hear several cycles in which the gap contains silence, first the wider gap, then the narrower. It is evident that the non-deleted glide portions reach a higher pitch in the second case. Next you will hear the same two conditions with noise inserted in the gaps. Again, it seems that the maximum pitch is higher in the second set of cycles than in the first. We can conclude that, in this case, at least, the auditory system does not extrapolate the rise and fall, but simply joins the highest ends of the glides that are physically present.

Technical details. The gliding pure tone goes from 1464 Hz to 2297 Hz (about 8 semit) and back again in 2 sec. The glides are exponential (equal pitch change per unit of time). The noise is band-passed 1500-3000 Hz). The spectrum level of the noise is about 3 dB higher than that of the tone. The duration of the wide and narrow gaps are 400 and 100 msec respectively.

Reading. The original research on this topic was done by Dannenbring (1976). The question of whether bottom-up ASA performs extrapolation or interpolation is discussed in ***ASA-90***, Ch.4, especially pp. 411-442.

31. The picket-fence effect with speech.



Loudness warning: To avoid damage to your ears or playback equipment, you should not exceed the volume settings established by the initial calibration procedure given on page ***9****.*

Demonstrations 28 to 30 showed how a steady or gliding tone with gaps could be heard as continuous when the gaps were filled with noise. The effect is not restricted to tones, but can also occur with speech. If the gaps in the speech are not too wide, listeners hear the speech as continuously present behind the noise, although what it is saying may not be clear. When the gaps are the width of a single phoneme, listeners may even think that they hear the missing sound when it is predictable from context. This is known as "phonemic restoration". The perceived continuity of speech through a series of gaps that are filled with noise has been called the "picket fence effect" by analogy with the visual case, shown in the figure, in which an object, viewed through a picket fence, seems continuous despite the occlusion of some of its parts by the fence.

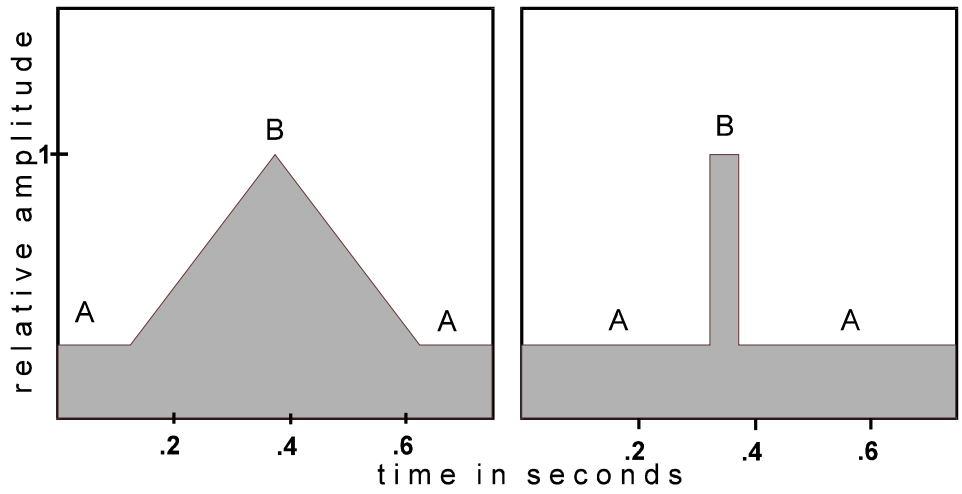
In this demonstration we illustrate the perceived continuity of speech when interrupted by white noise. In the first example, we play a sequence of two sentences in which half the sound has been eliminated by taking out every other one-sixth-second segment and replacing it with a silence. Then the same signal is played with loud noise bursts replacing the silences. Finally, the intact sentences are played.

In the second example, we again delete half the speech, but this time every other quarter-second segment. Despite the longer deletions, the listener will typically hear the sentences as continuous, though unclear. First the gaps are filled by silence, then with loud noise.

Technical details. The speech was digitized at a sampling rate of 22255 samples/sec. Then gaps were introduced by gating it with a raised square wave of 3 Hz in the first part, and 2 Hz in the second part, causing both the durations of the gaps and the surviving speech segments to be one-sixth and one-quarter seconds respectively. In the case where the gaps contain noise, the noise was turned on and off instantaneously at the beginning and end of each gap. The RMS amplitude of the noise is about twice that of the loudest speech segment. The intensity of the noise spectrum drops by 6 dB per octave.

Reading. Early research on this effect was done by Miller and Licklider (1950). Warren and his associates have done a number of studies of the effect (summarized in Warren, 1982; see also Bashford & Warren, 1979, 1987a, 1987b; Bashford, Meyers, Brubacker & Warren (1988). Other work includes Repp (1992), Samuel 1981a, 1981b, Verschuure & Brocaar (1983). Theoretical issues in the apparent continuity of speech are discussed in ***ASA-90***, p. 373. Apparent continuity in general is discussed in ***ASA-90***, pp. 344-383.

32. Homophonic continuity and rise time.



Homophonic continuity occurs when a steady noise burst, A, at a lower intensity, suddenly rises to a new intensity, and then returns to, and continues at, the lower level, as shown in the figure, especially Panel 2. Listeners hear the A sound as if it had continued unchanged in intensity for the whole time. Instead of hearing the rise in the intensity of A as such, it is heard as a second sound joining A.

This treatment of the signal can be interpreted as an instance of the old-plus-new heuristic, carried out as follows: When any spectrum suddenly changes in a way that permits the assumption of an unchanging old sound and an added new one, it will be heard in that way. A sufficient amount of energy to support the interpretation of a continuing A is extracted from the louder section, and the residual energy is heard as a new sound (B). The likelihood of hearing the two-sound interpretation is enhanced if the rise and fall of the intensity change is sudden (Panel 2).

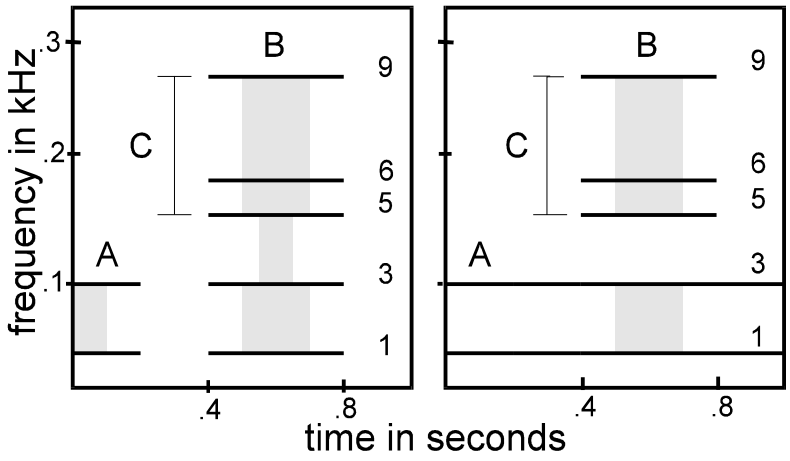
The present demonstration presents a repeating cycle of a softer A segment alternating with a louder B segment. In the first part the changes are gradual, with a long rise and fall in intensity (see Panel 1). In the second part the changes are sudden and the high-intensity segment is short (Panel 2). The signal shown in Panel 1 is typically experienced as a rise and fall of intensity. That of Panel 2 is typically heard as a long continuous noise burst, joined briefly by a second.

Technical details. The signals are bursts of white noise shaped by different amplitude envelopes. The amplitude of the soft steady noise at the outset is .25 of the peak amplitude. The timing parameters are as follows: (1) For the *gradually* rising signal, the

total duration is 744 msec. The initial low intensity continues for 120 msec, then the amplitude rises linearly for 252 msec to a peak, arbitrarily designated as having an amplitude of 1.0. Then it immediately begins to fall linearly for another 252 msec until it reaches .25, which it maintains for 120 msec. (2) For the *abruptly* rising signal, the total duration is 724 msec. The low intensity noise holds steady for 342 msec. Then it rises in 1 msec to the peak intensity, which it holds for 32 msec, then falls in 1 msec to the lower intensity, which it holds for 348 msec.

Reading. In ***ASA-90***, the old-plus-new heuristic is discussed on pp. 222, 224, 261, 371, and 655, and the formation of residuals is described on pp. 343, and 634. Homophonic continuity is discussed on p. 372 of ***ASA-90***, in Warren, Obusek, & Ackroff (1972), and in Warren (1982), p.141.

33. Creation of a high-pitched residual by capturing some harmonics from a complex tone.



You will recall that the old-plus-new heuristic decomposes spectra: when a spectrum, A, suddenly gets more complex or louder, yielding spectrum B, the latter should, if possible, be interpreted as a continuing A plus a new sound C. In short, B = A + C. In earlier demonstrations, our focus has been on the older sound, A, showing that it was heard as part of B. In the present demonstration we highlight the formation of C, the residual. We present a tone with a rich spectrum, B, alternating with a tone, A, that contains only some of its lower harmonics. Under some conditions, Tone A will capture out of B the harmonics that it shares with B, into a perceptual stream of low sound. The harmonics that are unique to B will be left behind as a residual, to be heard as a separate higher-sounding tone. In the figures, the B spectrum, and some of the A spectrum, before and after it, are shown. The harmonics are numbered.

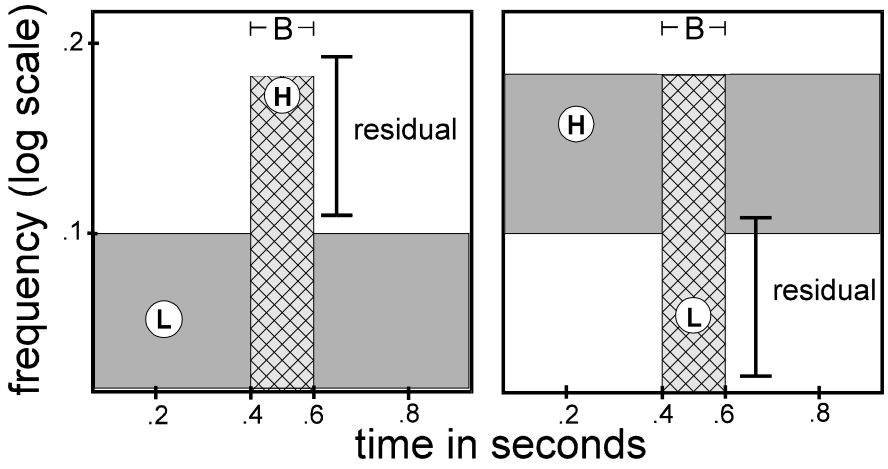
The strength of the decomposition of B into old + new depends on the length of the silent gap between A and B. With a 200-msec gap (Panel 1), the residual (C) is heard only weakly as a separate sound, if at all. When the delay is 0 msec (Panel 2), a clear residual can be heard. In the figure, the width of the shading shows the hypothesized strength of fusion between the harmonics that it connects. There are three signals presented in the demonstration, each repeated eight times. First we play the full B tone to give you an idea what it sounds like before it is broken into A and C. The second presents an alternation between A and B with a 200-msec gap. The third shows what happens when there is no silent gap between A and B.

Technical details. The spectrum of B contains the following harmonics of a 300-Hz fundamental (their intensities, relative to the fundamental, are shown in square brackets): 1st [0 dB, the reference intensity], 3rd [-20 dB], 5th [-28 dB], 6th [-33 dB], and 9th [-38

dB). The spectrum of A, the captor, consists of the 1st and 3rd harmonics, with intensities the same as in B. The duration of one cycle of the A-B alternation is 1 sec. In the case having the 200-msec gap before and after B, a 200-msec A alternates with a 400-msec B. In the case that has no gap, a 600-msec A alternates with a 400-msec B.

Reading. The old-plus-new heuristic is discussed in ***ASA-90***, pp. 222, 224, 261, 371, and 655. The formation of residuals is described in ***ASA-90***, pp. 343, 634.

34. Capturing a band of noise from a wider band.



This example resembles Demonstration 33, except that noise bursts are used rather than harmonics of a complex tone. The main stimulus is a band of noise, B, which can be thought of as containing two sub-bands, a lower, L, and a higher H (B = L + H). If we create a cycle in which a sound, consisting of only one of the sub-bands (say L), precedes and follows B, this sub-band captures the corresponding frequencies from the wide band, leaving behind a residual that resembles the other sub-band (e.g., H), as shown in Panel 1.

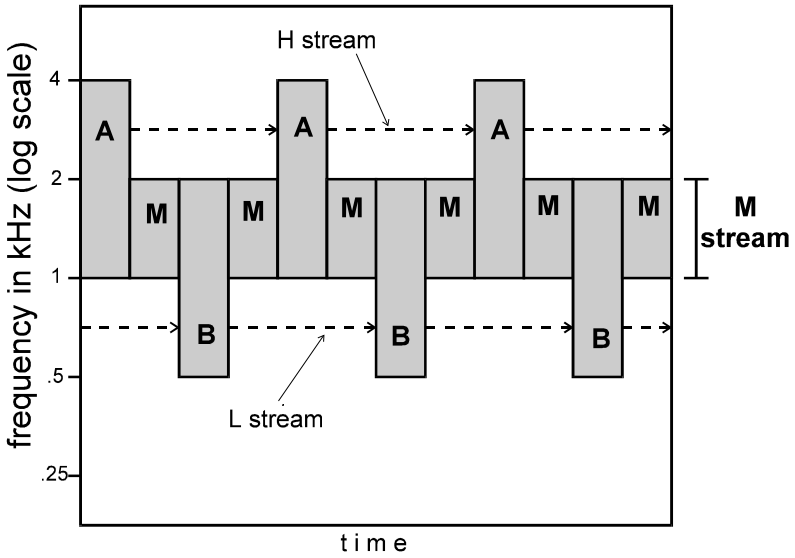
There are two signals in this demonstration. In the first (Panel 1), B is preceded and followed by L, without any gaps. This captures the lower part of B to serve as part of a long, unbroken, low-frequency noise. The higher frequencies of B are left behind to form a separate, short, high-sounding noise burst (H), appearing once on each cycle.

In the second signal, shown in Panel 2, we do just the opposite, using H to capture the higher frequencies of B which then serve as a part of a long, unbroken, high-frequency noise, leaving behind the L frequencies, which are heard as a short, low-sounding noise burst, appearing once on each cycle.
The first and second signals are presented twice in alternation, so that you can compare the sound of the short bursts. Remember that physically the short B bursts are identical in the two cases; so if they sound different in the two signals, it is because we are experiencing the residuals, rather than the full spectrum of B.

Technical details. The full B burst contains frequencies from 0 to 2 kHz, the L band contains frequencies from 0 to 1 kHz, and the H band has frequencies from 1 to 2 kHz. The B bursts are 200 msec in duration, while the H or L capturing sounds, preceding and following B with no silent gaps, are each 400 msec in duration. The noise bands were created by applying transversal (FIR) filters to a white-noise spectrum.

Reading. See Demonstration 33.

35..Perceptual organization of sequences of narrow-band noises.



Loudness warning: To avoid damage to your ears or playback equipment, you should not exceed the volume settings established by the initial calibration procedure given on page 9.

This demonstration considers a range of frequencies as being divided into three bands, low (L), medium (M), and high (H). By an appropriate alternation of noises built from these bands, we are able to create a complex version of the old-plus-new heuristic, which allows us to demonstrate both streaming and apparent continuity with the same signal.

Three kinds of noise bursts are presented in a cycle. One is burst A, which spans the high and medium bands (H+M). A second, burst, B, spans the low and medium bands (L+M). A third type, contains only the medium-frequency band (M). The order in the cycle is A,M,B,M**...**, as shown in the figure. When this cycle is presented repetitively, we can show that the high, medium, and low bands are heard as separate streams. This experience occurs because the old-plus-new heuristic extracts the shared M band from all sounds, leaving behind H and L as residuals. None of the three bands will group with another: Since A has been segregated from M, the two cannot form a stream. Also H and L are too far apart in frequency to form a stream. So each of the H, L, and M bands will group sequentially only with the next instance of itself, as shown by the broken lines in the figure. The continuation of the M band can be viewed as an instance of the old-plus-new heuristic, which looks for a continuation of M inside the A and B bands. The

grouping of each of the three bands into separate perceptual streams is an example of stream segregation.

First we present the total sequence, which sounds very complex. Then we demonstrate that you have segregated A and B into their sub-bands by presenting, as standards, the sequence of sounds that you would hear if only one of the bands were present. Then, with no break, the full sound pattern follows. It is possible to continue to hear each standard pattern (or something similar to it) right through the full sequence. We present this standard-full-sequence pattern using first the low band alone as a standard, then the high band alone, and finally the middle band alone.

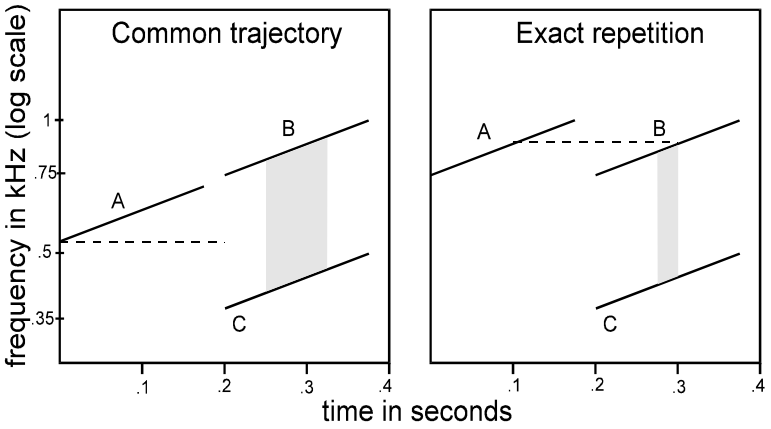
Note that when the cycles of the total pattern begin, it may be hard, for the first few cycles, to hear the standard as a separable part of the total. However, after several cycles of the full pattern, the parts become clearer and the ones that match the standard sequence are more easily extracted.
Another observation is that the high band, heard in the full cycle, seems higher than it does in the standard. This could be due to some inhibition exerted by the medium band on the lower frequencies of the high band.

After listening to the standards and the full cycle for a while, you will be able to pick out the three streams of sound without the aid of the standards. You may find that you can separate the sounds more easily using headphones instead of speakers.

Technical details. The H, M, and L frequency ranges are each an octave wide, L from 500 to 1000 Hz, M from 1000 to 2000 Hz, and H from 2000 to 4000 Hz. Three bands of noise, that combined these underlying ranges, were synthesized: band A, spanning ranges H and M (1000-4000 Hz), band M (1000-2000 Hz), and band B, spanning ranges L and M (500-2000 Hz). Each band of noise was synthesized by adding up the closely spaced harmonics of a 2 or 3 Hz fundamental (only those harmonics that fell in the desired band), with equal amplitude and random phase. This method gives bands with very sharp edges at the specified band limits.

Reading. See Demonstration 33 for references on the old-plus-new heuristic and the formation of residuals. Apparent continuity is discussed in ***ASA-90***, pp. 344-383.

36. Capturing a component glide in a mixture of glides.



We saw, in Demonstrations 25 and 26, how a pure tone, A, captures a second one, B, preventing it from fusing with a third one, C, that is more or less synchronous with it. But what does the auditory system take to be the relation between A and B when it groups them? Is B being interpreted as a continuation of A or as a repetition of it? In Demonstration 33, we saw that the capturing of B by A was best when there was no silent gap between them. Does this mean that in every case of sequential capturing in which A and B have identical spectra, the auditory system is treating B as a continuation of A, rather than a repetition? The question is hard to answer when steady-state tones are used as A and B, because the two possibilities would give the same results. However, if A, B, and C are gliding tones, as in the figure, the alternative theories lead to different predictions.
We use ascending pure-tone glides in this demonstration. A high one, A, is used to capture the higher glide, B, out of a mixture of two glides (BC), as shown in the figure. The captor glide, A, and the mixture (BC) are alternated to form a repeating cycle. When capturing succeeds, you hear a stream of two higher pure-tone glides (A B A B A B **...**) accompanied by a lower pure-tone glide repeating less often (- C - C - C - **...**). If capturing does not occur, you hear a pure sounding glide, A, alternating with a low richer-sounding glide, BC, both repeating at the same tempo.

The frequency of A can be set so that A and B are aligned on frequency-time coordinates (Panel 1), Alternatively, A can be set so that B is an exact repetition of it, with both glides starting and ending at the same frequencies (Panel 2). A's center frequency can be varied, in steps, from the position at which it is aligned with B to the one at which it is the same as B.

In the demonstration, the frequency of the A glide is first set to the position at which A and B fall on a common trajectory (Panel 1). Then the frequency of A is raised in steps until it is identical to that of B. This series of steps is presented twice.

As A's frequency approaches B's, the interpretation of B as a second high glide, resembling A, increases. This is particularly obvious on the second presentation of the series of positions of A. It seems that capturing is best when B is an exact copy of A. Their sequential grouping in this condition is indicated by the broken line that joins them in Panel 2. When the two are aligned on a common trajectory, as in Panel 1, B is not clearly segregated from A. Their continued fusion is indicated by the wide shaded area that joins them in Panel 1.

Could it be that the reason that their alignment was not effective was that there was a silent gap between them? Perhaps the perception of B as a continuation of A requires that there be an unbroken continuity. To find out, the A and B glides are next aligned on a common trajectory, either with a short silence between them or as a continuous long glide with no silence. The first group of cycles contain the silence, the next ones have no silence.

There are two repetitions of this part of the demonstration. Particularly on the second, it appears that the trajectory, AB, can be followed more easily when there is no silent gap. In the discontinuous case, listen for a long ascending broken glide. When the noise is present, listen for a long continuous glide.
The results of the listening test seem to show that for a sound to be heard as a continuation of another, there should be no silent gap between them. Ciocca and Bregman (1987) showed that the apparent continuity of a glide behind a noise burst was experienced more strongly when the part of the glide that exited from the noise was aligned on a common trajectory with the part that entered it. It seems, then, that filling the gap between A and B with noise, rather than leaving it silent, makes it possible for A and B to be heard as part of the same sound. Under these conditions, alignment is important.

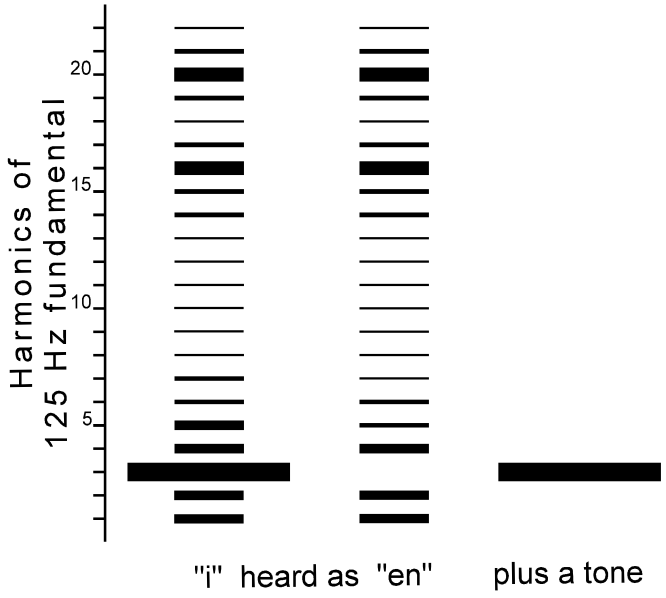
In the third part of the demonstration, we use identical glides as A and B, but the length of the silent gap separating them increases in four steps. It is quite evident that our ability to hear B as a repetition of A, rather than as fused with C to form BC, becomes worse as the gap duration increases.

Technical details. The waveforms of all glides are sinusoidal, and the glides sweep upward at a rate of 2.5 octaves per second. Each is 170 msec in duration, and the silent gap that follows it is 30 msec in duration, so the total onset-to-onset time is 200 msec. All glides are exponential (i.e., straight lines on log-frequency by time coordinates), as shown in the figure. Glide B is always an octave above C, and the two remain unchanged in every example: In 170 msec, B sweeps from 744 Hz to 998 Hz and C from 372 to 499 Hz. The frequency sweep of A varies from being identical that of B to being a half-octave below it.

Reading. The grouping of gliding tones is described in ***ASA-90***, pp. 96, 101, 108-113, 257-260, 581.

37. Changing a vowel's quality by capturing a harmonic.

Note: When this demonstration is played over loudspeakers, it is very sensitive to the acoustic characteristics of the room, so it is recommended that you listen over headphones.



This demonstration relates to the role of ASA in the perception of speech sounds. In complex, everyday environments, it is necessary that a speech sound's quality should be protected from modification by other ongoing sounds. One way to do this is to use the asynchrony of onset of the speech sound's components from other concurrent sounds to segregate the two, and the synchrony of onset among the vowel's own components to integrate them.

If the ASA system actually does this, we should be able to use this dependence on onset synchronies to fool it. In the present demonstration we play a synthetic vowel, whose third harmonic (a steady pure tone) is kept on all the time while the rest of the vowel goes on and off. In the figure, the line width represents the intensity of each harmonic. Because the tone does not turn on and off with the vowel, it does not integrate with it perceptually, and the vowel sounds as if the energy of its third harmonic, or at least some of it, is missing.

This signal can also be described as a pure tone at the third harmonic alternating with the whole vowel. From this perspective, we can think of the third harmonic as sequentially capturing its counterpart from the vowel, treating the third harmonic as the “old” sound continuing, and the remainder of the vowel as a new event starting. The “new event”, the

vowel minus its third harmonic, sounds like a different vowel. The two descriptions of the signal given in this and the preceding paragraph, are really only differences in terminology; they lead to the same prediction – a change in the vowel's quality.

In this demonstration, the synthetic vowel sounds like “i” (as in “hid”) when played in isolation with its third harmonic making its normal contribution to the vowel's quality. If the third harmonic is physically removed, the vowel has a more nasal sound, somewhere between the “e” in “bed” and the “en” in “bend”. These are our labels; you may find that others are better for you. The figure first shows the spectrum of the “i”, with the third harmonic lengthened to visually display the fact that its onset is not coincident that of the remaining harmonics. Then it illustrates the interpretation of this signal as “en” plus a tone.

First we play 4 cycles of the “i” sound (with a full-intensity third harmonic), then four of the “en” sound (with no third harmonic), so that you can become familiar with them. Then we play 8 cycles of the “i” sound with the third harmonic present all the time, even between repetitions. Its steady presence causes it to be treated as a separate sound and, when the vowel comes on, some of the harmonic's energy does not combine with the rest of the vowel. The result is a change in the vowel's quality in the direction of “en”.

To help you to judge how much of the third harmonic's energy has been captured by the tone, we run a series of trials. Each trial starts with the third-harmonic tone present all the time and the rest of the “i” going on and off eight times. While this is the same on every trial, a comparison vowel, that follows it, changes each time. This comparison vowel starts off as an “i” on the first comparison, but its third harmonic is progressively attenuated on successive trials, making it sound more like “en”. On the first trial there is no attenuation; on subsequent trials, the attenuations are 3, 6, 9, 12, and 18 dB. This allows you to compare the results of capturing to the results of physical attenuation. By deciding which comparison vowel sounds more like the cycling “i”, you can determine how much energy from the third harmonic is being prevented from contributing to the vowel's identity.

Technical details. The “i” spectrum, shown in the figure, was created by additive synthesis of harmonics, using formant peak values close to those given by Peterson and Barney (1952) for the first three formants, but modifying the intensity of harmonics empirically until variations in the intensity of the third harmonic made a clear difference in the identity of the vowel. The fundamental frequency was 125 Hz. The harmonics were all in sine phase. Spectral peaks were at 375, 2000, 2500, and 3500 Hz (3rd, 16th, and 20th harmonics). The vowel durations were 450 msec, including 50-msec rise and decay times. They repeated once per second. For cases in which the third harmonic (375 Hz) was present all the time, while the “i” vowel repeated, the use of additive synthesis allowed us to to avoid discontinuities in the third harmonic when the vowel went on and off.

Reading. The altering of a vowel's identity by starting one of its harmonics earlier than the rest was accomplished by Darwin (1984), but the harmonic was not on continuously

as in the present demonstration. He and his colleagues have done many other studies on the effects of perceptual organization on the perceived identity of vowels. Some are described in ***ASA-90***, pp. 422, 538-542, 565-569, 579-588. Others, published after the preparation of ***ASA-90***, include Carlyon, Darwin, & Russell (1993), Culling & Darwin (1992), Darwin (1990, 1995), Darwin & Ciocca (1993), Darwin & Gardner (1987).

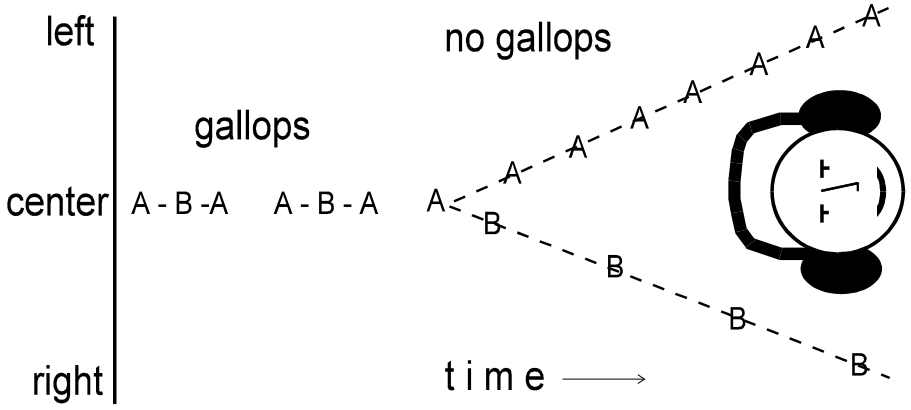
Dichotic demonstrations

While the whole disk is recorded on two channels, Demonstrations 1 to 37 are really monophonic, with the same material on both channels. Only the final four demonstrations are in true stereo, with different material sent to the two channels. You should listen over headphones, but if this is not possible, try to maximize the separation of the speakers.

Left-right balance.
We begin with a procedure for calibrating the balance of the intensities of Channels 1 and 2. First, make sure the playback equipment is set for stereophonic presentation. Then play the calibration signal found on Track 43. It consists of a series of ten beeps, which can be played while adjusting the left-right balance control on the equipment until these sounds seem to come from a centered position.

The stereo demonstrations illustrate the use of spatial separation by ASA in determining the perceived number of sources. We show how it is used by processes of both sequential and simultaneous grouping.

38. Streaming by spatial location.



One of the clues used by ASA in determining the number of acoustic sources is the spatial origin of the time-varying spectrum or of parts of this spectrum. Location can affect both sequential and simultaneous grouping.

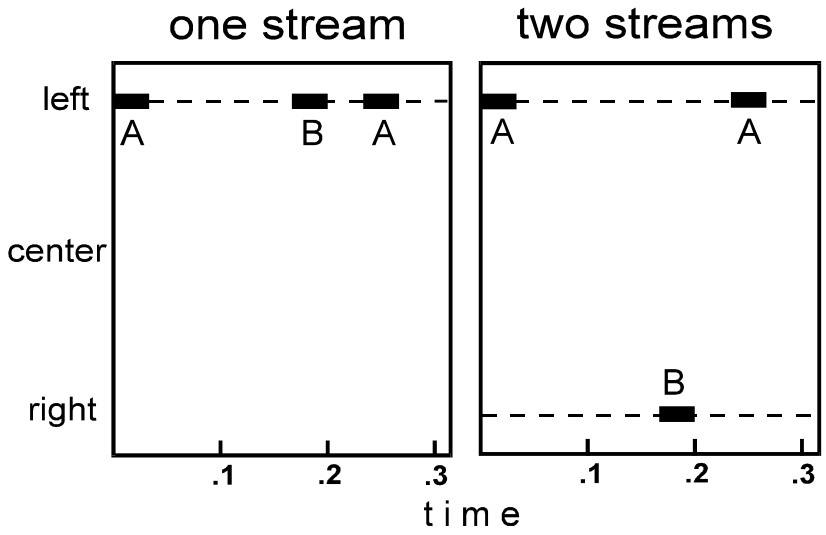
The present demonstration illustrates sequential grouping by location. It uses the same sort of galloping pattern as in Demonstrations 3, 13, and 14, which can be symbolized as ABA-ABA-**...**. Both A and B are bursts of white noise. The only difference between them is the channel of presentation (i.e., ear of presentation when listening over headphones). First, all the sounds appear to be coming from the center. Then gradually, the first and third bursts of each triplet (the A's) move to the left while the middle bursts (B's) move to the right, as suggested in the figure.

The triplet breaks up into two streams, and the galloping rhythm can no longer be heard. Instead we hear a fast, isochronous (evenly spaced) sequence of bursts on the left and a slower, isochronous sequence on the right. Then the bursts gradually come back to the center and you can again hear the galloping rhythm. This sequence is presented twice.

Technical details. The elements are white noise bursts with flat long-term spectra and abrupt 1-msec onsets and offsets. If the sequence is symbolized as A,B,A,-,A,B,A,-,**...**, the A's and B's represent 50-msec noise bursts, the commas represent 50 msec intertone silences and the hyphen (-) represents a missing beat (50-msec silent interval). Therefore the onset-to-onset time for successive A's is 200 msec and for successive B's is 400 msec. The apparent location of bursts A and B are controlled by their relative intensities in the left and right stereo channels. They appear to come from the center when they are each played equally strongly in both channels (i.e., to both ears, if listening over headphones). We make them appear to move to opposite sides by progressively increasing the relative intensity of A in Channel 1 and B in Channel 2, until A comes only to Channel 1 (left side) and B only to Channel 2 (right side).

Reading. The effects of spatial separation on the sequential grouping of sounds are described in ***ASA-90***, pp. 73-83, and by Barsz (1991), Hartmann & Johnson (1991), Lakatos (1993), ten Hoopen (1995), Radvansky, Hartmann, & Rakerd (1992).

39. Spatial stream segregation and loss of across-stream temporal information.



In Demonstration 13, we observed that when sounds segregate into separate streams, it becomes difficult to make accurate judgments about the timing relationships among them. In that demonstration the segregation was due to a frequency difference. In this one, it is due to a difference in perceived location.

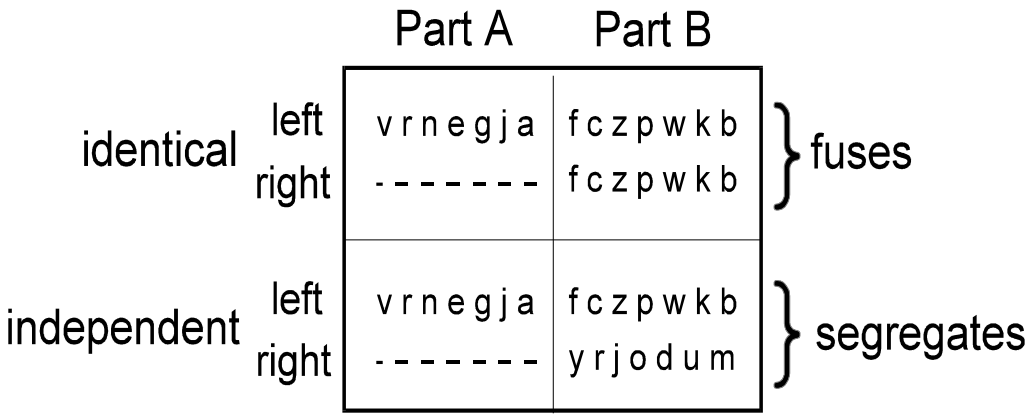
Short noise bursts are played in a galloping rhythm, ABA-ABA-.... The middle burst may or may not occur exactly halfway between the other two in time. You can estimate the effects of spatially based segregation by contrasting the difficulty of two judgments, one using a monaural sequence, the other using a dichotic sequence. In the monaural case (Panel 1), all the bursts are presented to Channel 1 (to the left ear, if listening over headphones). Two examples of the galloping sequence are played, the first with B exactly halfway between the two A's, and the second with B a bit delayed. Try to remember how hard it is to hear that it is delayed. The example is played twice. The second judgment (Panel 2) is to be made on a pair of examples in which A and B are sent to different channels, the A's to Channel 1 (left) and the B's to Channel 2 (right). You must judge whether the middle tone is exactly halfway between the others in the first example or in the second. The sequence of two examples is played twice. **The answer is given at the end of the booklet under "Answers to listening tests".**

It is possible by careful listening to make the judgment correctly on the spatially separated stimuli, but less easily than with the single-location stimuli, in which the judgment is based on a more direct appreciation of the rhythm.

Technical details. Before and after each galloping rhythm, there is a sequence of A bursts alone, which serve to increase the amount of streaming. All bursts are white noise and are 40 msec in duration. In the ABA pattern, there is a 200-msec silent gap between successive A bursts (offset to onset) of which 40 msec are used up by B, leaving 160 msec of silence. When B is placed halfway between A's, there is an 80-msec gap both before and after B. When B is delayed relative to the halfway point, there is a 130-msec silence before it and only a 30-msec silence after it.

Reading. The effects of segregation on gap discrimination are discussed in ***ASA-90*** on p.159. Segregation by spatial location is described on pp. 73, 75, 77, and 644. Loss of temporal information is described on pp. 143-164, and 650.

40. Fusion of left- and right-channel noise bursts, depending on their independence.



If we always fused the sounds arriving at one ear with those at the other, we would frequently make errors by integrating parts of sounds that had come from different sources. It appears that the ASA system has ways to reduce the likelihood of such errors by deciding, in effect, how much of the stimulation at each ear to assign to a common perceived sound.

One of its methods is to look at the correlation between the changes occurring at the two ears. If they are highly correlated, the two should be combined to derive the spatial position of a single sound. If they are uncorrelated, two separate locations should be derived.

In this demonstration, we use white noises as signals. They are illustrated in the figure by rows of randomly ordered letters that represent the material in the two channels. We present the noise in one of two ways. The first is “mono” (called “diotic” when heard over headphones), presenting identical signals to the left and right channels, as shown in Part B of the top half of the figure. The second is “stereo” (called “dichotic” when heard over headphones), presenting independently generated white-noise signals to the two channels, as shown in the bottom half of Part B.

Part 1. The first set of signals is illustrated in part B of the figure (ignore Part A for the moment). In the mono case (top half of Part B), a single white noise is split, sending an identical signal to the two channels. The exact correspondence favors the fusion of the signals. In the stereo part (bottom half of Part B), an independently computed sample of white noise is sent to each channel. Although the left- and right-channel signals have

identical long-term spectra, they are different in detail because noise is never exactly the same twice. This independence favors the segregation of the noises in the two channels from one another.

In the first part of the demonstration, the two types of signal, mono and stereo, are presented twice in alternation. The perception of the result depends on whether you are listening over headphones or loudspeakers (placed near the listener). With head-phones, when you hear the mono signal, the identical inputs to right and left ears is merged by your brain to derive the perception of a single noise burst located at the center of the head. When this signal is heard over loudspeakers, the mono noise is heard as coming from a position somewhere between the two speakers.

When the stereo version is presented dichotically over headphones, one sound appears to come from the left and another from the right, with possibly a softer signal coming from the center of the head. The latter may represent a partial failure to use the independence to segregate the sounds. When heard over speakers, it does not seem to come from two distinct locations as it does with headphones, but seems to be spread out in space and less integrated than the diotic sound. This difference occurs because with headphones, each ear receives sound from only one channel, whereas with loudspeakers, each ear receives some sound from both speakers, which reduces the independence of the signals at the two ears.

Part 2. The effects of identical versus uncorrelated noise at the two ears can also be illustrated by the use of a "capturing" paradigm. We show that the left-channel signal can be more easily captured into a left-side stream when it is uncorrelated with the right-channel signal. The sounds are made into a "capturing sequence" by first presenting the signal shown in Part A of the figure, and, immediately afterward, the one shown in Part B. First a noise is presented to the left channel only (as shown in Part A). Its purpose is to serve as a "captor" for the subsequent left-channel material. Then, with no pause, noise is sent to both channels as illustrated in Part B. So the signal consists of Part A followed by Part B. While we have described the signal as a left-side captor, capturing the subsequent left-side component of the noise, we can also describe this stimulus without reference to capturing. We can see it as a signal in which the left-channel signal is on continuously, and is joined, after 2 seconds, by the right-channel signal. However, the principles of grouping make the same predictions regardless of the choice of description. Both "capturing and "asychrony of onset" predict the segregation of the two channels.

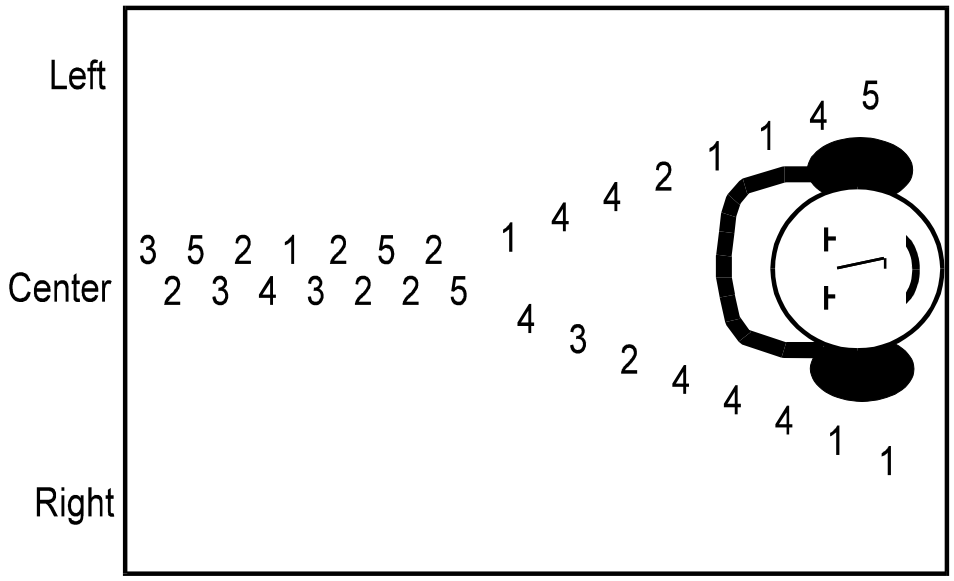
On the disk, the two cases are alternated: mono, stereo, mono, stereo. In the mono case, as soon as the right-channel signal joins the identical left-channel one, they fuse and seem to be coming from the center. There may be a slight tendency to continue to hear something on the left as well, but it quickly disappears. However, in the stereo case, the signals fuse less strongly and you can hear a continuing signal, on the left, all the way to the end of the signal, accompanied by a separate right-hand signal.

Technical details. All signals are white noise bursts. In Part 1, the signals are 2 sec in duration, with 10-msec rise and decay times. In Part 2, the left-channel signal is on

continuously for 4 sec, joined after 2 sec by the right-channel signal. Both have 10-msec rise and decay times.

Reading. The role of correlation in the fusion of stereophonic signals is discussed in ***ASA-90***, pp. 296-297, 311. The old-plus-new heuristic is discussed in Demonstrations 25 to 37.

41. Effects of a location difference of the parts in African xylophone music.



Note: This demonstration is not satisfactory when played over loudspeakers.

We present a demonstration quite similar to Demonstration 38. The difference is that the sounds in 38 are noise bursts, and here they are the notes of the East African xylophone (amadinda) playing the piece, “SSematimba ne Kikwabanga”, which was described in Demonstrations 7, 8, and 9. You will recall that this piece of music is formed of two interleaved cycles of notes, drawn from a pentatonic scale (numbered 1 to 5), as shown in the figure. Each of these cycles is isochronous (notes equally spaced in time), but the interleaving of the two cycles creates high and a low perceptual streams with irregular rhythms.
The piece starts with the parts of the two players interleaved with the result sent to both left and right channels with equal intensity. As a result, we hear the characteristic irregular rhythm of the piece, indicating that the two parts have been integrated. Then the first player's part disappears gradually from the right channel so that it appears to come from the left. At the same time the second part disappears gradually from the left channel, so that it appears to come from the right. The net effect is that the two parts seem to move apart in space, until part 1 is on the left and part 2 on the right. This leads to the separate perception of the two parts with their own isochronous rhythms and the loss of the high and low emergent streams with their irregular rhythms.

This demonstration and Demonstration 38, taken together, imply that when any rhythmic pattern is played fast enough, with sounds arriving, in alternation, from two distinct locations in space, listeners will segregate the sounds arriving from two locations and

lose the global rhythmic pattern in favor of rhythms that reside in each stream taken alone. Although Demonstration 38 involves noise bursts and the present one uses percussive tones, the results are the same.

Demonstrations 8, 9, and the present one use pitch, timbre, and space, respectively, to segregate the parts in this xylophone music.. Pitch and timbre (as well as timing) have been used extensively to control the "layering" in music. The present demonstration suggests that a relatively unexploited resource, space, offers a new set of opportunities to the composer.

Technical details. This example was synthesized in Berlin by Dr. Ulrich Wegner, using digital samples of the sound of the amadinda, and was sent to us on DAT tape. It was re-sampled in our laboratory at 22255 samples per second. Each part is played at a rate of 4.2 tones per second, so the combined rate, when both parts are present, is 8.4 tones per second, which is in the range where stream segregation is obtainable. The sounds are very percussive, having a roughly exponential decay with a half-life ranging between 15 and 30 msec. Their peak intensities are all within a range of 5 dB. See Demonstartion 7 for further details.

Reading. The effects of spatial separation on the sequential grouping of sounds are described in ***ASA-90***, pp. 73-83, and by Barsz (1991), Hartmann & Johnson (1991), Lakatos (1993), ten Hoopen (1995), Radvansky, Hartmann, & Rakerd (1992). A related issue, the decomposition of polyrhythms, is discussed in ***ASA-90***, p.158. See also Demonstration 7.

Answers to listening tests

Demonstration 5
The hidden tune was “Mary had a little lamb"

Demonstration 23
The sine-wave speech demonstration starts with the sentence, “Please say what this word is”, and then presents the list, “sill, shook, rust, wed, pass, lark, jaw, coop, beak”.

Demonstration 39
In the dichotic ABA-ABA- comparison, in the first sequence B is delayed relative to the point halfway between the A's, and in the second, B is exactly at the halfway point.

References

ASA-90 (see Bregman, 1990)

ASYST V4.0 (1982) Keithley Instruments, Inc. Taunton, MA.

Barsz, K. (1991) Auditory pattern perception: The effect of tone location on the discrimination of tonal sequences. *Perception and Psychophysics, 50 (3),* 290-296.

Bashford, J.A., Meyers, M.D., Brubaker, B.S., & Warren, R.M. (1988) Illusory continuity of interrupted speech: Speech rate determines durational limits. *Journal of the Acoustical Society of America, 80,* 1635-1638.

Bashford, J.A.,Jr., & Warren, R.M. (1979) Perceptual sythesis of deleted phonemes. In J.J. Wolf & D.H. Klatt (Eds.)*, Speech Communication Papers*, New York: Acoustical Society of America, 423-426.

Bashford, J.A., & Warren, R.M. (1987a) Multiple phonemic restorations follow the rules for auditory induction. *Perception and Psychophysics, 42*, 114-121.

Bashford, J.A., & Warren, R.M. (1987b) Effects of spectral alternation on the intelligibility of words and sentences. *Perception and Psychophysics, 42*, 431-438.

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

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

Bregman, A.S., Ahad, P., & Kim, J. (1994) Resetting the pitch analysis system. 2: Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones. *Journal of the Acoustical Society of America, 96,* 2694-2703.

Bregman, A.S., Ahad, Kim, J., & Melnerich, L. (1994) Resetting the pitch analysis system. II: Effects of rise time of tones in noise backgrounds or of harmonics in a complex tone. *Perception and Psychophysics, 56 (2)*, 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. (1973) The effect of continuity on auditory stream segregation. *Perception and Psychophysics, 13*, 308-312.

Bregman A.S., & Rudnicky, A. (1975) Auditory segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception and Performance, 1,* 263-267.

Bregman, A.S., & Tougas, Y. (1989) Propagation of constraints in auditory organization. *Perception and Psychophysics, 46 (4)*, 395-396.

Carlyon, R.P. (1991) Discriminating between coherent and incoherent frequency modulation of complex tones. *Journal of the Acoustical Society of America, 89 (1)*, 329-340.

Carlyon, R.P., Darwin, C.J., & Russell, I.J. (Eds.) (1993) *Processing of complex sounds by the auditory system.* New York: Oxford University Press

Chalikia, M.H., & Bregman, A.S. (1993) The perceptual segregation of simultaneous vowels with harmonic, shifted, or random components. *Perception and Psychophysics, 53*, 125-133.

Ciocca, V., & Bregman, A.S. (1987) Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception and Psychophysics, 42*, 476-484.

Cole, R.A., & Scott, B. (1973) Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology, 27*, 441-449.

Culling, J.F., & Darwin, C.J. (1992) Perceptual separation of simultaneous vowels: within and across-formant grouping by F0. Reviewed for *Journal of the Acoustical Society of America*, October 11, 1992

Dannenbring, G.L. (1976) Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian Journal of Psychology, 30*, 99-114.

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. (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. (1990) Environmental influences on speech perception. In *Advances in speech, hearing and language processing. 1*, 219-241.

Darwin, C.J. (1995) Perceiving vowels in the presence of another sound: a quantitative test of the "Old-plus-New" heuristic. In ,C. Sorin, J. Mariani, H. Meloni, and J.Schoentgen (Eds.). *Levels in speech communication: Relations and interactions: a tribute to Max Wajskop.* Amsterdam: Elsevier.

Darwin, C.J., & Ciocca, V. (1993) Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component. *Journal of the Acoustical Society of America, 93 (5)*, 2870-2878.

Darwin, C.J., & Gardner, R.B. (1987) Perceptual separation of speech from concurrent sounds. In M.E.H. Schouten (Ed.) *The psychophysics of speech perception*. Martinus-Nijhoff, NATO-ASI series.

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

Hall, J.W., Haggard, M.P., & Fernandes, M.A. (1984) Detection in noise by spectro-temporal pattern analysis. *Journal of the Acoustical Society of America, 76*, 50-56.

Hartmann, W.M., & Johnson, D. (1991) Stream segregation and peripheral channeling. *Music Perception, 9(2)*, 155-184

Henke, W. L. (1990) *MITSYN: A synergistic family of high-level languages for time signal processing. Version 8.1* Belmont, Mass.: Author.

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

Lakatos, S. (1993) Temporal constraints on apparent motion in auditory space. *Perception and Psychophysics, 54 (2)*, 139-144.

Marin, C.M.H., & McAdams, S (1991) Segregation of concurrent sounds. II: Effects of spectral envelope tracing, frequency modulation coherenece, and frequency modulation width. *Journal of the Acoustical Society of America, 89 (1)*, 341-351.

McAdams, S. (1984) *Spectral fusion, spectral parsing, and the formation of auditory images*. Ph.D. dissertation, Stanford University.

Miller, G.A., & Licklider, J.C.R. (1950) Intelligibility of interrupted speech. *Journal of the Acoustical Society of America, 22*, 167-173.

Moore, B.C.J. (1990) Co-modulation masking release: spectro-temporal pattern analysis in hearing. *British Journal of Audiology, 24*, 131-137.

Moore, B.C.J., Glasberg, B.R., & Peters, R.W. (1986) Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America, 80*, 479-483.

Pastore, R.E., Harris, L.B., & Kaplan, J.K. (1982) Temporal order identification: Some parameter dependencies. *Journal of the Acoustical Society of America, 71 (2)*, 430-436.

Peterson, G.E., & Barney, H.L. (1952) Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 115-84.

Radvansky, G.A., Hartmann, W.M., & Rakerd, B. (1991) Structural alterations of an ambiguous musical figure: The scale illusion revisited. *Perception and Psychophysics*, *52*, 256-262.

Remez, R.E. (1994) A guide to research on the perception of speech. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. New York: Academic Press.

Remez, R.E., Pardo, J.S., & Rubin, P.E. (1992) *Making the auditory scene with speech.* Unpublished manuscript. Dept. of Psychology. Barnard College, New York, NY.

Remez, R., Rubin, P.E., Berns, S.M., Pardo, J.S., Lang, J.M. (1993) On the perceptual organization of speech. *Psychological Review, 101*, 129-156.

Remez, R.E., Rubin, P.E., Pisoni D.B., & Carrell, T.D. (1981) Speech perception without traditional speech cues. *Science, 212*, 947-950.

Repp, B.H. (1992) Perceptual restoration of a "missing" speech sound: Auditory induction or illusion. *Perception and Psychophysics, 51*, 14-32.

Samuel, A.G. (1981a) Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110*, 474-494.

Samuel, A.G. (1981b) The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1124-1131.

ten Hoopen, G. (1995) Auditory spatial alternation transforms auditory time (again): Comments on Lakatos (1993), "Temporal constraints on apparent motion in auditory space." *Perception and Psychophysics, 57 (4)*, 569-572.

Tougas, Y. & Bregman, A.S. (1985) The crossing of auditory streams. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 788-798.

Van Noorden, L.P.A.S. (1975) *Temporal coherence in the perception of tone sequences*. Doctoral dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands.

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.

Verschuure, J., & Brocaar, M.P. (1983) Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise. *Perception and Psychophysics, 33*, 232-240.

Warren, R.M. (1982) *Auditory perception: A new synthesis*. New York: Pergamon Press.

Warren, R.M. (1984) Perceptual restoration of obliterated sounds. *Psychological Bulletin, 96*, 371-383.

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

Wegner, U. (1990) *Xylophonmusik aus Buganda (Ostafrika).* Wilhelmshaven:Florian Noetzel Verlag. (Musikbogen. Wege zum Verstaendnis fremder Musikkulturen, 1) [*Xylophone music of Buganda, East Africa*. (Cassette and booklet.) In the series, Musical bow: Roads to the understanding of foreign musical cultures.

Wegner, U. (1993) Cognitive aspects of amadinda xylophone music from Buganda: Inherent patterns reconsidered. *Ethnomusicology, 37*, 201-241.