AUDITORY SCENE ANALYSIS AS A SYSTEM

Albert S. Bregman
Dept. of Psychology, McGill University, 1204 Penfield Ave., Montreal, QC, Canada H3A 1B1. al.bregman@mcgill.ca

1. INTRODUCTION

A serious problem faced by any listener is that the ears receive a mixture of all the environmental sounds that are present at a given moment. Yet in order to generate appropriate responses, the auditory system must be able to build representations of the individual sounds that have created the mixture; this accomplishment is called auditory scene analysis (ASA). It seems to be accomplished by a two-stage process. First it analyses the incoming signal into its frequency components, both at a given time and extending over time. Then bottom-up processes of perceptual grouping use various acoustic relations among the components to build up evidence favouring the grouping of certain subsets of components, each subset representing a single environmental sound, with its own spectral and temporal properties (Bregman, 1990). Top-down processes use these "grouping recommendations" in building a representation of the streams.

1.1 Laboratory phenomena related to ASA.

There are a number of phenomena, including stream segregation, illusory continuity, fusion and decomposition of complex sounds, which are best viewed as glimpses of a single, coherent ASA system. Stream segregation is typically studied by alternating two types of tones (call them A and B). When the difference in the feature that distinguishes A from B tones is large enough and the speed of the sequence is fast enough, the listener perceives two parallel but independent streams of sound, one restricted to the A tones and the other to the B tones. Differences between A and B tones can be in terms of frequency (for pure tones), or in pitch and timbre (for complex tones), or in how the properties of a tone change over time, or for separation in the spectrum (for band-limited noises), or where they seem to come from in space (for all tones).

Illusory continuity is the apparent continuity of a sound, A, through a loud interruption, B, despite the fact that A is turned off during the interruption. It is stronger when the interrupting sound, B is shorter, and when it would have masked A even if A had actually been present during the interruption. It is viewed as a perceptual compensation for masking, a way of dealing with a sonic environment composed of many sounds, where one can temporarily mask another.

Fusion and decomposition of complex sounds occurs when many different frequency components are detected at the same time in a sensory input. Fusion into a single sound is favoured by harmonic relations among the components, a common spatial origin, their proximity in frequency and correlation in how they change over time.

These phenomena are best viewed as glimpses of a single, coherent system. The alternative view would see them as distinct, each with its own physiological basis.

2. PHYSIOLOGICAL INVESTIGATION

From a physiological perspective, we may well discover analysis mechanisms for the different acoustic features that favour the grouping of components. For example, Fishman, Arezzo, and Steinschneider (2004) presented an ABAB... sequence to awake monkeys while neural activity was recorded in primary auditory cortex (A1). Using pure tones, they varied the frequency separation of the A and B tones, the tone presentation rate, and the duration of each tone. Recordings were made at the cortical site which responded maximally to tone A (the "A-site").

In human psychophysics it has been found that as the speed of the sequence is increased, the segregation of the ABAB… sequence into A and B streams becomes stronger. In the cortex of monkeys, at slow speeds the A-site responded also to the B tones (but to a lesser degree). The greater the frequency difference between A and B, the less the A-site responded to B. But more interestingly, as the tone rate increased, while the response at the A-site to the A tones was somewhat reduced, the A-site's response to B was even more greatly reduced, so that the A-site yielded a neural response pattern dominated by responses to the A tone, occurring at half the alternation rate. In other words it becomes more selective in favouring the A stimuli.

2.1 Interpretation of the results

These neural activities were taken by the. authors as an important cause of perceptual segregation. In effect the A-site "sees" the A stream only (a B-site would see only the B stimuli). They also pointed out that since no response was required of the monkeys, the research was studying an automatic or obligatory process.

One is tempted to conclude that the physiological mechanism of stream segregation has been discovered. One of the difficulties with such a straightforward interpretation is that the results depend on the use of pure-tone stimuli. Because

the earliest research used pure tones as A and B, and manipulated their similarity be making them different in frequency, most of the physiological and animal research has also used pure tones. As a consequence, the explanations are often specific to pure-tone stimuli.

However, A and B can be made different in many other ways. For example they can appear to emanate from different spatial locations, e.g., by the manipulation of interaural differences in time of arrival, or be different in loudness, factors that will promote segregation. If A and B are complex periodic tones, a difference in their pitches can also favour segregation. Also for these tones, differences in the placement of formants (peaks in their spectra) can favour their segregation. It is even true that A and B tones can have the same pitch and loudness, come from the same location, have components located in the same spectral region, with identical amplitude spectra and still be made to segregate – by having the phase relations among the components of each tone be different, altering their timbres.

One might argue that for each feature there are separate cortical sites that respond to different values of that feature, each one working in the same way as the frequency sites described by Fishman et al. (2004). The actions of any or all of these sites could produce stream segregation. But this argument does not take into account the fact that the various sorts of differences between A and B tones tend to interact. Any particular acoustic difference only promotes, rather than directly determining, the grouping of components to represent individual environmental sounds. The more factors that favour a particular grouping, the stronger the total evidence for that grouping will be. The various acoustic differences can also compete with one another, some favouring one grouping, and others favouring different ones. It seems that in such a case the grouping with the most evidence in its favour will be the one that is perceived.

In addition to the interaction of the various bottom-up acoustic analyses, these also interact with top-down processes, such as knowledge of the class of signal involved. Such stored knowledge can influence the attentional processes that participate in the building of the perceptual representations of individual streams of sound. The overriding role of top-down processes is particularly evident in the case of speech, where sine-wave-analog speech, despite being a stripped-down cartoon of speech, can still be understood, despite the fact that some of the bottom-up acoustic cues favouring integration of the components of the "speech" signal are missing.

So one cannot say that any particular acoustic relation, such as the frequency difference studied by Fishman et al. (2004) is *the* physiological cause of grouping, even in the simple laboratory example of the alternation of two tones.

3. ASA AS A COHERENT SYSTEM

In the introduction, we described three different phenomena as being glimpses of the action of the ASA system: stream segregation, perceptual fusion of simultaneous components, and illusory continuity. We can consider the argument that each of these has its own distinct physiological basis.

One argument against it is that these three phenomena respond to the same variables. For example, two narrow-band noise bursts, A and B, can be created where A has a higher pass band than B, and these can be alternated, each burst separated from the next by a wideband noise (W), in the pattern AWBWAWBW…. (Bregman, Colantonio & Ahad, 1999), We can use this stimulus to look at both stream segregation and illusory continuity. When the centre frequencies of the A and B bands are close together they will not segregate, and will also appear to connect up behind the wide-band interruption, yielding illusory continuity. When their centre frequencies are further apart, both segregation and continuity are affected. A and B are heard as separate streams, and also fail to connect up behind the noise. We have argued that both stream segregation and the continuity illusion are products of a single ASA system. We have also argued that the process called the "old-plus-new heuristic" can explain both illusory continuity and the decomposition of the spectrum into separate sounds (Bregman, 1990)

The question is whether the phenomena described in this paper result from the activity of an integrated ASA system or are the result of a haphazard set of physiological processes, each perhaps having evolved independently to favour the correct parsing of the incoming sound. There are a number of arguments for preferring the view of ASA as a coherent system. They include (1) the desire for parsimonious explanation: The alternative is to believe that a bundle of unrelated phenomena just happen to be that way for idiosyncratic physiological reasons. This is an unparsimonious explanation – a bristly hypothesis that needs a shave with Occam's razor; (2) the fact that these phenomena result from processes that serve a common function; (3) the fact that they interact in ways that are predictable under the assumption of a unified system; (4) that they respond to many of the same variables; and (5) the heuristic value, for researchers, in finding the factors that may affect all of the various grouping phenomena..

REFERENCES

Bregman, A.S. (1990). *Auditory scene analysis: the perceptual organization of sound.* Cambridge, Mass.: The MIT Press, 1990 (Paperback 1994).

Fishman, Y.I., Arezzo, J.C., & Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J. Acoust. Soc. Am. 116*, 1656-1670.

Bregman, A.S. Colantonio, C. & Ahad P.A. (1999). Is a common grouping mechanism involved in the phenomena of illusory continuity and stream segregation? *Perception & Psychophysics, 61 (2),* 195-205.