Three directions in research on auditory scene analysis

Albert S. Bregman

Citation: Proc. Mtgs. Acoust. **19**, 010021 (2013); doi: 10.1121/1.4799217
View online: https://doi.org/10.1121/1.4799217
View Table of Contents: https://asa.scitation.org/toc/pma/19/1
Published by the Acoustical Society of America

Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Animal Bioacoustics
Session 2pAB: Listening in the Natural Environment

2pAB1. Three directions in research on auditory scene analysis

Albert S. Bregman*

***Corresponding author's address: Psychology, McGill University, 5710 Melling Ave., Cote Saint-Luc, H4W 2C4, Quebec, Canada, al.bregman@mcgill.ca**

Research on auditory scene analysis (ASA) began with some simple laboratory phenomena such as streaming and illusory continuity. Subsequently, research has gone in three directions, downwards towards underlying mechanisms (by neurophysiologists), upwards towards system organization (by computer scientists), and sideways towards other species (by neurobiologists). Each direction has its challenges. The downward approach sometimes takes a phenomenon-oriented view of ASA, leading to simple explanations of a single ASA demonstration, such as streaming, with no obvious connection to any larger system. Research done by the upward approach usually takes the form of a computer program to achieve ASA in a working system, often neglecting known facts about human ASA, in favor of mathematically understood principles. The sideways approach often finds that non-human animals can respond to an important sound despite the presence of other interfering sounds. However, there is no reason to believe that a frog, a fish and a human accomplish this by means of the same mechanisms. So finding out how some animal does this, while interesting in its own right, may shed little light on how humans do it. I will describe some properties of the human ASA system that should be borne in mind when manufacturing explanations.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4799217]
Received 19 Jan 2013; published 2 Jun 2013

Since the introduction of the term in 1984, auditory scene analysis (ASA) has come to mean different things to different people, and to be pursued in different directions. The research began with some simple laboratory phenomena such as streaming, illusory continuity, and perceptual fusion of concurrent sounds. Subsequently, research has gone in at least three directions, downwards towards underlying mechanisms (by neuroscientists), upwards towards system organization (by computer scientists), and sideways towards other species (by neurobiologists). I would like to describe some of this research and make some suggestions about future research.

Neuroscientists

Neuroscientists have taken a number of approaches. One is simply to measure the large-scale electrical activity of the brain to determine when (and sometimes where) some aspect of auditory scene analysis takes place. Two phenomena have been pursued in this way.

Elyse Sussman and her colleagues have developed a technique using the event-related potential (ERP) to determine whether or not a subset of tones has become an integrated stream [1]. Claude Alain and his associates have found a marker in the ERP that indicates whether or not concurrent tonal components have been integrated into a global sound [2]

Of course, measurement techniques are useless unless they discover new facts. Sussman's technique has shown (a) that contrary to the idea that sound is organized into a single foreground stream and an undifferentiated background [3], at least three organized streams can be formed at the same time [4], (b) that whereas attention strengthens the formation of an auditory stream, it is not essential for its formation (e.g., [5]), (c) that newborns, a few days of age, already form their auditory inputs into integrated streams [6], and (d) that top-down effects can modify an initially stimulus-driven auditory organization [7].

Claude Alain's technique has yielded a number of facts about auditory organization, for example that there are both bottom-up and top-down influences on the perceptual organization of concurrent sounds [8], and that different neural circuits are involved in the perceptual integration of a sequence of sounds on the one hand and the integration of sets of simultaneous sounds on the other [9].

One of the advantages of using event-related potentials to measure organization is that the fine-grained time course of auditory organization can be observed, making it possible to differentiate between aspects of organization that are entirely stimulus-driven and those that operate in a top-down manner. Also, since the influences of attention on the ERP are well-known, it can be confirmed that stream segregation can occur without attention to the details of the signal.

I have mentioned the research of Sussman and Alain because both of them have been clear about the fact that the phenomenon that they are focused on in a particular experiment is part of a larger system, and that being able to find the influences of the experimental conditions on a particular phenomenon is not the end of the story.

So far I have described the measurement of large-scale brain activity in the form of event-related potentials to study auditory organization. However, there is a more fine-grained approach that looks at the activity of individual neurons during auditory organization. Since this has to be performed on non-human animals, before I consider these experiments I will discuss the study of auditory scene analysis in non-humans.

ASA in Non-humans

One of the pioneers in this field was the late Stewart Hulse, who, with his associates at Johns Hopkins, studied the auditory abilities of European starlings. They discovered that the birds could distinguish between song segments from two different starlings, even in a background of four other starling songs, a clear demonstration of auditory scene analysis [10]. The Johns Hopkins group went on to use sequences of pure tones, alternating in frequency, to show that the starlings formed auditory streams, defined by frequency, in the way that humans do [11]. Research on goldfish has shown, using classical conditioning, that they are able to hear a mixture of two pulse trains – each with its own spectral profile and repetition rate – as two separate streams [12]. Japanese monkeys also form streams, grouping together tones that lie within the same frequency range [13]. Moss and Surlykke have also found ASA in the echolocation of bats [14]

There are two kinds of studies on nonhuman animals, which have different implications. The first type is naturalistic, showing that the animal can detect, among a mixture of signals, one that is significant in its own life. The second is analytic, using simple, well-defined signals, such as pure-tone sequences, and demonstrating that the animal forms streams in much the same way as people do.

Studies that illustrate an animal's ability to detect important signals from their own species in a background of interfering sounds show that *some* process of organization exists, but not how it works in the brain. So we don't know whether it resembles ASA in other species or in people. For example female frogs can pick out the mating call of an individual male and move towards it despite a dense chorus of calls from other frogs and toads. Separation in space between the target and other sounds helps, just as it does in humans [15]. No doubt this capacity of frogs to segregate some sounds from others may have the same advantage as ASA does in humans and other species. However, it is not necessary to believe that the neural processes and the methods that they embody to achieve ASA are the same as in other species. Some simpler animals may just employ a sort of neural filter, tuned to the distinctive features of the particular sound of interest, such as a mating call, and perhaps tune this filter, in real time, to the call of a particular conspecific. Of course, we humans can also filter sounds, using our knowledge of the properties of specific sounds to select them from a mixture, but can do so for a great number of classes of sounds, with which we have become familiar, such as words in our own language, the sound of an automobile engine, and the barking of our own dog – a process that I have called "schema-based" ASA. But in addition, we employ stimulus-driven, Gestalt-like organizing processes that are useful for any class of sounds. Other animals may have bottom-up processes as well, but the set of methods used by one species may be different from the set used by another. This has to be discovered on a case-by-case basis until generalizations emerge. One could offer, as a first hypothesis, that as the lifestyle of the animal becomes more varied, and its brain becomes more complex, the ASA mechanisms become less specialized (i.e., less single-purpose).

Neural Basis of ASA

One important reason for studying auditory organization in non-human animals is that it allows the neural processes underlying ASA to be studied directly, and some animal studies have done so, exposing some details of neural activity in ASA.

Keller and Takahashi, for example, studied the ability of barn owls to segregate two simultaneous noise bursts located at different points in space [16]. They showed that specific neurons in the inferior colliculus responded to one of the sound sources whenever it was within their receptive fields, and registered the temporal changes of that sound source, disregarding temporal changes in the other sound source.

The neural basis of stream segregation has been studied in the primary auditory cortex of awake macaque rhesus monkeys (e.g., [17], [18]) and in the auditory forebrain of awake European starlings [19]. In each of these studies, the researchers first determined the best frequency, A, for a selected neuron (call it the A neuron). They then presented a sequence of alternating tones, A and B, to the animal, varying the frequency of the B tones. These three experiments, as a group, found that the firing of the A neuron to tone B corresponded to some behavioral phenomena: (a) it became weaker with greater frequency separations between the A and B tones, which is not surprising, but this could make it the neural basis of a stream of A tones. (b) the response of the A neuron to tone B decreased with the speed of the sequence [18] [19] and with more repetitions [19], just as a stream of A tones segregates more from the B tones as the sequence get faster, and with more repetitions. Since the observed narrowing of the tuning of Neuron-A responded to some of the same variables as does the perception of a separate A stream, this suggested to the researchers that they had found an important neural substrate of stream segregation, the A neuron representing a single stream of A tones.

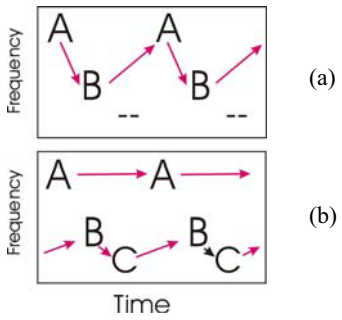

FIGURE 1. Panel (a) shows a repeating cycle of three elements, Tone A, Tone B, and a silence. In Panel (b), a third tone, C, replaces the silence. The red lines connect the elements that are heard as being in the same stream.

It seems that the narrowing of the spectral sensitivity of a single cell can explain some of the facts about the streaming of alternating high and low sounds, but as soon as the situation gets a bit more complex, this explanation proves insufficient. Figure 1, Panel a, shows two iterations of a repeating cycle consisting of pair of alternating tones, a high-frequency A and a lower-frequency B, with a silence, equal to the length of one tone, coming right after B. The red lines connect tones that are in the same stream. A and B are just close enough in frequency to form a single stream. This means, according to the neural explanation, that the A neuron must be responding not only to A but to B. However, let's add a third tone, C, to the cycle, as in Panel b, replacing the space that comes just after B, with C being below B in frequency. If C captures B into a BC stream, B may no longer group with A (c.f., [20]). According to the explanation in terms of a change in the A neuron's receptive field, this means that the A neuron no longer responds to B. In metaphorical terms, B has been captured by C into its own BC stream. Yet no acoustic change has taken place in the frequency range to which the A neuron responds; so there is no reason to expect it to stop responding to B. Perhaps neuron *B* may stop responding to A due to the narrowing of its receptive field induced by Tone C. However, the continued response to both tones by neuron A should keep the AB stream going, unless the B neuron can somehow tell the A neuron that it is no longer responding to tone A, so that the A neuron might just as well stop responding to B.

In other words, there would have to be a temporary connection between neurons A and B, representing their presence in the same stream, a connection that can be broken by either neuron. In general, if there is a competition among the frequency proximities of all the sounds that are close to each other in time, and this competition determines the perceptual grouping, there would have to be extensive communications among the neurons that responded to the different sounds.

All this presupposes that stream segregation is fully determined at a fairly early stage in the brain's processing of sound. However, we know that top-down processes can influence stream segregation. For example, if listeners are told to focus on Tone A, they can succeed in doing so over a wide range of frequency separations between tones A and B, and of rates of alternation [21]. In a more naturalistic setting, a voice speaking one's own language is easier to pull out of a mixture of voices than one speaking a foreign language.

The neuron-sensitization theory has the limitation of having been built on a single laboratory effect, the stream segregation that occurs when two tones alternate. If we continue with this approach, looking for a neural correlate for each simple laboratory phenomenon, we will miss the larger picture – how a system of neurons can work together.

CASA

As I said earlier, the upward direction of research, towards system organization has mainly been carried out by computer scientists under the name *computational auditory scene analysis* (CASA). The definition of CASA is "the field of computational study that aims to achieve human performance in ASA by using one or two microphone recordings of the acoustic scene" [22]. In other words, CASA researchers want to solve the "cocktail party problem" posed by such early students of speech recognition as Cherry [23]. However, this general goal has often taken the specific form of having a computer take in an acoustic mixture of the sounds from two or more acoustic sources, such as a person talking and a siren sounding, and outputting separate signals, each of which contains the sound from only one of the sources. While the idea is to then feed each of the separated signals into a recognition process, in practice the recognition is usually done by a human listener, who judges whether the segregated signal representing the target voice is easier to understand than the original mixed signal. The preliminary separation is intended to prevent the errors that occur when a signal that is actually a mixture is treated as coming from a single source. The constraint of working with actual acoustic signals and achieving a demonstrable result has forced the researchers to consider the properties of the system as a whole.

However, the intention of emulating the human achievement of ASA is not the same as a commitment to the exclusive use of the methods of the human auditory system to do so. Some approaches to CASA do use some of the methods used by humans, but these human-like approaches are often subordinated to the goal of getting a system that actually works and is achievable on existing computers. It had been hoped that the development of CASA would test the adequacy of the methods proposed for human ASA [24], but to the extent that emphasis shifts away from human methods, this becomes less likely.

An important fact about human ASA is that it employs a variety of redundant cues for separating mixtures. In an environmental situation in which some of these are unavailable (e.g., when there are no harmonics as in whispered speech, or when sound has traveled around corners) or some cues have been distorted (e.g., by reverberation), the system doesn't simply stop functioning, but utilizes whatever cues are available to the extent possible. In other words, it degrades gracefully. I don't think the criteria of interchangeability of cues and of graceful degradation have

always been given the importance they deserve in CASA. Graceful degradation is possible when a number of methods, running in parallel, are collaborating to achieve a result. An early example was the Hearsay-II model of Raj Reddy [25] . It was based on the idea that if several processes are to collaborate in arriving at a result, the system has to have a shared data structure that represents the current hypotheses, so that the processes can work together to support or disconfirm them. This common data structure has been called a *blackboard*. Such a structure has been used in CASA by Dan Ellis in his doctoral research to allow multiple bottom-up cues and a top-down model of speech to collaborate [26]

Do our brains use blackboards in solving the ASA problem? This is equivalent to asking whether a number of parallel analyses converge on a representation that merges their results and allows them to cooperate and compete in the allocation of sounds to streams.

If the Gestalt psychologists are to be believed, any brain region that acts as a "blackboard" should be an active process, not merely a passive medium that records the results of other brain operations. For example, suppose each part of the brain region represented a different combination of values of frequency, time, and intensity, in a three-dimensional array, where proximity in the brain reflected the nearness between these represented values. The elements of the "blackboard" would not merely summarize the results of analyses of the signal performed by other processes, but would reach out to nearby elements at the same level of abstraction – to nearer ones more strongly than to more distant ones – and build alliances with them, forming units and streams. Unfortunately, the problem with implementing a Gestalt "field" in the brain by using distance to represent similarity is that there can be large number of bases for similarity among pieces of sound (fundamental frequency, spectral centroid, spatial location, intensity, harmonicity, roughness, onset abruptness, and perhaps more), and these properties would have to translate into a correspondingly large number of dimensions of distance in the brain. It is not clear how this could be achieved in a three-dimensional brain.

If a neural blackboard is to represent auditory streams, what would this require? There are two jobs that a stream does: (1) it asserts that a particular set of sounds belongs to it, and (2) it acts as an entity that can itself have emergent properties (such as a melody or a rhythm) that are not properties of the individual component sounds, but of the stream as a whole. The first job, uniting a set of sounds, is not so hard. Neural representations of the individual sounds could be temporarily given a common property, such as (a) being connected, as a group, to the same element at a stream-representing level of the auditory system, or (b) each sound being represented by an oscillating circuit, and the set of circuits oscillating in a common phase if they are parts of the same stream, as Deliang Wang has described [27], expanding the ideas of van der Malsburg [28].

However the second job, acting as the basis for form-extracting processes such as those that find rhythms or melodies, requires that the stream not only group the individual sounds, but preserve their acoustic and temporal properties. A single property, such as being connected to a stream-identifying element, or having a common form of oscillation, could not, in itself, serve as the basis for the extraction of structural properties. A blackboard would preserve the identities and order of the individual sounds, while at the same time marking them as belonging to a common stream. Furthermore, it would represent all the current sounds, not just the ones of a single stream, so that other processes – perhaps those concerning familiar melodies or speech sounds – could revise the stream membership if the evidence was strong enough to warrant such a change. In a word, assigning a description, such as stream membership, must not act as a barrier between the lower-level features and higher-level analyses. Such use of a blackboard prevents the system from being strictly hierarchical, and provides a meeting ground for top-down and bottom-up interpretations of the signal.

Ears in Motion

Both CASA and the study of ASA in humans have banished an important source of information. This information will become more critical when we evolve from present-day information technology, in which a computer that sits still is asked to deal with a mixture of acoustic sources that also sit still, and progress to a future technology in which a moving robot is required to deal with a mixture of sounds from moving sources (in order, for example, to serve drinks at a cocktail party). At first glance, introducing motion seems to make the ASA problem harder, but in fact it may make it easier. As humans or robots move their heads and bodies, this will induce changes in the mixture of sound received at their ears (or microphones) in systematic ways that will contribute to scene analysis. For example, those spectral regions that all get louder together, as the listener moves, probably come from the same environmental source or a tight cluster of sources. The classical psychophysical approach – fixing the head of subjects as they listen to sounds – treats head movement as a source of error to be eliminated. But in ASA, head movement contributes information that is undoubtedly exploited. One hopes that researchers will liberate the head of the listener and show exactly how motion information is used, not just for resolving ambiguities in perceived

location but in separating sounds. Head motion, and its changing effects on the head shadow, should be helpful for ASA with even a single ear or microphone.

Questions about ASA

The final part of this paper mentions some questions that should be answered by any theory of ASA, whether derived from a comparison among species, embodied in an array of neurons, or in a computer system. What are some of these questions?

- What mechanism is responsible for the fact that patterns such as melodies, words, or rhythms tend to be formed from elements that have been assigned to the same primitive stream?
- Why are judgments of timing more precise when they involve a comparison of elements that fall within the same auditory stream that when they involve elements that are in different streams?
- In synthetic speech, if you change the fundamental frequency in the middle of a syllable such as "wa", why does the part after the change sound like it begins with a stop consonant? [29]
- Why does camouflage work? Familiar patterns that one is listening for, such as melodies, can be made very hard to hear if they are broken up by bottom-up grouping, even if one's attention is trying its best to hear the pattern? In other words, how does the interpretation based on raw acoustic features interact with the interpretation based on attention and stored schemas?
- Is there a "blackboard" (shared data structure) that makes this interaction possible? If so, what are the properties of this data structure?
- How does the "old-plus-new heuristic" work? This is a method that comes into play when an ongoing sound or mixture of sounds is joined by a newly arriving one. Suppose that a spectrum – call it the "old" spectrum – suddenly becomes more complex, energy having been added at various frequencies to form what we can call the "changed" spectrum. The auditory system analyses the signal to determine whether the old spectrum is still there, as part of the changed spectrum. If it is, the system partitions the changed spectrum into two perceived parts: (i) the old sound, which is perceived to continue right through the change, and a "new" sound that comes on at the moment of change, supplying the spectral energy which, when added to the old spectrum, yields the changed spectrum. In other words, the changed spectrum is heard as the sum of two spectra, a continuing old one and an added new one, this new one actually being a residual, derived by subtracting the spectrum of the old sound from the changed spectrum. The illusory continuity of a sound, when a short part of it is replaced by a loud noise burst, studied extensively by Richard Warren and his associates (e.g.,[30]), is an example of the "old-plus-new heuristic" at work. A blackboard – either in a classroom or in a brain – makes it easier to describe such a process. A simple hierarchical arrangement of neural analyzers has a much harder time. How are the properties of the residual isolated from the properties of the changed spectrum as a whole? Why, after the change, do we hear two sounds and not just one? How could a neural system compute the spatial position of a residual? In forming the residual, does the neural process remove, from it, all the features (and energy) that it has allocated to the older sound? How does it deal with the problem of allocating spectral energy at a certain frequency to the old and new sounds?
  While I have framed this question in terms of the decomposition of a mixture of single, complex sounds, the same sort of reasoning applies to mixtures of sequences, when a new sequence, such as a spoken utterance, joins another one that was already in progress.

The point that I am trying to make with these questions is that the understanding of ASA involves much more than the understanding of how auditory streams are formed by the alternation of high and low tones in the laboratory. Explanations of ASA – be they in terms of brain processes, computer systems, or the evolution of the nervous system – need to be tested against a wide range of facts about the perceptual organization of sound. And any claim that primitive ASA in non-humans corresponds to primitive ASA in humans also needs to be tested against a wide range of phenomena to see how far the correspondence holds up.

ACKNOWLEDGMENTS

The author's research has been supported by the Natural Sciences and Engineering Research Council of Canada for the past 47 years. The assistance of Pierre Ahad for more than two decades of this period is gratefully acknowledged.

REFERENCES

1. E. Sussman, W. Ritter and H. G. Vaughan, "An investigation of auditory stream segregation using event-related brain potentials.," *Psychophysiology,* vol. 36, pp. 22-34, 1999.
2. C. Alain, S. Arnott and T. Picton, "Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials.," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 27, no. 5, pp. 1072-1089, 2001.
3. R. Brochard, C. Drake, M. Botte and S. McAdams, "Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation.," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 25, pp. 1742-1759, 1999.
4. E. Sussman, A. Bregman, W. Wang and F. Khan, "Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments.," *Cognitive, Affective, & Behavioral Neuroscience,* vol. 5, no. 1, pp. 93-110, 2005.
5. I. Winkler, E. Sussman, M. Tervaniemi, J. Horváth, W. Ritter and R. Näätänen, "Preattentive auditory context effects.," *Cognitive Affective and Behavioral Neuroscience,* vol. 3, pp. 57-77, 2003.
6. I. Winkler, E. Kushnerenko, J. Horváth, R. Ceponiene, V. Fellman, M. Huotilainen, R. Näätänen and E. Sussman, "Newborn infants can organize the auditory world.," *Proceedings of the National Academy of Sciences,* vol. 100, no. 20, pp. 11812-11815, 2003.
7. I. Winkler, R. Tagegata and E. Sussman, "Event-related brain potentials reveal multiple stages in the perceptual organization of sound.," *Cognitive Brain Research,* vol. 25, no. 1, pp. 291-299, 2005.
8. C. Alain, S. R. Arnott and T. W. Picton, "Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 27, no. 5, pp. 1072-1089, 2001.
9. C. Alain and A. Izenberg, "Effects of attentional load on auditory scene analysis.," *Journal of Cognitive Neuroscience,* vol. 15, no. 7, pp. 1063-1073, 2003.
10. A. Wisniewski and S. H. Hulse, "Auditory scene analysis in European starlings (Sturnus vulgaris): Discrimination of song segments, their segregation from multiple and reversed conspecific songs, and evidence for conspecific song categorization.," *Journal of Comparative Psychology,* vol. 111, no. 4, pp. 337-350, 1997.
11. S. MacDougall-Shackleton, S. Hulse, T. Gentner and W. White, "Auditory scene analysis by European starlings (sturnus vulgaris): Perceptual segregation of tone sequences.," *Journal of the Acoustical Society of America,* vol. 103, no. 6, pp. 3581-3587, 1998.
12. R. Fay, "Auditory stream segregation in goldfish (Carassius auratus)," *Hearing Research,* vol. 120, no. 1-2, pp. 69-76, 1998.
13. A. Izumi, "Auditory stream segregation in Japanese monkeys.," *Cognition,* vol. 82, pp. B113-B122, 2002.
14. C. Moss and A. Surlykke, "Auditory scene analysis by echolocation in bats.," *Journal of the Acoustical Society of America,* vol. 110, no. 4, pp. 2207-2226, 2002.
15. A. Feng and R. Ratnam, "Neural basis of hearing in real-world situations.," *Annual Review of Psychology,* vol. 51, pp. 699-725, 2000.
16. C. H. Keller and T. Takahashi, "Localization and identification of concurrent sounds in the owl's auditory space map.," *Journal of Neuroscience,* vol. 25, no. 45, pp. 10446-10461, 2005.
17. Y. Fishman, D. Reser, J. Arezzo and M. Steinschneider, "Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey.," *Hearing Research,* vol. 151, pp. 167-187, 2001.
18. C. Micheyl, B. Tian, R. P. Carlyon and J. P. Rauschecker, "Perceptual organization of tone sequences in the auditory cortex of awake macaques.," *Neuron,* vol. 48, pp. 139-148, 2005.
19. M. A. Bee and G. M. Klump, "Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain.," *Journal of Neurophysiology,* vol. 92, no. 2, pp. 1088-1104, 2004.
20. A. Bregman, "Auditory streaming: Competition among alternative organizations.," *Perception & Psychophysics,* vol. 23, pp. 391-398, 1978.
21. L. P. A. S. Van Noorden, *Temporal coherence in the perception of tone sequences.*, Eindhoven, The Netherlands: Doctoral Dissertation, Eindhoven Institute of Technology, 1975.
22. D. Wang and G. Brown, *Computational Auditory Scene Analysis : Principles, Algorithms, and Applications.*, Piscatway, NJ: IEEE Press, 2006.
23. E. Cherry, "Some experiments on the recognition of speech with one and with two ears.," *Journal of the Acoustical Society of America,* vol. 25, no. 5, pp. 975-979, 1953.
24. A. S. Bregman, *Auditory scene analysis: the perceptual organization of sound.*, Cambridge, MA: The MIT Press, 1990.

25. L. Erman, F. Hayes-Roth, V. Lesser and D. Reddy, "The Hearsay-II speech-understanding system: Integrating knowledge to resolve uncertainty," *ACM Computing Surveys,* vol. 12, no. 2, pp. 213-253, 1980.
26. D. Ellis, *Prediction-driven computational auditory scene analysis.*, Cambridge, MA: Dept. of Electrical Engineering, MIT., 1996.
27. D. Wang, "Primitive auditory segregation based on oscillatory correlation.," *Cognitive Science,* vol. 20, pp. 409-456, 1996.
28. C. von der Malsburg, "The correlation theory of brain function.," in *Models of Neural Networks II: Temporal aspects of Coding and Information Processing in Biological Systems*, New York, Springer-Verlag, 1994, pp. 1-26. (Published originally as Internal report 81-2. Department of Neurobiology, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany, 1981.)
29. C. Darwin and C. Bethell-Fox, "Pitch continuity and speech source attribution.," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 3, pp. 665-672, 1977.
30. R. Warren, "Perceptual restoration of obliterated sounds.," *Psychological Bulletin,* vol. 96, pp. 371-383, 1984.