Auditory Streaming as a Paradigm of Synergetic Pattern Formation in Brain and Behavior

Viktor Jirsa[1]*, Felix V. Almonte[2], Ajay S. Pillai[3], and Betty Tuller[4]

[1] Institut de Neuroscience des Systemes UMR1106 Inserm,
Aix-Marseille Universite, 13005 Marseille, France
http://ins.univ-amu.fr/research-teams/theoretical-neurosciences-group
[2] Center for Complex Systems & Brain Sciences,
Florida Atlantic University, Boca Raton, FL 33431, USA
[3] Human Motor Control Section, National Institutes of Health,
10 Center Drive, Rm 7D42, Bethesda, Maryland 20892, USA
[4] National Science Foundation, 4201 Wilson Blvd., Arlington VA 22230, USA

Abstract. Synergetics has established a well-known top-down approach to the modeling of perceptual phenomena in psychology and cognitive sciences. This phenomenological approach is deply rooted in the theory of pattern formation and offers a formal justification that in the proximity of transitions from one pattern to another a low-dimensional description via canonical models is permissible. We exploit this thinking in the context of auditory scene analysis, specifically auditory streaming, where the brain network integrates or segregates sounds that arise from two or more distinct sources. We interpret the process of integration and segregation as a pattern formation process and demonstrate through mathematical modeling, behavioral experiments and functional magnetic resonance imaging (fMRI) that selected networks in the brain get differentially activated as a function of the percept. We propose a functional architeture composed of brain areas with tonotopic organization (auditory cortex) and non-tonotopic organization (various parietal areas including right superior parietal lobule and precuneus). The dynamics of this functional architecture extends beyond auditory streaming and suggests the existence of informational convergence zones in the brain that get selectively activated in a nonlinear all-or-none fashion. This dynamics is reminiscent of phase transitions as discussed in synergetics and generalizes concepts well established in multisensory integration.

Keywords: auditory streaming, synergetics, perception, emergence, fMRI, BOLD signal

1   Introduction

Synergetics is an interdisciplinary field of research founded by Hermann Haken [1] and explains how macroscopic self-organized pattern formation occurs in open

systems operating away from equilibrium. Examples of such pattern formation range from the formation of Rayleigh-Bénard convection rolls in liquids to various chemical instabilities and morphogenesis in biological systems [2]. Though synergetics has its roots in physical systems, its concepts and mathematical apparatus generalize to other disciplines describing the collective self-organizing dynamics of multi-component systems.

One of the more recent applications of synergetics is found in the psychological sciences. In particular, perceptual phenomena related to state transitions are amenable to synergetic analysis. In this context, we will discuss the phenomenon of auditory scene analysis, particularly auditory streaming. Auditory stream segregation, or streaming, has been used as a model for how the auditory system integrates or segregates sounds that arise from two or more distinct sources. For example, when listening to bass and soprano vocalists singing simultaneously, the two voices are perceived as separate from each other but each voice is simultaneously perceived as an integrated perceptual event. In the laboratory, a similar effect can be created using sequences of tones. In a typical streaming experiment, two sequences are created using alternating high and low tones. Sequences vary in the frequency difference between the tones and presentation rate. In general, when the frequency separation is relatively small and/or the rate is relatively slow, listeners perceive a single integrated melody (or stream) and can accurately report the ordering of the tones. But when the frequency separation is relatively large and/or the rate relatively fast, listeners report hearing two auditory streams, one with higher pitch than the other. They can easily attend selectively to one or the other stream but they are unable to hear the tones as a single integrated stream and cannot report the relative order of individual events between the two streams. In this sense, streaming may be regarded as a pattern formation process, where the perceptual patterns emerge dependent on the details of the input sequences. The parameters characterizing the input sequence serve as unspecific control parameters, equivalent to the control parameters in a physical system. This is in stark contrast to response-driven approaches where the specific characteristics of the input signal determine the time course of the driven system. Prevailing models from auditory streaming studies focussed on the examination of auditory cortex responses and posit that streaming will be evoked whenever the tones of the input excite non-overlapping populations of neurons. Parametric variations of stimulations or stimulus features could produce neural activity patterns, which vary linearly with the sigmoidal firing rate of neural populations. While such patterns have been widely reported in vision, only limited evidence for such a mechanism exists in the auditory system. More importantly, such mechanisms rely heavily on the tonotopic organization of the auditory cortex. Yet evidence from neural recordings in humans suggests that activations during auditory streaming paradigms are significantly more widespread, involving brain areas outside of the auditory cortex that have no tonotopic organization [3, 4]. This suggests that the large network activated during the formation of streaming-related percepts results in the emergence of

brain pattern dynamics.

In this chapter we will remind the reader of our model of auditory streaming [5], which predicted that widespread brain networks with mixed organization (tonotopic and non-tonotopic) would be involved in auditory streaming. Next we provide behavioral evidence for a particular prediction of this architecture, namely, the phenomenon of amplitude streaming, in which tone sequences that differ only in the amplitude of consecutive tones can form integrated or segregated streams. This distinguishing feature poses a particular challenge to functional architectures comprised of only tonotopically organized networks. Through functional Magnetic Resonance Imaging (fMRI), we tested the neuro-architectural predictions of our auditory streaming model and report the findings in the final section of this chapter.

1.1   Auditory stream segregation

Auditory stream segregation has long been a focus of psychophysical research (e.g., [6]). The phenomenon has proven to be robust and may be fundamentally related to other integration and segregation phenomena in perception. Related studies have been conducted in humans (e.g., [6–15]), monkeys [16], and bats [17], and several theories have been proposed to account for the phenomenon. In Gestalt theory, streaming is viewed as arising from fundamental principles inherent in the input patterns, such as proximity (of the tone frequencies), similarity, and spatiotemporal cohesion [18–25]. Bregman (1990) appeals to Gestalt principles in explaining auditory grouping mechanisms, and auditory scene analysis in general, but recruits other explanatory concepts such as integrative schemas when Gestalt principles fail. Other general theories invoked to explain auditory streaming include filter, or channel models, e.g., [26], that hold that streaming is based on selective attention to a single perceptual dimension such as pitch. Nevertheless, spectral separation and other differences in power spectrum are not necessary conditions for perceptual stream segregation (e.g., [14, 15]). Streaming also occurs in sequences of amplitude modulated, harmonically complex tones (e.g., Joris, Schreiner, and Rees; 2004). Moreover, channel models fail to accommodate streaming effects that are dependent on relationships among tones, such as quality, higher-order frequency relationships [27] and timbre [28]. Others (e.g., [7, 29]) propose that listeners attend to frequency motions in perceiving auditory streams. In fact, some researchers claim that Gestalt principles are generally inadequate explanations for entire classes of acoustic and visual grouping, for example in speech perception [30] and vision [22]. Although in nature there are many kinds of gestalten, that is organized patterns of perception or behavior, the means by which those patterns arise is as yet unclear.

Other work has highlighted the dynamic nature of streaming phenomena by demonstrating the importance of initial percept (van Noorden, 1975) and longer-term temporal effects [31, 32]. In one influential series of experiments, van Noorden presented tones with two different pitches in the order low-high-low-low-high-low (ABAABAABA). Tones were of equal amplitude and duration and
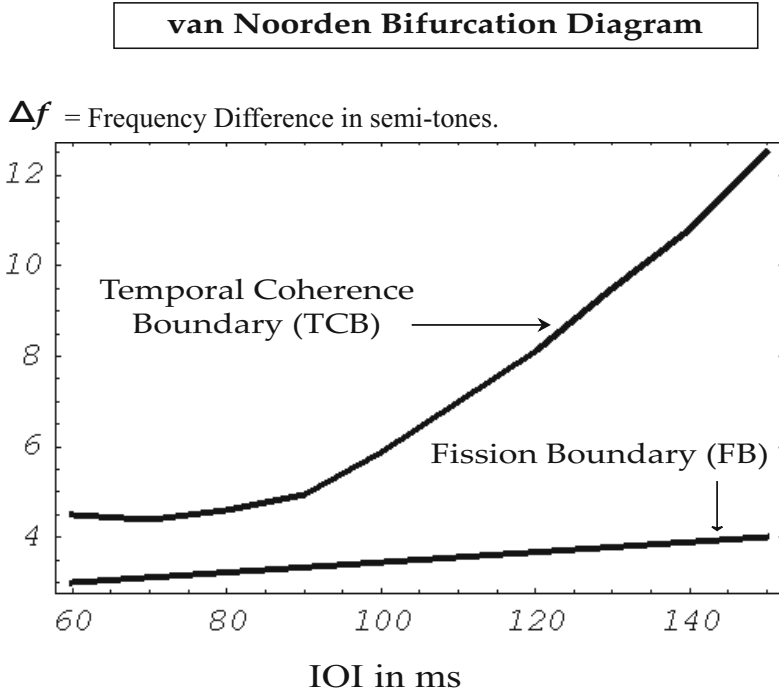
Fig. 1. *Stimulus sequences.* Variations of frequency difference or interstimulus intervals (ISI) are used as control parameters to manipulate the emergence of a particular percept.

the onset interval between successive tones was identical. Van Noorden presented listeners with tone sequences in which the frequency difference between the A and B tones was small and asked them to follow the gallop rhythm formed (so that perception was of an integrated tone sequence). He also presented tone sequences in which the frequency difference between the A and B tones was much larger and asked the listeners to focus their attention on the string of low tones (so that two segregated sequences of different frequency tones were perceived). See Figure 1 for example stimulus sequences, which were used in the experiments reported in this chapter. van Noorden (1975) then manipulated the frequency difference between A and B tones and their interstimulus-onset-interval (IOI) toward the other sequence type (integrated or segregated) and mapped the perceptual changes (see Figure 2). He found (1) a frequency-time boundary beneath which all sequences were heard as integrated, regardless of instructions (the Fission Boundary, FB) and (2) a frequency-time boundary above which all sequences were heard as segregated, regardless of instructions (the Temporal Coherence Boundary, TCB). In the bistable region between these two boundaries, a sequence could be heard as either integrated or segregated
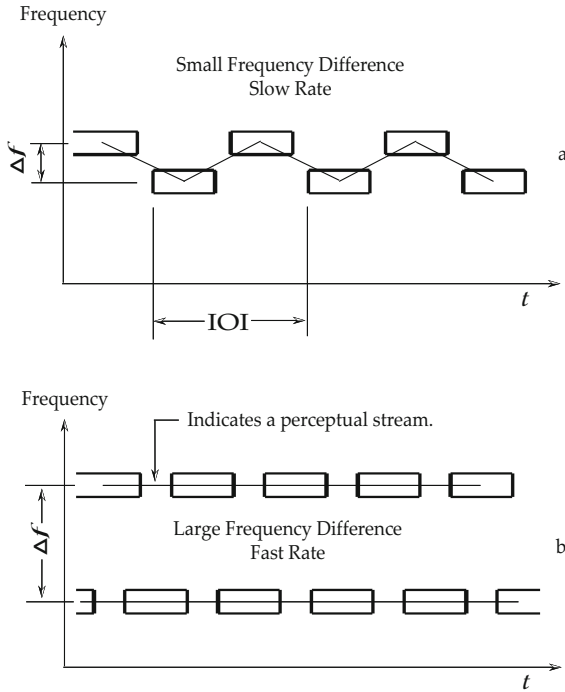
Fig. 2. *van Noorden's bifurcation diagram.* Variations of two control parameters, the frequency difference of tones and their interstimulus interval (ISI), allow to span a two-dimensional parameter space. The parameter space is partitioned into three regimes, one region with the percept one stream, another region with the percept two streams and a region in between which permits both.

depending upon initial instructions, with hysteresis phenomena observed when traversing the bistable regime.

1.2   Physiological correlates of auditory streaming

Complementary to the psychophysical approaches to the integration and segregation of sounds, there are currently two predominant neurophysiological theories of how the nervous system integrates environmental signals. The first theory is referred to as the binding theory and assumes that an integrated percept arises when activity in cortical areas becomes synchronized [33]. The second theory is grounded in the field of multisensory integration and assumes the existence of informational convergence zones. These convergence zones are made of cortical and subcortical networks such as the network consisting of the superior colliculi, the inferior parietal areas and the insula which is activated during the integration of speech and vision [34] among other behaviors. Both sources of neurophysiological evidence suggest that activations of larger networks are relevant for perceptual integration. Importantly, they also involve neural areas with

no known tonotopic architecture. For example, the left inferior parietal cortex has been hypothesized to be involved in the integration of visual-acoustic information to a common percept parametrized by temporal disparity [35]. Other researchers have found similar networks to be involved in the spatiotemporal integration of visual cues during collision judgments [36]. In the more related, but also more complex situation of speech perception, Hickock and Poeppel [37] argued that the left inferior parietal cortex is also involved in the integration of auditory-motor processes. Recruitment of these higher areas is largely non-specific to the modalities involved, which implies a general mechanism for the integration of signals to a coherent percept.

Specific to the phenomenon of auditory streaming, work by Snyder and colleagues (2008, 2009) confirms the presence of hysteresis in that perception and Event Related Potential (ERP) magnitude during an auditory streaming task depend on whether the prior sequence was perceived as integrated or segregated. Cusack [4] reported increased BOLD activity in the anterior intraparietal sulcus during two versus one-stream percepts, but did not find differences in the auditory cortex based on percept or frequency separation. The latter is in contrast to findings by Gutschalk and colleagues [38] who found covariations in magnetoencephalographic (MEG) signals with both frequency separation and percept. In event-related fMRI experiments of auditory streaming, Kondo and Kashino [39] found activations in auditory cortex, posterior insula, medial geniculate body, and supra-marginal gyrus but no contrasts were carried out regarding effects of perceptual organization or frequency separation. These results underline the need for further examination of the neurophysiological basis of auditory streaming, in particular the involvement of brain areas outside of the auditory cortex.

2   The Almonte et al. model of auditory streaming

The functional architecture proposed by Almonte and colleagues [5] was inspired by the large literature on integration phenomena across sensory modalities, in which multisensory convergence zones (represented by certain brain areas such as the insula) integrate information from different sensory modalities through an increased activation covariant with the percept of multisensory integration [40]. Almonte and colleagues explored the possibility of a homeomorphic architecture for auditory streaming comprising two layers. One layer, a neural field [41–44], is tonotopically organized such that the frequency of the acoustic stimulus maps onto a location in neural space. The second layer is a non-tonotopically organized subsystem and classifies the spatiotemporal neural field dynamics along very much the same lines as convergence zones in multisensory paradigms. The classification itself is not just a measurement (else the application of a simple measure to the neural field would suffice) but is itself a dynamic integrating process. In fact, bistability and hysteresis are properties of the classification dynamics rather than properties of the neural field dynamics. From the view point of synergetics, the second subsystem defines the order parameter of the percep-

tual pattern forming system, whereas the first subsystem performs a form of preprocessing of the input stimulus sequence.

The dynamics of the neural field $\mu(x,t)$ are given by the Jirsa-Haken wave equation [44] accommodating auditory inputs $s(x,t)$ as follows:

$$\left[\frac{1}{\gamma^2}\frac{\partial^2}{\partial t^2} + \frac{2}{\gamma}\frac{\partial}{\partial t} + 1 - r^2\nabla^2\right]\mu_v = \left(1 + \frac{1}{\gamma}\frac{\partial}{\partial t}\right)\varsigma(\mu_v + s(x,t)), \tag{1}$$

where, as a reminder, $\gamma = c/r$, c is the speed of spike propagation and r parameterizes the spatial decay of lateral interactions. The external input or stimulus to the neural sheet is $s(x,t) : \mathbf{R}^2 \to \mathbf{R}$, which contains all the spatiotemporal characteristics of the auditory input stream. Periodic boundary conditions, $\mu(0,t) = \mu(L,t)$, $t \geq 0$, are used. The second subsystem is not tonotopically organized, hence its spatial dimension is of no relevance, when we consider only the competition of two streams. In fact, the ability to show multistable pattern formation is the only relevant property of this subsystem. A simple multistable subsystem with its scalar state variable $y(t)$ is given by the equation

$$\dot{y} = \varepsilon y - y^3 - I_0 + I(t), \tag{2}$$

where ε is a constant that captures all linear contributions. I_0 contains all constant contributions given rise to the rest state activity. The functional $I(t)$ is specified as

$$I(t) = \int_0^L h(\mu(x,t))dx \qquad h(n) = \left\{\begin{array}{ll} 0, & n \leq \Omega \\ n, & n > \Omega \end{array}\right\}, \tag{3}$$

where Ω is a neural activity threshold. The equations (1), (2) and (3) define the dynamics of a stream classification model in one of its simplest forms. Figure 3 illustrates the architecture of the model.

To understand van Noorden's results, we parametrize a sequence of consecutive tones by their frequency difference, Δf, and their inter-onset interval, IOI. As the neural field evolves, it is integrated across space and time yielding the time dependent, but scalar, activity $I(t)$ driving the second system. $I(t)$ represents the relevant "information" from the neural field μ as a spatiotemporally integrated activity measure, which depends on the amount of dispersion over space and time. The greater the dispersion, the greater will be the value of $I(t)$ at a given time point. Figure 4 shows the contour lines of neural field activity over space x and time t for the bistable situation.

The final state reached by the second system defined in Eq. (2) with activity y will depend on $I(t)$ and its own intrinsic dynamics. For an intermediate value of $I(t)$, there is a bistable regime in which y can assume either one of the fixed points. The negative fixed point is identified with perceiving one stream and the positive fixed point with perceiving two streams. The time series for y are shown in Figure 5 for several different initial conditions of the activity y. After a transient the activity becomes stationary, displaying three possible scenarios (see Figure 5 from top to bottom): one stream only, or the bistable situation, in
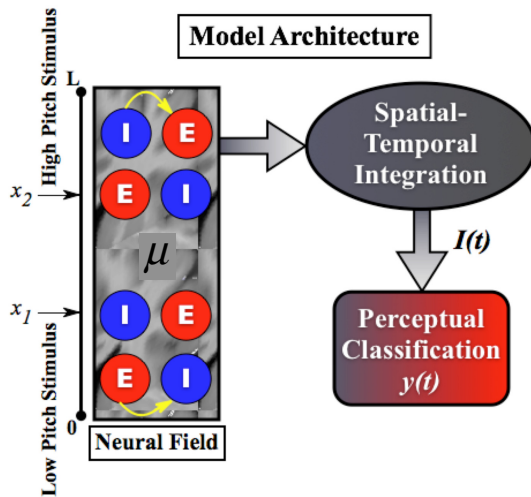
Fig. 3. *Cortical architecture of the model.* The neural field is illustrated by the rectangular box showing the neural activity $\mu(x, t)$ composed of inhibitory and excitatory neurons. The input $s(x, t)$ is provided at locations x_i via the Gaussian localization function $e^{-(x-x_i)^2/\delta_i}$ with width $\sqrt{\delta_i}$. The explicit model parameters used in the simulations are given in Almonte et al (2005).

which either one integrated stream or two separate streams may be perceived, or finally two streams only. For each choice of Δf and IOI, the model equations (1) and (2) are solved numerically and their stationary states determined. The results are plotted in the two-dimensional parameter space shown previously in Figure 1.

3 A novel paradigm: amplitude streaming

Do the mechanisms leading to the emergence of auditory streaming need to be explained through tonotopically structured networks? Or may there be a crucial role to be played within the functional architecture by subsystems with no tonotopic organization? In what follows, we study specifically whether auditory streaming involves interacting neural subnetworks, some tonotopic but others non-tonotopically organized and acting as information convergence zones for the classification of the perceptual world. To this end, we first show that perceptual streaming occurs for sequences of pure tones with constant frequency but systematic amplitude modulation across tones. This demonstration is a necessary step because our model does not depend on the system being organized
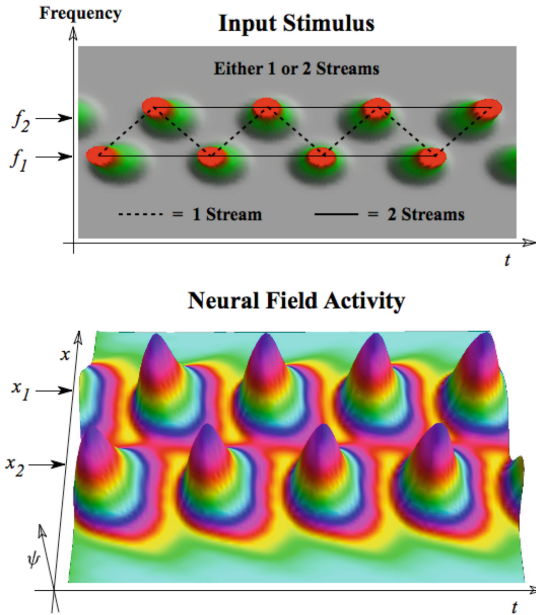
Fig. 4. *Bistable regime of auditory streaming.* The stimulus sequences (top) and its resulting neural field dynamics (bottom).

tonotopically. We then, in the next section, examine the BOLD responses from listeners who perceive auditory streaming 1) on the basis of frequency differences among equal amplitude tones and 2) on the basis of amplitude differences of same-frequency tones.

In order to assess perceptual streaming on the basis of amplitude differences alone, we constructed sequences containing 40 sinusoidal tones of constant frequency (600 Hz) and duration (100ms) and an inter-onset interval (IOI) of 110ms (i.e., an interstimulus interval of 10ms). The only parameter varied across sequences was the amplitude ratio (AR) between adjacent tones (T1 and T2). AR was kept constant within each of the 20 T1-T2 pairs in a sequence. Values of AR relative to the highest amplitude tone were 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 1 (AR = 1 denotes equal amplitude of T1 and T2). Amplitude was linearly ramped over the first and last 10ms of each tone.

Pairs of 40-tone sequences were presented 5 times to subjects in random order, with 500ms between members of each sequence pair. The subject's task was to judge which of the two sequences in a pair sounded slower. When tone sequences of the type used here are perceived as two streams, they are perceived to be slower than sequences perceived as a single stream. Subjects responded by pressing one of three labeled keys on a keypad: Pressing the number 1 meant that the first sequence was perceived as slower; pressing 2 meant that the second sequence was perceived as slower; and pressing = meant that the two sequences
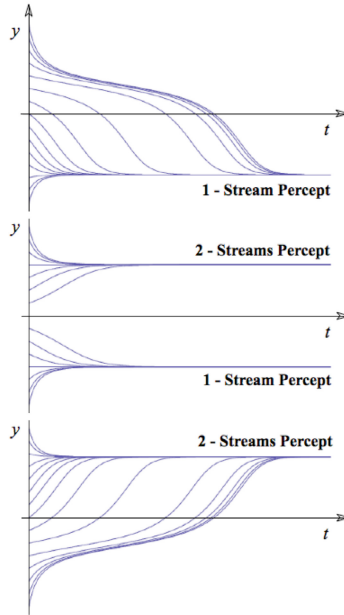
Fig. 5. *Percept formation.* For multiple initial conditions the time series of $y(t)$ are plotted for the three regimes, one stream only (top), bistable (middle) and two streams only (bottom).

were perceived as equal in rate. All combinations of sequences were included, with the exception of sequences having AR = 0.3, AR = 0.5, and AR = 0.7 paired with itself, for a total of 230 sequence pairs. Order of sequence presentation for each pair was counterbalanced.

A control condition was included in which each trial consisted of three short tone sequences presented with 500ms ISI between sequences. Each sequence was either 1) a T1-T2 sequence with AR = 0.2 and IOI = 110ms, or 2) a sequence with T1=1 and IOI = 220ms, i.e., T2 was omitted. Triad orders were sequences (a) 1, 2, 1; (b) 2, 1, 1; (c) 1, 2, 2; and (d) 2, 1, 2. Sequence 1 contained 12 tones. Sequence 2 contained 6 tones, with double the IOI of sequence 1, in order to verify that subjects reliably heard the lowest amplitude tones. The procedure was an ABX, forced-choice task in which subjects judged whether the rate of the third tone sequence (X) was more like the first sequence (A) in the triad or more like the second sequence (B) in the triad. Subjects responded by pressing one of two labeled keys on a keypad: Pressing the number 1 meant that the rate of the third (test) sequence was more like the first sequence. Pressing 2 indicated that the rate of the third sequence was more like the second sequence.

All subjects (N=11) had normal hearing according to self-report and were naive to the purpose of the experiment. They were told that there were no right
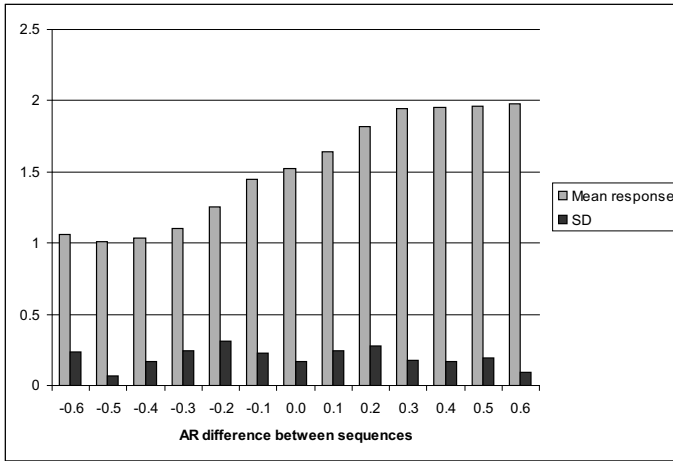
Fig. 6. *Amplitude Streaming.* Behavioral results are shown based on amplitude variation of consecutive tones.

or wrong answers and that our interest was purely in how they perceived the tone sequences.

Responses from the control condition were examined to verify that all subjects were able to hear the lowest amplitude tone reliably. All subjects responded with at least 85% accuracy (9 of the 11 responded with 95% accuracy or better). Thus, the low amplitude tone was indeed perceived and contributed to the perception of a faster rate sequence.

Figure 6 shows the means and standard deviations of the responses across subjects for the perception of streaming in sequences of amplitude-modulated pure tones. The x-axis represents the AR difference between the two sequences in each pair. When the difference is negative, the amplitude difference between successive tones within a sequence is greater in the first sequence than the second. In this case, subjects judge the first sequence to be slower than the second, indicating that streaming has occurred in the first sequence but not in the second. When the AR difference is positive, the amplitude difference between successive tones within a sequence is greater in the second sequence than the first. In this case, subjects judge the second sequence to be slower than the first, indicating that streaming has occurred only in the second sequence. For small AR differences, subjects judge the rates of the sequences to be equal, indicating that streaming occurred in both sequences (denoted on the y-axis by 1.5 in Figure 6). Moreover, there is no order effect. AR differences of -0.6, for example, are equivalent to AR differences of 0.6. The lack of an order effect was confirmed statistically by a 2-way repeated measures ANOVA with order and AR difference as factors. For order, $F_{(1, 5)} = 2.67$, $p > 0.1$ and the interaction of order and AR difference was similarly not significant, $F_{(5, 5)} = 1.88$, $p > 0.1$.

Amplitude Streaming (AS)		
Condition: Name	f_1, f_2 **in Hz**	A_1, A_2
1: AS	1000, 1000	0.2, 1
3: Control	1000, 1000	1, 1
4: Control	1000, 0	1, 0
5: Control	1000, 1000	0.2, 0.2
6: Control	1000, 1000	0.2, 0
Frequency Streaming (FS)		
Condition: Name	f_1, f_2 **in Hz**	A_1, A_2
2: FS	500, 1000	1, 1
3: Control	1000, 1000	1, 1
4: Control	1000, 0	1, 0
7: Control	500, 500	1, 1
8: Control	500, 0	1, 0

Table 1. Parameter values for experimental conditions consisting of tone sequences *ABAB...AB*. Conditions 3 & 4 function as controls for both amplitude and frequency streaming. In order to examine the neural effects of auditory streaming, several control conditions needed to be built into the experimental design. Condition 1 is the condition most likely to result in amplitude streaming, since the tones are of the same frequency but with very different amplitudes. Condition 2 is most likely to result in frequency streaming. Conditions 3 and 5 were used to eliminate rate effects as the only cause of a significant amount of neural activation in amplitude streaming (condition 1). Conditions 4 and 6 were used to eliminate amplitude effects as the only cause of a significant amount of neural activation in amplitude streaming (condition 1). Control conditions 3 and 7 were used to eliminate rate effects as the only cause of a significant amount of neural activation in frequency streaming (Condition 2). Conditions 4 and 6 were used to eliminate amplitude effects as the only cause of a significant amount of neural activation in frequency streaming (condition 2).

In summary, in spite of the fact that all tones were of equal frequency and IOI was constant across sequences, amplitude difference alone was sufficient to cause the pure-tone sequences to split into two perceptual streams. This is in agreement with the model's prediction of the possibility of perceiving one or two streams depending on the AR of the tone sequences.

4   Amplitude streaming and its associated BOLD responses

Next we consider the BOLD responses from listeners who perceive auditory streaming 1) on the basis of frequency differences among equal amplitude tones and 2) on the basis of amplitude differences of same-frequency tones. Eight subjects (4 females and 4 males) between 23 and 42 years of age participated in the fMRI experiment. All subjects were in good health with normal hearing (by self-report) and no past history of psychiatric or neurological disease. Informed

consent was collected from each subject prior to the experiment and the study was approved by the Florida Atlantic University Institutional Review Board.

Because scanner noise is a primary concern when using fMRI for auditory perceptual experiments, we used a sparse sampling (or clustered volume) scanning technique [3]. The long TRs used by this technique directly influence the possible experimental design, especially, experiment duration, which affects the number of trials that can be collected for each experimental run. Auditory stimuli were presented to the subjects in 12-second on-blocks followed by 12 seconds of silence. The on-blocks consisted of pure tone sequences. The only parameters that were varied were the frequency difference between adjacent tones, the silent gaps between tones, and the amplitude ratio of adjacent tones. Tone durations were always 50ms with 5ms amplitude ramping at the beginning and end of each tone. Interstimulus silent intervals were either 100ms or 250ms, depending on the condition (see Table 1). Each on-block contained the maximum number of [tone, silentgap] pairs that fit in the 12-second interval; any remaining time was divided into two equal durations of silence at the beginning and end of each tone sequence. fMRI scans lasted 2sec and began 2sec before the end of each block, so that the tone stimuli were still present during the scanning period. Each of the eight conditions was repeated twenty times during the experiment for a total run time of 64 minutes. The order of stimulus presentation was randomized across presentation blocks and conditions, but not across subjects, with the constraint that no condition was repeated consecutively.

Before the beginning of the experiment, subjects went through a short training session with sequences from Conditions 1 and 2, each repeated 3 times in random order. The purpose of the training session was to allow the subjects to become familiar with the stimulus sequences and to determine via verbal report whether they perceived 1) amplitude-based streaming in Condition 1 (equal frequency tones of different amplitude) and 2) frequency-based streaming in Condition 2 (different frequency tones with equal amplitude). A General Electric (GE) 3T Signa scanner was used to acquire T1-weighted structural images and functional EPI images for the measurement of the blood oxygenation level-dependent (BOLD) effect. The acquisition scheme and parameters used for the functional scans were as follows: echo-planar imaging (EPI), gradient recalled echo, TR = 12 s, TE = 35 ms, flip angle = 90 degrees, 64 x 64 matrix, 30 axial slices per scan each of thickness 5 mm acquired parallel to anterior-posterior commissural line. The data was preprocessed and analyzed using Statistical Parametric Mapping software [10] (SPM2 from Wellcome Department of Cognitive Neurology, London, UK). Motion correction to the first functional scan was performed within subject using a six-parameter rigid-body transformation. The 8 subjects had less than 7mm of translation in all directions and less than 6.0 degrees of rotation about the three axes. The mean of the motion-corrected images was then coregistered to the individual 30-slice structural image using a 12-parameter affine transformation. The images were then spatially normalized to the Montreal Neurological Institute (McGill University, Canada, http://www.bic.mni.mcgill.ca/) template brain by applying a 12-parameter affine transformation, followed by

a nonlinear warping using basis functions. These normalized images were interpolated to 2 mm isotropic voxels and subsequently smoothed with a 4 mm isotropic Gaussian kernel. A random-effects, model-based, statistical analysis was performed with SPM2 [10] in a two level procedure. The first level consists of estimating a General Linear Model (GLM) [11] of the form: $Y = Xv + e$ for each subject where X is a (m x n) experimental design (basis) matrix (m = 320 = Total number of stimulus presentations and n = 8 = Number of conditions), and v is a constant vector representing weights for each basis vector of X. Each column vector of X consists of a series of entries of zeros for the off-blocks and ones for the on-blocks and represents different stimulus conditions in each functional run and six motion parameters obtained from the realignment. v is estimated using the method of least squares and minimizing the error e. Individual contrast images were created by correlating the brain response with the aforementioned covariates for each subject. Global differences among subjects were controlled by proportional scaling. The individual contrast images were then entered into a second-level analysis, using a separate one-sample t-test for each term in the general linear model. The summary statistical maps were thresholded at p < 0.05 (uncorrected for multiple comparisons). These maps were overlaid on a high-resolution structural image in the Montreal Neurological Institute (MNI) orientation.

In previous work, no explicit contrasts were carried out to test for effects of perceptual organization or frequency separation (such as [39]) reinforcing the need for further examination of the involvement of areas outside the auditory cortex in streaming. Here we develop a set of constraints based on set theoretical operations allowing us to address differential effects related to streaming or streaming in combination with amplitude, frequency, or rate changes. Figure 7 shows voxels which correlate with streaming percepts. The voxels in panels A (amplitude streaming), B (frequency streaming), & C (the intersection of amplitude and frequency streaming) are related only to streaming percepts. Anatomical areas activated by amplitude streaming (panel A in Figure 7) include primary auditory areas as well as parietal areas (right superior parietal lobule (SPL) and precuneus). Similarly, for frequency streaming (panel B) there are activations in primary auditory areas and beyond, including the left inferior parietal lobule (IPL), supramarginal gyrus (SG), precuneus, parietal lobe subgyral (Brodmann area 40), right inferior and superior parietal lobule, and middle and superior temporal gyri. The implication of the right parietal cortex is consistent with related findings in which intraparietal sulcus (IPS) showed greater activity when two streams were perceived rather than one [4] and reports that right parietal cortex influences the auditory perceptual scene - more specifically, sound movement (Griffiths et al. [9]). The findings are also consistent with the results of Bushara and colleagues [45] who found that the hierarchical organization of the auditory system extends beyond the temporal lobe to include areas in the posterior parietal and prefrontal regions, especially for auditory spatial processing.
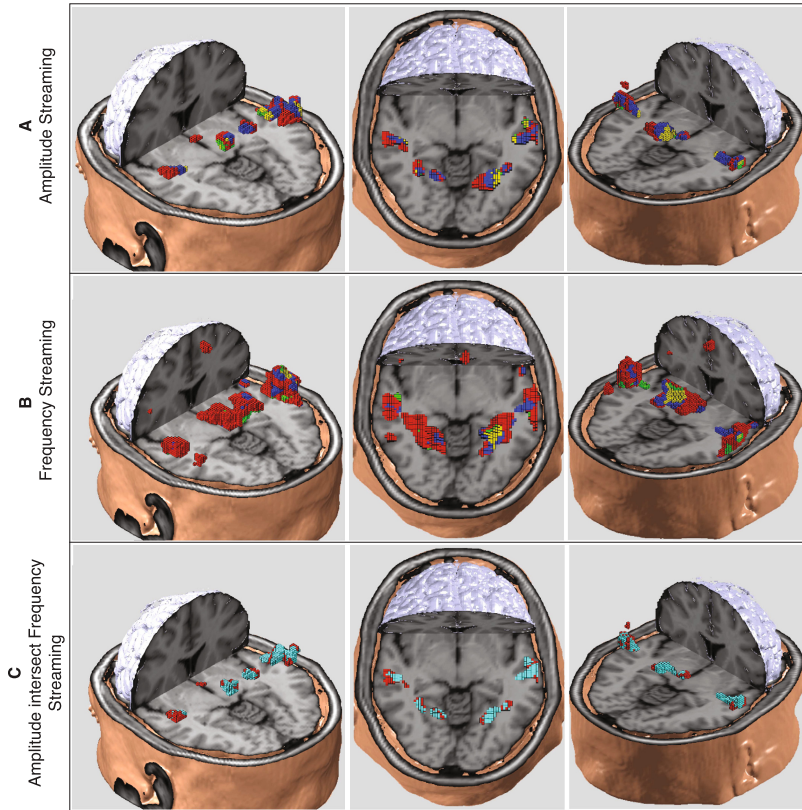
Fig. 7. *Streaming results in the BOLD response.* Three different activation patterns are shown. Active voxels related to the percept of amplitude and frequency streaming and their intersection at p < 0.05. The voxels are color coded to show contributions due only to the streaming percept (red voxels) and any other effects not related to amplitude, frequency, and rate changes in the stimuli. Whereas the blue, cyan, green, and yellow voxels are related to amplitude, frequency, or rate changes in the stimuli.

Panels A and B in Figure 7 show areas of significant BOLD increases for amplitude and frequency streaming. In contrast, there are relatively few voxels with significant activity that are common to both amplitude and frequency streaming (panel C). There is ∼44% overlap among the voxels of panels A and C and ∼8% overlap among voxels of panels B and C. The non-empty intersection shown in panel C defines the necessary condition for the existence of convergence zones that influence the perception of both amplitude and frequency streaming. If we subtract (voxel by voxel) these common networks from the voxels in panels A and B we obtain specific networks corresponding only to amplitude and frequency streaming percepts respectively. The networks represented by the voxels in panels A, B, and C are independent of the changes in the control parameters since the effects of these parameter changes were subtracted out. Complemen-

tary to this conclusion, there also exist specific neural networks that influence the perception of either amplitude or frequency streaming but not both since there exist voxels in panels A and B complementary to the intersection shown in Panel C.

5   Conclusions

We have shown in the auditory domain that by specifying primary experimental stimulus conditions (amplitude and frequency modulation, in our case), together with appropriate controls, we can identify convergence zones related to the percepts induced by the primary stimulus conditions. Convergence zones were determined by subtracting effects present in individual control conditions using contrasts dependent on relationships among the primary conditions and controls (see Felix Almonte's PhD thesis, Florida Atlantic University 2006, for details). Convergence zones were found mainly within the right parietal areas. Thus, even though parietal areas may be multimodal, they still account for unimodal auditory temporal integration. Convergence zones are interpreted as networks that integrate higher order information leading to specific perceptual states, thus making them plausible representational candidates for the neural substrates of awareness. From the synergetic view point, convergence zones basically track the perceptual dynamics as we propose here, but they also participate in a widespread brain network responsible for the overall brain pattern formation. Hence, when convergence zones change their activation dynamics, it corresponds to a larger spatiotemporal reorganization of the brain patterns as described by phase transitions in synergetic systems. From a more traditional neuroscience perspective, convergence zones form dominant parts of cell assemblies in a neural representational strategy, for instance as proposed by Singer, relying on the dynamic association of feature-specific cells into functionally coherent cell assemblies. The cell assemblies represent the constellation of features defining a particular perceptual object. These possibilities combined with our experimental results provide support for theoretical models of perception that employ the idea of convergence zones as mediators for the organization of percepts.

References

1. Haken, H.: Synergetics. An Introduction. Springer, Berlin (1983)
2. Cross, M.C., Hohenberg, P.C.: Pattern formation outside of equilibrium. Rev. Mod. Phys. **65**, 851–1112 (1993)

3. Dykstra, A.R., Halgren, E., Thesen, T., Carlson, C.E., Doyle, W., Madsen, J.R., Eskandar, E.N., Cash, S.S.: Widespread brain areas engaged during a classical auditory streaming task revealed by intracranial EEG. Frontiers in Human Neuroscience **5**, 74 (2011)
4. Cusack, R.: The intraparietal sulcus and perceptual organization. J. Cogn. Neuroscience **17**, 641–651 (2005)
5. Almonte, F., Jirsa, V.K., Large, E.W., Tuller, B.: Integration and segregation in auditory streaming. Physica D **212**, 137–159 (2005)
6. Miller, G.A., Heise, G.A.: The trill threshold. The Journal of the Acoustical Society of America **22**, 637–638 (1950)
7. van Noorden, L.P.A.S.: Temporal coherence in the perception of tone sequences. Ph.D. thesis, Eindhoven University of Technology (1975)
8. Bregman, A.S.: Auditory streaming: Competition among alternative organizations. Perception and Psychophysics **23**, 391–398 (1978)
9. Bregman, A.S.: Auditory Scene Analysis: The perceptual organization of sound. MIT Press (1990)
10. Lackner, J.R., Goldstein, L.M.: Primary auditory segregation of repeated consonant-vowel sequences. The Journal of the Acoustical Society of America **56**, 1651–1652 (1974)
11. Neisser, U.: Cognitive Psychology. Appleton-Century-Croft (1967)
12. Warren, R.M., Warren, R.P.: Auditory illusions and confusions. Scientific American **223**, 30–36 (1970)
13. Helenius, P., Uutela, K., Hari, R.: Auditory stream segregation in dyslexic adults. Brain **122**, 907–913 (1999)
14. Grimault, N., Bacon, S.P., Micheyl, C.: Auditory stream segregation on the basis of amplitude-modulation rate. The Journal of the Acoustical Society of America **111**, 1340–1348 (2002)
15. Vliegen, J.V., Oxenham, A.J.: Sequential stream segregation in the absence of spectral cues. The Journal of the Acoustical Society of America **105**, 339–346 (1999)
16. Izumi, A.: Auditory stream segregation in japanese monkeys. Cognition **82**, B113–B122 (2002)
17. Moss, C.F., Surlykke, A.: Auditory scene analysis by echolocation in bats. The Journal of the Acoustical Society of America **110**, 2207–2226 (2001)
18. Wertheimer, M.: Gestalt theory. Social Research **11** (1924). Translation of lecture at the Kant Society, Berlin
19. Koffka, K.: Principles of Gestalt Psychology. Harcourt, Brace and Company, Inc. (1935)
20. Köhler, K.: Gestalt Psychology. Liveright (1947)
21. Hartmann, G.W.: Gestalt Psychology. The Ronald Press Company (1935)
22. Kubovy, M., Pomerantz, J.R. (eds.): Perceptual Organization. Lawrence Erlbaum Associates, Inc., Publishers (1981)
23. Gobar, A.: Philosophic Foundations of Genetic Psychology and Gestalt Psychology. Martinus Nijhoff, The Hague, The Netherlands (1968)
24. Gordon, I.E.: Theories of Visual Perception. John Wiley & Sons (1989)
25. Lewin, K.: Principles of Topological Psychology. McGraw-Hill Book Company (1936)
26. Broadbent, D.: Perception and Communication. Pergamon Press (1958)
27. Idson, W.L., Massaro, D.W.: Cross-octave masking of single tones and musical sequences: The effects of structure on auditory recognition. Perception and Psychophysics **19**, 155–175 (1976)

28. Singh, G.P., Bregman, A.S.: The influence of different timbre attributes on the perceptual segregation of complex tone sequences. The Journal of the Acoustical Society of America **102**, 1943–1952 (1997)
29. Jones, M.: Time, our lost dimension: Toward a new theory of perception, attention, and memory. Psychological Review **83**, 323–355 (1976)
30. Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S., Lang, J.M.: On the perceptual organization of speech. Psychological Review **101**, 129–156 (1994)
31. Snyder, J.S., Carter, O.L., Lee, S.K., Hannon, E.E., Alain, C.: Effects of context on auditory stream segregation. J. Exp. Psychol. Hum. Percept. Perform. **34**, 1007–1016 (2008)
32. Snyder, J.S., Holder, W.T., Weintraub, D.M., Carter, O.L., Alain, C.: Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. Psychophysiology **46**, 1208–1215 (2009)
33. Engel, A.K., Fries, P., Singer, W.: Dynamic predictions: Oscillations and synchrony in topdown processing. Nature Reviews: Neuroscience **2**, 704–716 (2001)
34. Calvert, G.A.: Crossmodal processing in the human brain: Insights from functional neuroimaging studies. Cerebral Cortex **11**, 1110–1123 (2001)
35. Bushara, K.O., Grafman, J., Hallett, M.: Neural correlates of auditory-visual stimulus onset asynchrony detection. The Journal of Neuroscience **21**, 300–304 (2001)
36. Assmus, A., Marshall, J.C., Ritzl, A., Noth, J., Zilles, K., Fink, G.R.: Left inferior parietal cortex integrates time and space during collision judgments. NeuroImage **20**, S82–S88 (2003)
37. Hickok, G., Poeppel, D.: Towards a functional neuroanatomy of speech perception. Trends in Cognitive Sciences **4**, 131–138 (2000)
38. Gutschalk, A., Micheyl, C., Melcher, J.R., Rupp, A., Scherg, M., Oxenham, A.J.: Neuromagnetic correlates of streaming in human auditory cortex. J. Neuroscience **25**, 5382–5388 (2005)
39. Kondo, H.M., Kashino, M.: Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. J. Neuroscience **29**, 12695–12701 (2009)
40. Calvert, G.A., Spence, C., Stein, B.E. (eds.): The Handbook Of Multisensory Processes. MIT Press (2004)
41. Amari, S.: Biomathematics: Current Status and Future Perspectives, chap. A Mathematical Theory of Self-Organizing Nerve Systems, pp. 159–177. North-Holland (1982)
42. Wilson, H.R., Cowan, J.D.: A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. Kybernetik **13**, 55–80 (1973)
43. Jirsa, V.K., Haken, H.: Field theory of electromagnetic brain activity. Phys. Rev. Lett. **77**, 960–963 (1996a)
44. Jirsa, V.K., Haken, H.: Derivation of a field equation of brain activity. J. Biol. Phys. **22**, 101–112 (1996b)
45. Bushara, K.O., Weeks, R.A., Ishii, K., Catalan, M.J., Tian, B., Rauschecker, J.P., Hallett, M.: Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. Nat. Neuroscience **2**, 759–766 (1999)