Lecture for CIRMMT (June 19, 2008)

"ASA at McGill"

I am delighted to be here today, among my colleagues and friends in CIRMMT. It is an opportunity to review my longstanding connection with McGill's Faculty of Music.. My talk today will be as much about people and scientific conversations as it is about auditory perception and music.

In memory of Earl Schubert. I'd like to dedicate my talk today to the memory of Earl Schubert, a man whom most of you have probably never heard of. A Stanford professor of Hearing Science, who had been a musician earlier, he spent his later years at CCRMA – Stanford's computer music center.

He was a fine student of auditory perception and a generous "guru" who was open and willing to discuss other people's research with them, often over lunch at his home. His counsel was always worth having. He told me about one of his guiding rules, as a West Coast professor: "Publish, or you'll have to go back to the Midwest."

He is one of my two heroes. The other is James J. Gibson, the influential advocate of "Direct Perception". Both these men were fine examples to younger academics, because they gave as high apriority to spirited dialog and good-natured human relations, as to their academic work.

A student. Let me begin with a story about one of my students. In the mid 1970's, a young man got in touch with me from Hawaii. He had been studying music at De Anza Junior College in California and asked whether McGill would be a good place to come if he wanted to study the perception of sound and its relation to music. I encouraged him to apply and in the fall of 1974, he showed up as an enthusiastic undergraduate student. In a single-minded way, he went on to take courses in every subject that the university had to offer that could help him understand auditory perception as it related to music, including psychology courses in thinking and perception, and physiological mechanisms; he also studied physics, mathematics, chemistry, computer science, statistics, and electronic music. He was an extraordinarily goal-directed and ambitious student.

He joined my lab for his research projects and worked on how perceptual grouping was affected by similarities of timbre. When he graduated, he went on to study for his Master's degree at Northwestern University, a period that he later described as the two worst years of his life. Happily, he went on to Stanford for his Ph.D. where he worked under the benevolent supervision of Earl Schubert. It was through this student that my first contacts were made with CCRMA, Stanford's computer Music Centre, where I later spent three extended and fruitful periods. During part of his doctoral career he worked at IRCAM in Paris. It was he who introduced the research on the perceptual organization of sound that we were doing at McGill to the computer music community, through an article we co-authored, but that he initiated.

I'm delighted that this youngster is sitting before us today as the Director of CIRMMT. He has come back to McGill, his adopted home, and I wish him many productive and enjoyable years here. It's true that Montreal is not Paris, but McGill is McGill, an excellent place to do one's work.

Other McGill colleagues and students. CIRMMT is a meeting place for many other old friends and former students. I believe I first met Wieslaw Woszczyk when he visited CCRMA, at a time when I was there. We kept in touch after I came back to McGill, and we have remained friends and colleagues since that time, and have produced two papers together.

Many students from the Faculty of Music came (or, as I suspect, were sent) up the hill from the McGill's Music Faculty, joined by students from other McGill faculties and students from other universities, ordered by their advisors to take my course in auditory perception. There were a large number of them, including Ichiro Fujinaga, René Quesnel, Bruno Gingras, Cory McKay, Olivier Bélanger, Mark Ballora, Norma Welch, Douglas McKinnie, Patrick Bermudez, Sylvie Hébert, and John Usher.

In the 1980's, one of them, James Wright, was much taken by the ideas about perceptual grouping that he had learned in the course, and found that he could apply them to music. We spent many enjoyable hours talking together, with him teaching me about music and how the ideas he had encountered in the course could be applied in this field. Eventually we published a paper together and I based a good part of a chapter in my book, *Auditory Scene Analysis,* on our discussions.

How my research began. As many of you may know, my research on auditory organization began by accident. I was setting up an experiment on learning, using rapid sequences of very short extracts from continuous sounds, such as a dentist's drill or water splashing in a sink, or a voice saying "ah". However, when I listened to the tapes, the sounds appeared not to be in the order in which I had put them there.

My problem in perceiving their order reminded me of the Gestalt Psychologists' ideas about perceptual grouping. As a Master's student at the University of Toronto, many years earlier, I had written an essay on the research and theories of the Gestalt psychologists who had been working on visual perception in the late 19th and early 20th centuries. Some of their visual examples showed that that similar shapes would group together and segregate from dissimilar ones. If my sounds were being grouped by similarity, this would explain why they were not perceived in the correct order. [1.03]

I thought I would like to study this phenomenon in more detail, so I went down the hill from the Psychology Department to visit Istvan Anhalt, one of McGill's rising electronic music composers, whose recently established McGill electronic music studio was situated in a stone coach house at 3500 Redpath Street.

Because Stockhausen was one of the strongest influences on the New Music of that era, Anhalt's studio was equipped with:

- a multi-track tape recorder,
- a variable-speed tape recorder,
- a filter, a mixer, and an oscillator bank,
- an electro-mechanical multi-track sequence recorder, and…
- other gizmos, many of them custom-built by Hugh Le Caine of the National Research Council in Ottawa, who had built the world's first electronic music synthesizer

I tried out some of these devices and found that while they were accurate enough for music, I felt that a greater accuracy was needed for scientific research. So I abandoned Anhalt's lab and found that I could do better by splicing tape manually into short loops that I could play on a tape recorder.

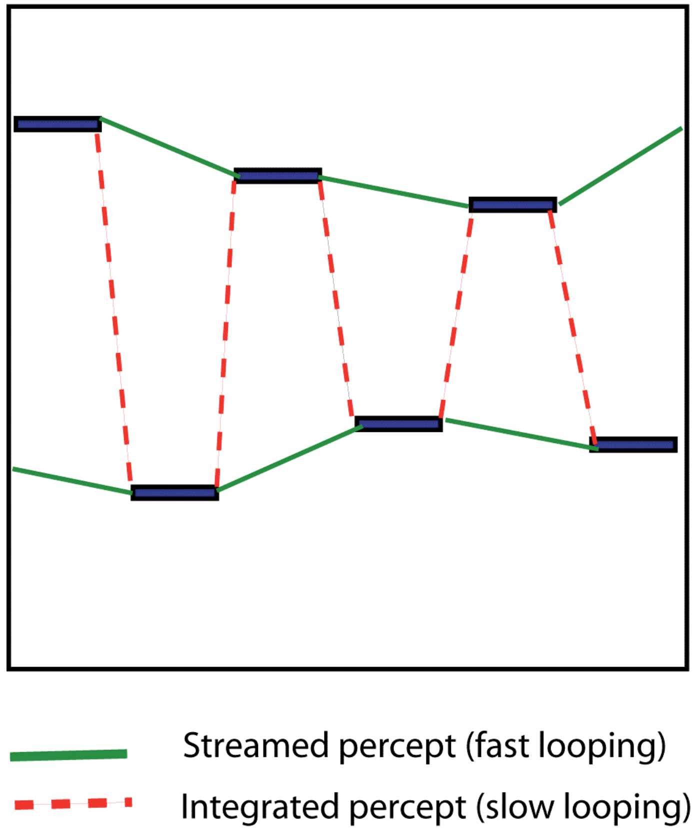
Later I made my stimuli on a PDP-8 computer in the department of Electrical Engineering with the help of a graduate student there, and then on a Link-8 computer at the Montreal Neurological Institute, with the help of a young Jean Gotman, recently arrived from France, who now has a long grey beard, and is still at the MNI, investigating the mechanisms behind the generation of discharges in the brains of epileptic patients. He helped me make long random sequences of tones of different frequencies, played very quickly, so I could hear the groupings of tones that emerged.

Eventually the Psychology department got its own computer, a PDP-11, and money for a young programmer, who. at the time, had just earned his M.A. in Biomedical Engineering. Now he's the Director of Network and Communication Services at McGill – Gary Bernstein.

Simple short repeating sequences. In the late 1960's I read the research of Richard Warren at the University of Minnesota, who had studied the perception of tape loops of 3 or 4 sounds, all of different types. He had discovered that listeners couldn't tell the order of sounds, sometimes even when they were as slow as a 1/3 second per sound. I thought the reason might be the formation of inappropriate perceptual groupings.

So I decided to simplify my research and work with loops of six tones, three high and three low ones, with the high and low tones interleaved in time, at about a hundred milliseconds per tone. When I listened to one of these tapes, I was amazed. I was hearing two parallel streams of sound, a high one and a low one, which apparently had nothing to do with one another, other than that they were happening at the same time. Half of the McGill students who listened to these sequences described their order as three high tones followed by three low ones, or the reverse, despite the fact that it was a strict alternation of high and low tones. I said to myself, "There's at least one good publication in that." This is what it sounded like:

Six-tone loop



Track #1: Stream segregation in a cycle of 6 tones:

Naming the effect. When I published my first report on this, I gave the name "auditory streams" to the separated sound sequences, and called the phenomenon "primary auditory stream segregation" (Bregman & Campbell, 1971). I inserted the word "primary" because in the literature on the theory of animal learning, the word "primary" referred to unlearned phenomena, such as primary drives, and I believed that the perceptual capacity to form auditory streams was present at birth, though later learning might assist it.

The Gestalt psychologists had established that the principles of visual grouping were innate and were present in non-human species. I believed the same would be found for auditory stream segregation. This expectation has now been borne out by recent research that has shown that brain recordings made on infants only 2 to 5 days old indicate the presence of segregated auditory streams. Other research has found stream segregation in many animal species.

My first study on stream segregation was carried out in 1969, and in the 1970's I discovered the work of Leon van Noorden, [Leon recently gave a talk at CIRMMT, but about other work.] In his 1975 doctoral thesis (of which everybody in the field now owns a copy), he had studied, in a very systematic way, the perceptual segregation or integration of a pair of alternating tones of different frequencies (Van Noorden, 1975).

Subjectivity and objectivity. At this point, I want to interject a few words about subjectivity and objectivity in psychological research. The personal experience of the researcher has not fared well as acceptable data for scientific psychology. Since the failure of Titchener's Introspectionism, a very biased form of report of one's experience, in the early twentieth century, and the rise of Behaviourism to replace it, scientific psychology has harboured a deep suspicion of the experience of the researcher as an acceptable tool in research.

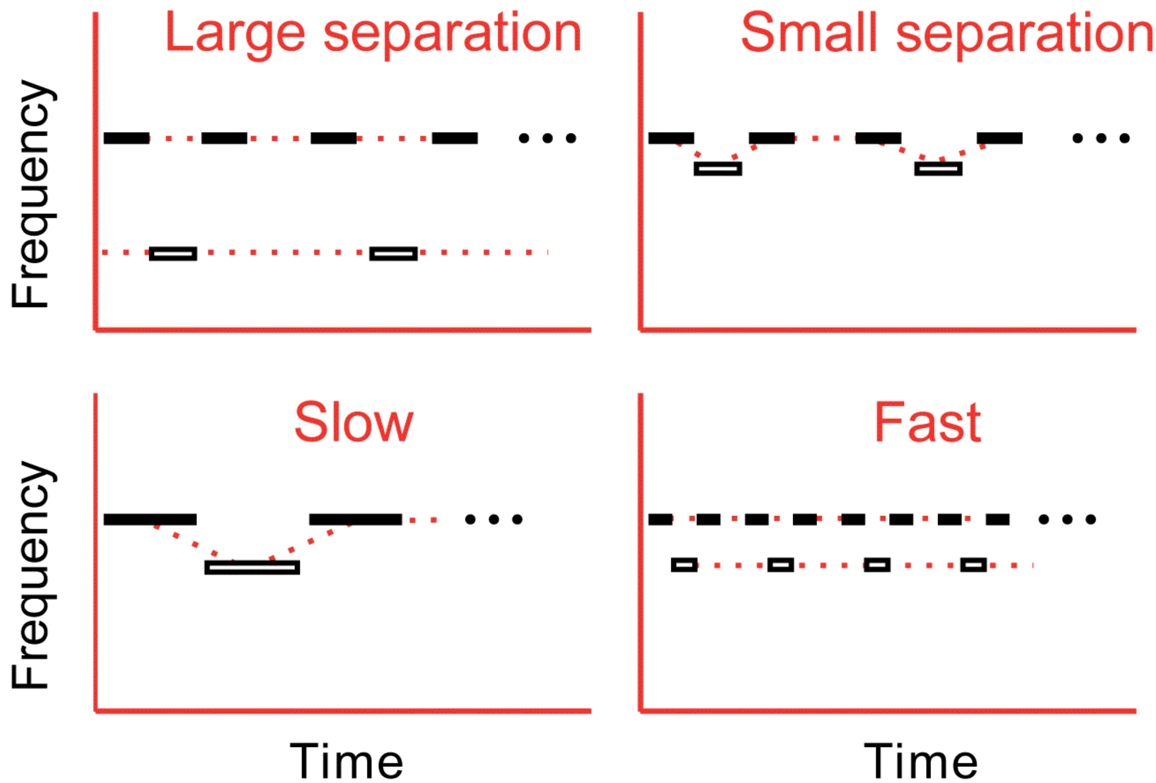
You would think that the study of perception would be exempt from this suspicion, since the subject matter of the psychology of perception is supposed to be about how a person's experience is derived from sensory input. Instead, academic psychology, in its behaviouristic zeal, redefined perception as the ability to respond differently to different stimuli – bringing it into the behaviourist framework. We may be doing research nowadays on cognitive processes, but the research methods are, on the whole, still restricted to behaviouristic ones. Since it was a perceptual experience of my own (the rapid sequence of unrelated sounds) that set me off on a 40-year period of study. of perceptual organization, I have always questioned the wisdom of this restriction.

In my many years of research on how and when a mixture of sounds will blend or be heard as separate sounds, my own personal experience and those of my students has played a central role in deciding what to study and how to study it. When I encouraged students to spend a lot of time listening to the stimuli and trying out different patterns of sound to see which ones would show the effect we were interested in, far into the academic year, and nearing the time that they should have been carrying out their experiments, they would get nervous and ask when they would start doing the "real research". I told them that what they were doing now was the real research, and the formal experiment with subjects and statistics was just to convince other people.

Furthermore the role of subjectivity has often been criticized by journal reviewers: In the reviews of my first published article on auditory stream segregation, which showed that a rapid alternation of high and low sounds segregated into two perceptual streams, one of the skeptical reviewers proposed that there was something wrong with my loudspeakers – perhaps they continued to give out sound after the tone went off – and insisted that I test them.

I was convinced that if the reviewers had merely listened to the sounds, their objections would have evaporated, but in those days you didn't send in audio examples with your manuscript, and I'm not sure it would be acceptable for most journal editors even today. The demonstration I would have included is the one I played to you earlier.

Here is another example of stream segregation based on the galloping rhythm pioneered by Van Noorden., created by repeating a high-low-high triplet: HLH-HLH-HLH …. in a galloping rhythm. When the high and low tones are close in frequency and fall into a single stream, galloping rhythm is heard. However, when the frequency separation is made larger, The galloping rhythm of lost and replaced by two isochronous rhythms, one high and one low in pitch.



Track # 3.

The advantage of using the galloping pattern is that the difference between the galloping rhythm and the simpler ones is easy to hear. For this reason it has been widely adopted for studies of stream segregation, even in the study of nonhuman species.

Anyway, I got around the taboos about subjective data by giving many talks accompanied by auditory examples and by eventually publishing my own Compact Disk of auditory demonstrations. However, the CD didn't come until 23 years after the first research paper. Nowadays you could put demonstrations on the web and refer reviewers to the website.

Another thing that reviewers have criticized was the use of a subjective rating scale, asking listeners, for example, to rate on a 1 to 7 scale how clearly they could hear a sound in a mixture. Perception journals on the whole prefer tasks that involve accuracy. This is in keeping with the behaviouristic view of perception as the ability to make different responses to different stimuli. According to this view, you should be able to score the answers of the subjects as either correct or incorrect (For example by asking whether a particular sound was or was not present in a mixture of sounds) rather than simply accepting the listeners' answers when they rate the clarity with which a target sound can be heard.

Sometimes we have used both types of measures, subjective rating scales and measures of accuracy, either in the same experiment or in a pair of related experiments. The two measures have given similar results, but the subjective rating scales have been more sensitive. I think the reason for their superiority is that they are a more direct measure of the experience, whereas turning one's experience into the ability to form a discrimination between sounds brings in many other psychological processes that are involved in comparison and decision making.

As a result of my belief in experience as an important part of Psychology, I'm going to try to describe some of my research on auditory perception, but I won't give any data. Instead, I'm going to support my arguments with audio demonstrations to the extent that time permits. The track number following the citation of each one refers to its numbering on my CD of auditory demonstrations (Bregman & Ahad, 199

Cues for sequential organization. Let's begin with the question "What features of the acoustic signal lead to stream segregation?" Research has shown that my first intuition was correct. Virtually any perceptual difference that is large enough can lead to the segregation of subsets of tones in a sequence. Also, I found that there was an important interaction with speed: The faster the sequence was played, the stronger the segregation into streams. It had happened, by sheer good luck, that I had made my original sequence of environmental sounds out of 1/10-second snippets of sound, to resemble the average phoneme length in English. It later turned out that this speed was just about optimal for yielding stream segregation.

There are a number of acoustic factors other than frequency separation and speed that promote stream segregation. Among them is timbre.

There are many acoustic variables involved in the timbre of a sound. Among them are:

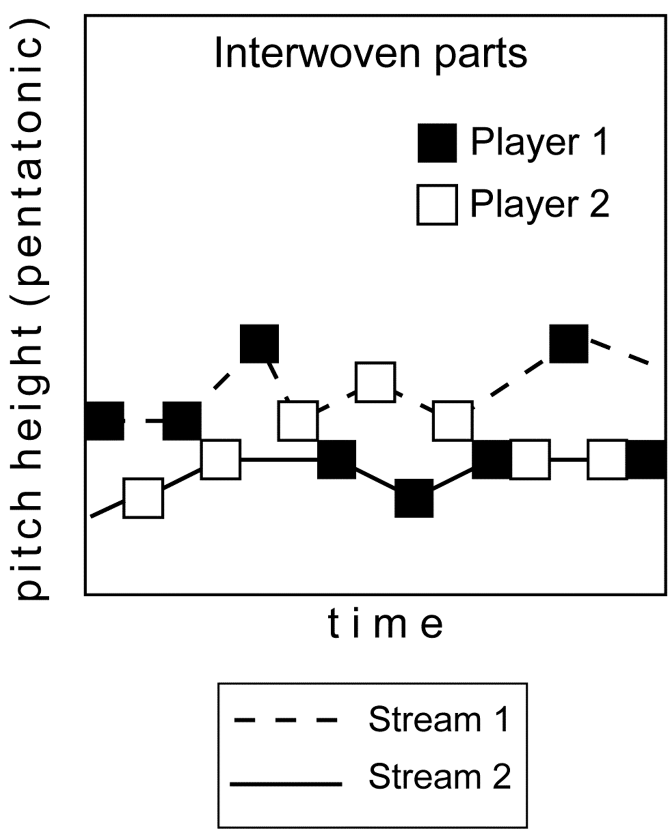
- The shape of the spectrum (e.g., "ah" vs. "ee").
- "Temporal envelope" (e.g.,, abruptness of the onsets)

Apart from timbre, other differences that promote stream segregation are:

- Frequency region (e.g., with noise located in two different frequency bands, the further apart these bands are in frequency, the more easily you can hear the high and low bands as separate sounds).
- Location in space
- Repetition. (Say a loop contains sounds that will segregate. When it first starts to play, it is heard as integrated. After a few cycles, the segregation starts to occur and gets stronger with each repetition.)

Effects of stream segregation. Stream segregation affects the perception of a sequence of sounds in many ways. For example each stream has its own melody and rhythmic pattern that is independent of those in another stream. The following illustration demonstrates this fact. The rhythmic pattern was created on the traditional xylophone of Uganda, by two players, hitting the instrument in strict alternation. Each player plays with an isochronous rhythm (a metronomically regular beat, each tone equally spaced from the one before it and after it). Yet when we listen to the result of the two players playing together, we hear a very complex rhythm. How can two players playing isochronous rhythms yield a complex irregular rhythm? The answer comes from principles of perceptual grouping. We do not hear each player in a separate stream. Because each player plays a full range of pitches, including high and low ones, the streams that are formed depend on the pitch proximities, the high notes of one player grouping with the high ones of the other. Similarly, their lower notes group together. Because of a prearranged irregular pattern of higher and lower notes for each player, the separate high and low streams that are formed each contains its own complex rhythm.

African xylophone music



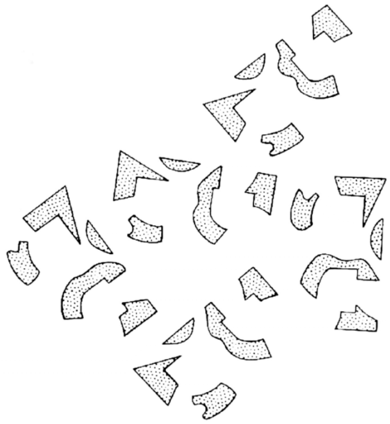
Track #7

Other effects of the formation of separate streams:

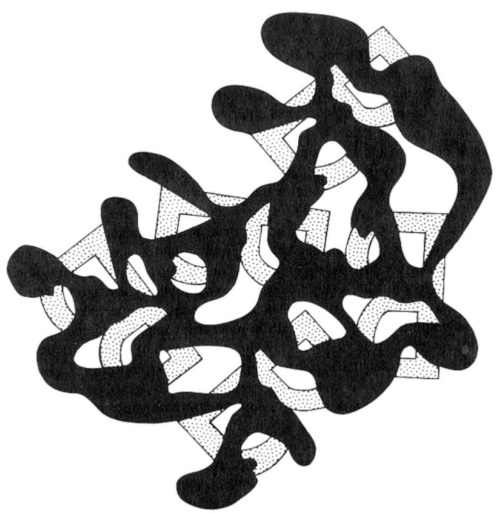
- Pattern recognition is easier within streams. For example, we could play a piece by Bach to one ear of a listener, and a piece by Beethoven to the other. If we asked the listener which note in the Bach immediately followed a particular C# in the Beethoven, this would be a difficult or impossible task. However if we asked which note in the Beethoven followed the C# in the Beethoven, the task would be much easier.
- Fine temporal relations (e.g., the order of the tones in one stream relative to those in another stream) can be lost When I first played a loop of three different high notes and three different low ones to students, with high and low notes strictly alternating with one another, and asked them to report the order, about half of them said that three high notes were followed by three low ones, or vice-versa. This happened because the fine timing relations between auditory streams has lost when the streams segregated..

Scene analysis: visual and auditory. Early in the research, I took the grouping of tones as simply an auditory analog of the grouping of visual figures. But then I asked myself whether this grouping served a function in the life of the individual. I was reminded of the problem in computer vision known as "the scene analysis problem". An example is shown in the next figure which was made by taking some familiar shapes an overlaying them with a highly irregular inkblot, then cutting away all the parts of the underneath forms that are occluded by the inkblot, and then taking the inkblot away. This removes the continuity of the occluded forms, so you don't know which parts to group to make a form. However, all you have to do to restore the perception of the shapes is to put the inkblot back over the forms. The brain understands occlusion and what it does to shapes; so it can now recognize them. This is an example of first a failure, then a success of <u>visual</u> scene analysis.

Figure: Occluded B's   (Fig #004G).



Part 1: (Disconnected fragments)

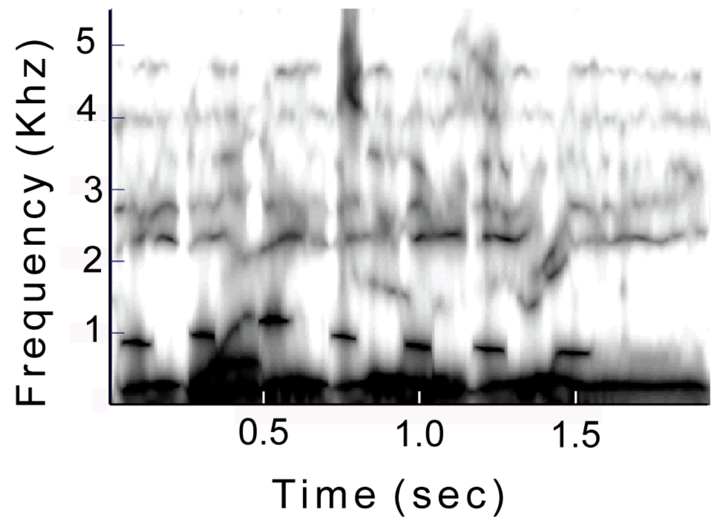


Part 2: Fragments with the occluder superimposed.

In everyday life we frequently see objects that are partially occluded from our vision by closer objects. In the early days of attempting to program a computer, armed with an artificial "retina," was asked to report on the shapes of the individual objects in a stack of objects. They ran into the same sort of trouble that you had when I showed you the visual fragment. Which visible parts should it connect with other visible parts as part of the same object? This was known as the "scene analysis" problem.

An analogous problem exists in auditory perception. It can be appreciated by looking at a spectrogram of a mixture:

Spectrogram of a mixture



Mixture: words "One, two, three"
+ singing: "da-da-da"
+ whistling
+ computer fan

The dark regions show energy at different frequencies. But we know that each component sound could have frequency components over a wide range of the spectrum and also has parts or continuations that occur at different moments of time. The problem of allocating the right temporal pattern of energy to each putative environmental sound can be seen as one of grouping. Our brains must group the right combination of auditory sense data to reconstruct the simple signals that were mixed together. I named this process "auditory scene analysis."

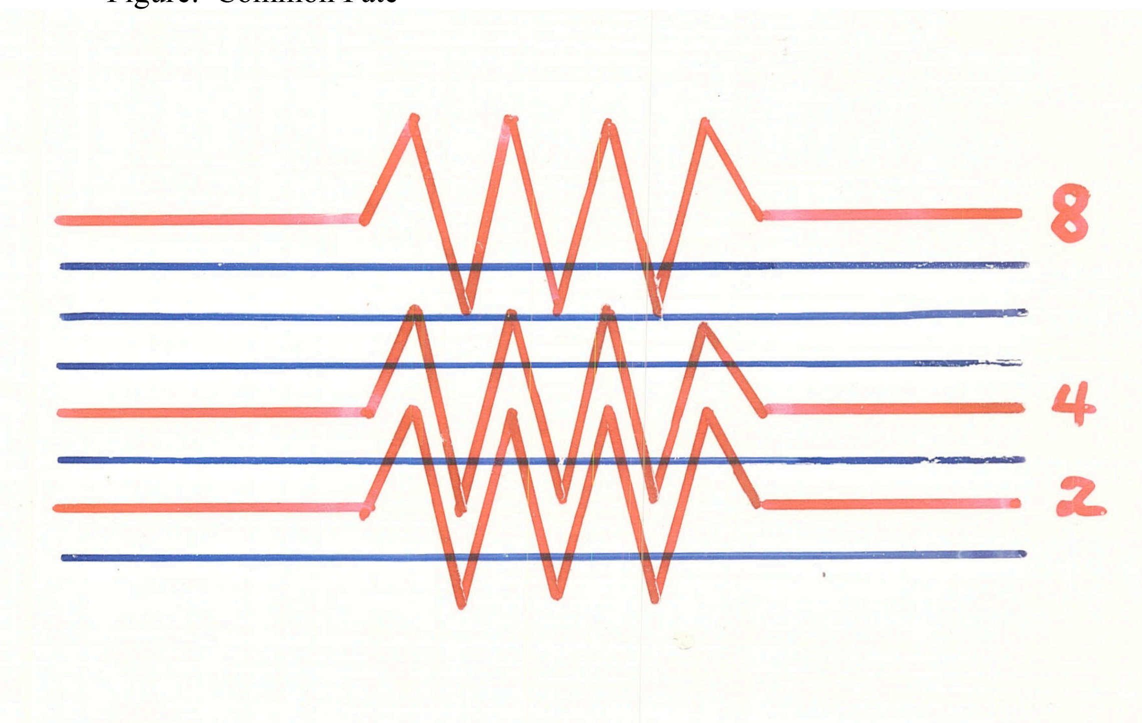
As soon as we recognize that the grouping phenomena are linked to scene analysis, we realize that the research I have described so far has neglected to study the segregation of signals that are partly or completely overlapped in time. In spectral grouping, the goal is to group the simultaneous sensory input as resulting from one or more distinct environmental sounds.

Grouping of simultaneous components. As we began to study this process, we found that the auditory system used a number of features of the signal to decide which components playing at the same time are to be considered parts of the same sound. The features that the brain uses seem to have an ecological basis. Whenever a set of frequency components come from the same sound-producing event (such as a person speaking), they tend to have certain relations between them. Here are some examples:

- When a complex sound begins, all its frequency components tend to start at the same time, and end at the same time. The fact that listeners do use the synchronous onsets and offsets of components to bind them together was established by a series of experiments in the late 1970's and early 80's, beginning with one carried out by Steven Pinker, assisted by Steve McAdams.

- Components arising from the same source tend to change in frequency together preserving harmonic relations within the set of tones. This is illustrated in the next figure and audio demonstration. A complex tone is steady for a while, then three harmonics, the second, fourth and eighth rise an fall in log frequency together, maintaining the frequency ratios among these partials as they change. A visual analog is also presented: the outline of a dog is hidden in a field of short curved line segments. Only when the dog moves, (or the background, but not the dog, moves) as a coherent whole, does the dog stand out from the background. Both are examples of the Gestalt principle of *common fate*, which states that parts will be grouped when they change at the same time in parallel ways.

Figure: Common Fate

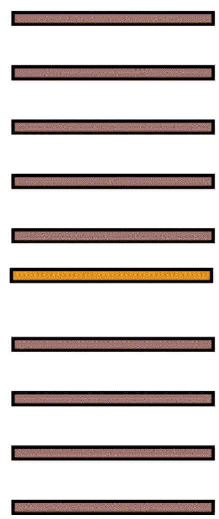


Track #19

Other features that cause partials to group are:

- coming from the same place in space,
- being parts of the same harmonic series. (If one of the partials of a harmonic tone is mistuned enough, it will no longer be hear as simply contributing to the timbre of the global tone, but will stand out as a separate tone.)

Mistuning a component relative to a harmonic series



These regularities and perhaps others, are used by the auditory system to solve the grouping problem.

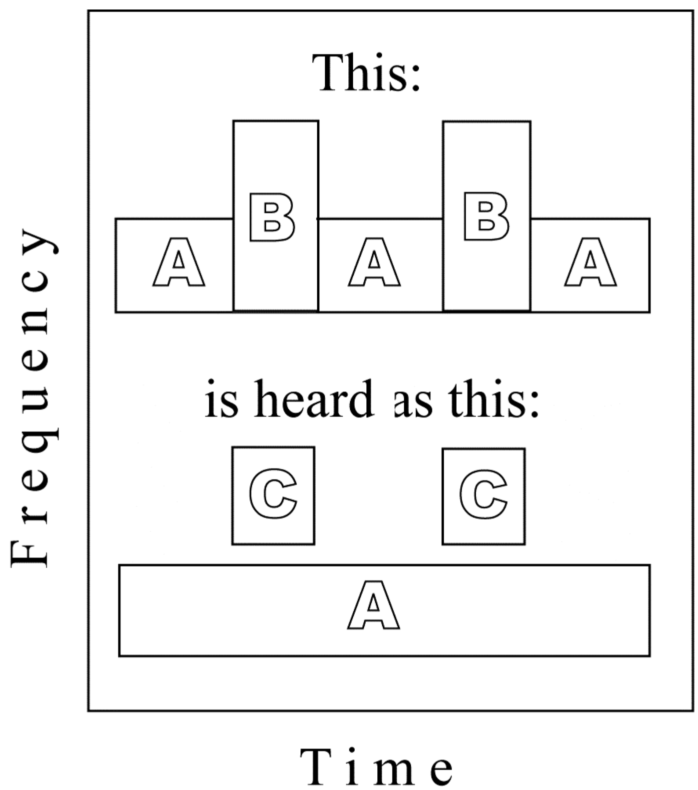
Effects of grouping simultaneous components. The perceptual choice as to whether or not to group different components together has a powerful effect on the perceived properties of the sensory input. For example:

- I might interpret the input as an overlap of simpler sounds, each with its own timbre and loudness, or a single sound with a more complex timbre and a greater loudness..
- I might interpret it a single sound from a single location, or a mixture of two sounds from different locations.

The pitch, timbre, and loudness, and location of the sound that we actually experience depend on how the overlapping energy has been allocated.

Interaction of sequential and simultaneous organization: The "old plus new" heuristic. We can also segregate concurrent sounds by the *old-plus-new* heuristic rule, which goes as follows: "When a spectrum becomes more complex or louder, especially if it changes suddenly, analyse the new spectrum to see if the old signal can be found in it. If so, subtract it out. What remains is the newly added signal. Hear it as a separate sound." If this correctly describes the workings of the auditory system, it is evident that the moment of onset of a sound is critical in separating it from its acoustic background. This process is illustrated in the next demonstration. Two cases are presented. In each one, a 200-ms noise burst that is band-limited to frequencies from 0 to 2000 Hz is presented in alternation with a 200 ms narrower-band noise burst restricted to a 1000-Hz band of frequencies. In the alternation, there is no silence between the 200-Hz-wide band and the narrower-band bursts.

In each case, the wider-band burst contains all the frequencies contained in the narrower-band burst, plus some extra ones. So the old-plus-new heuristic decides that a second sound must have joined an unchanging one. Although the two sounds are alternated in time, we hear a single low sound present throughout. The low burst has captured the lower frequencies of the wider-burst sound into a long unbroken sound. When the wider-band bursts are played, they are not heard as such, but as high sounds that join the unbroken low sound periodically. This is the case shown in the following figure. In the second case, we alternate a burst that has frequencies from 1000 to 2000 Hz with the wider-band one. In this case we hear a continuous high sound, joined periodically by a lower burst sound. Notice that this affects both the number and the height of perceived sounds.



Track #34.

Application to music. All these effects of perceptual grouping can be (and have been) used by composers to control the perception of timbre and rhythm, the separation of melodic lines or layers, and the perception of, or suppression of, the qualities of simultaneous notes such as their harmonies and dissonances.

I've been gratified that certain music theorists have found these ideas worth pursuing: Three come immediately to mind: Stephen McAdams, David Huron and Rosemary Mountain.

In conclusion let me say that the association I have had with people studying music over the years has been very rewarding. Among these people are psychologists, three in my own department – Dan Levitin, Caroline Palmer, and Robert Zatorre, as well as others in different universities, Carol Krumhansl, Diana Deutsch, Yoshitaka Nakajima, and Lola Cuddy. Others have been computer scientists, such as Dan Ellis and DeLiang Wang, composers such as John Chowning and Chris Chafe, and music theorists such as Eugene Narmour, James Wright and Fred Lehrdahl. If my work has been of use in the growing interdisciplinary field of Music Science, I am very satisfied.

References

Bregman, A.S., & Ahad, P. (1996) *Demonstrations of Auditory Scene Analysis: The Perceptual Organization of Sound.* Audio compact disk. (Distributed by MIT Press).

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.*

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