ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Physica D xx (xxxx) xxx–xxx

PHYSICA D

www.elsevier.com/locate/physd

Integration and segregation in auditory streaming

Felix Almonte*, Viktor K. Jirsa, Edward W. Large, Betty Tuller

Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL 33431-6424, USA

Received 29 March 2005; accepted 30 September 2005

Communicated by Y. Kuramoto

Abstract

We aim to capture the perceptual dynamics of auditory streaming using a neurally inspired model of auditory processing. Traditional approaches view streaming as a competition of streams, realized within a tonotopically organized neural network. In contrast, we view streaming to be a dynamic integration process which resides at locations other than the sensory specific neural subsystems. This process finds its realization in the synchronization of neural ensembles or in the existence of informational convergence zones. Our approach uses two interacting dynamical systems, in which the first system responds to incoming acoustic stimuli and transforms them into a spatiotemporal neural field dynamics. The second system is a classification system coupled to the neural field and evolves to a stationary state. These states are identified with a single perceptual stream or multiple streams. Several results in human perception are modelled including temporal coherence and fission boundaries [L.P.A.S. van Noorden, Temporal coherence in the perception of tone sequences, Ph.D. Thesis, Eindhoven University of Technology, The Netherlands, 1975], and crossing of motions (Bregman, 1996). Our model predicts phenomena such as the existence of two streams with the same pitch, which cannot be explained by the traditional stream competition models. An experimental study is performed to provide proof of existence of this phenomenon. The model elucidates possible mechanisms that may underlie perceptual phenomena.

PACS: 43.66.Ba

Keywords: Auditory; Network; Perception; Streaming

1. Introduction

Intuitively, auditory streaming or stream segregation is like listening to bass and soprano vocalists singing simultaneously. Although the two voices overlap in time, they clearly form two distinct perceptual events.

* Corresponding author.

E-mail addresses: almonte@ccs.fau.edu (F. Almonte), jirsa@ccs.fau.edu (V.K. Jirsa), large@ccs.fau.edu (E.W. Large), tuller@ccs.fau.edu (B. Tuller).

0167-2789/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.physd.2005.09.014

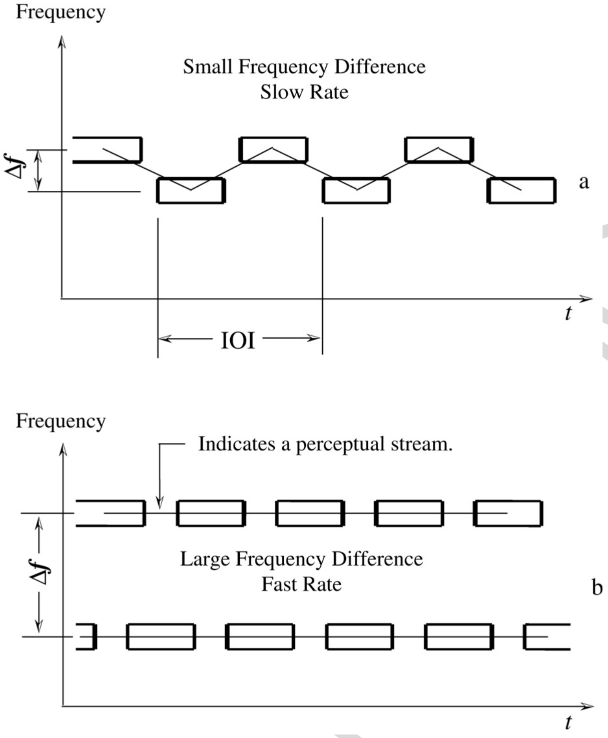
Frequency
Small Frequency Difference
Slow Rate
Δf
a
t
IOI
Frequency
Indicates a perceptual stream.
Δf
Large Frequency Difference
Fast Rate
b
t

Fig. 1. A frequency–time plot of input tone sequences where Δf is the frequency difference and IOI is the stimulus Inter Onset Interval. a: small frequency differences with one perceptual stream indicated by the solid lines. b: large frequency differences with the percept of two streams.

In the laboratory, a similar effect can be created using sequences of tones. In a typical streaming experiment, two sequences are created using sets of high and low tones as illustrated in Fig. 1. Sequences vary in presentation rate and the frequency difference between the tones. The basic finding (see, e.g., [1]) is as follows. (1) 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 (see Fig. 1a). (2) When the frequency separation is relatively large and/or the rate relatively fast, people clearly perceive two auditory streams, one with higher pitch than the other (see Fig. 1b). In this case they are able to attend selectively to one or the other stream with great efficiency and they are unable to hear the two tone sequences as integrated. For example, when listeners cannot integrate, they cannot report the relative order of individual events between the two streams.

During streaming experiments, van Noorden manipulated the initial perception of the listeners using the following strategy [2]. In one scenario he began with an integrated pattern and asked these listeners to follow the trill rhythm [3] formed by alternating sequences of high and low tones and to “hold on to this rhythm for as long as possible”. In other words, the subjects perceived one stream of alternating tones. In another complementary scenario, he began with perceptually segregated melodies and asked the subjects to “focus attention selectively on the string of low tones”. In each of the two scenarios, van Noorden [2] manipulated the frequency difference (Δf) and the inter-onset interval (IOI), toward the other sequence type (integrated or segregated). In essence, the initial condition is chosen such that the subject’s percept is monostable, i.e., there is no switching of percept. van Noorden [2] found three main phenomena. (1) There is a frequency–time boundary (known as the Fission Boundary (FB)) beneath which all sequences are heard as integrated, regardless of instructions. (2) There is a frequency–time boundary (known as the Temporal Coherence Boundary (TCB)) above which all sequences are heard as segregated,

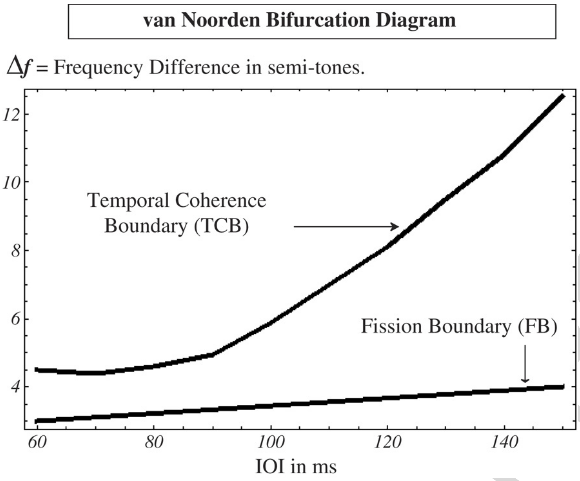

Fig. 2. Partial illustration of van Noorden's bifurcation diagram as a function of the frequency difference Δf and the IOI. The parameter space is partitioned into three regimes, one region with the percept one stream, another region with the percept two streams and a region in between which permits both.

regardless of instructions. (3) Between these two boundaries there exists a bistable region in which a sequence can be heard as either integrated or segregated depending upon instructions. Hysteresis phenomena are observed when traversing the bistable regime from either the Fission Boundary or Temporal Coherence Boundary. Refer to Fig. 2 for an illustration of van Noorden's results.

Numerous psychophysical studies investigating the stream segregation phenomenon (e.g., [4,1,2,5–8]) show its robustness and hint at its fundamental relationship to other integration and segregation phenomena in perception. Related studies have been conducted in humans (e.g., [3,2,4,1,9–13]), monkeys [14], and bats [15], 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 [16–23]. Bregman [1] appealed to Gestalt principles in explaining auditory grouping mechanisms, and auditory scene analysis in general, but recruited 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., [5], that hold that streaming is based on selective attention to a single perceptual dimension such as pitch. However, channel models fail to accommodate numerous streaming effects that are dependent on relationships among tones, such as quality, higher-order frequency relationships [24] and timbre [25]. Others (e.g., [26,2]) 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 [27] and vision [20]. Although in nature there are many kinds of gestalten, organized patterns of perception or behavior, the means by which those patterns arise is as yet unclear.

Complementary to the above psychophysiological approaches, 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 [28]. 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 [29] among other behaviors. Both sources of neurophysiological evidence suggest that activations of larger networks are relevant for perceptual integration. In particular, they also involve neural areas with no known tonotopic architecture. Injuries of these secondary non-tonotopically organized areas typically result in failure of integration of the environmental information, despite proper signal processing in the tonotopically organized primary areas. Given this neurophysiological evidence, we wish to explore if the architecture suggested by neurophysiology can shed light on the mechanism of streaming. We propose an architecture of two subnetworks, of which the first is tonotopically organized and the second is not. Together the two subnetworks achieve the integration of environmental signals. We will refer to models of this architecture as *stream classification* models.

For proof of concept, we propose a specific minimal realization of a stream classification model employing two self-organized systems. In the first system, a neural activity distribution is defined over space and disperses over time. This definition essentially has been referred to as a neural field [30–33]. We assume that the neural field is tonotopically organized such that the frequency of the acoustic stimulus maps onto a location in the neural space. The second non-tonotopically organized system classifies the resulting spatiotemporal neural field dynamics. This classification process is not just a measurement (else the application of a simple measure to the neural field would suffice) but is itself a dynamic process. In fact, bistability and hysteresis are properties of the classification process rather than properties of the neural field dynamics.

The dynamics of the classification system is motivated from models used in pattern recognition which themselves were inspired by simplified principles of neural functioning [34–37]. In the present case, categories available in the classification system correspond to the perception of either one stream or two. This is in contrast to a variety of models which implement the competition of individual streams as a mechanism for stream formation rather than the competition of classification patterns. Such *stream competition* models are composed of subnetworks each representing the material substrate of a stream. When input signals are provided to the subnetworks, they compete for the energy in the input signals until one, two or even more streams emerge as final states of the competition. These streams will all differ in pitch, because the underlying subnetworks have a winner-takes-all architecture and make it impossible to have stationary activity at the same pitch. Examples of stream competition networks are the cochlear models and filter banks [38,39] and the models based on Adaptive Resonance Theory [40–43].

By construction, the stream competition models cannot explain phenomena such as the existence of two streams with the same pitch [44] or duplex perception in stream formation [45]. Both phenomena, however, have recently found entry into the streaming literature. As a proof of existence of multiple streams with the same pitch but different amplitude, as well as duplex function of tonal elements belonging to two streams simultaneously, we provide additional experimental results that unambiguously demonstrate multiple streams. We implement the phenomenon computationally within the stream classification model introduced here and establish evidence for its capability of multiple stream segregation for same-pitched streams. The additional phenomena against which we test our model are the streaming phenomena established in van Noorden's bifurcation diagram and the "crossing scales" phenomenon or "bouncing" percept [1] in which a sequence of interleaved rising and falling tone trajectories are perceived as either "bouncing" or "crossing" when the frequencies of the two trajectories become relatively close. Whether bouncing or crossing is perceived is dependent on the frequency–time relationships among the tones.

In this article, we compare stream competition and stream classification models conceptually in Section 2 in order to clarify the underlying mechanism for perceptual integration. Here we take a minimalist approach in that we consider the most basic differences of the integration mechanisms in the individual models. In Section 3 we develop a specific realization of a stream classification model and in Section 4 we test it against a variety of well-documented phenomena.

2. Models for stream formation

It is often quite difficult to uncover the assumptions underlying dynamic models that attempt to explain streaming phenomena. In part, this is due to the implementation of many processes (such as filter banks, envelope functions, and pitch versus frequency discrimination) which are not essential for the stream formation process per se, but are obviously necessary to ensure proper functioning of the model for realistic stimuli. For example, recent models of auditory streaming have incorporated comprehensive models of auditory perceptual processing, including cochlear filtering, binaural processing and pitch perception (e.g., [46,47]). For conceptual clarity, we have opted for a simpler pitch–time input representation that assumes a certain amount of pre-preprocessing of acoustic signals. Although extensive modeling of earlier auditory processes may facilitate simulation for certain types of input signals, our simplification allows us to focus solely on those aspect of acoustic stimulation that are most relevant to the current proposal. Future modeling efforts will incorporate more comprehensive auditory modeling. At this stage, our strategy is simply to reduce the most common models, stream competition and stream classification, to the minimum needed to generate the main streaming phenomena for the simplest stimuli. The goal is to uncover the most basic assumptions of each model and hence their conceptual basis.

As illustrated in Fig. 1, stimulus patterns have spatial and temporal features and each model must capture or code these features. We associate the formation of a percept with the evolution of some neural activity $\psi(x,t)$, where x denotes a location in the neural space, Γ, and t an arbitrary time point. In particular, we assume that the neural space Γ is tonotopically organized. At this stage, for the purpose of a conceptual discussion, it is sufficient to state that the neural activity $\psi(x,t)$ processes the spectral stimulus features and varies as a function of time. In what follows we discuss the two major conceptualizations of how percepts form from neural activations. In the first conceptualization, the percept is realized as a projection of the neural activity $\psi(x,t)$ on a perceptual pattern $v_i(x)$. For example, if $v_1(x)$ is the pattern corresponding to the perception of one stream, then it has a non-zero component at a location x_1 and is zero otherwise. If $v_2(x)$ is the pattern corresponding to the perception of two streams, then it has non-zero components at x_1 and x_2 and is zero otherwise. The projections of $\psi(x,t)$ onto these patterns are in competition with each other. Most models fall in this class [41,39,48] and are stream competition models. In the second conceptualization, the percept is not inherent in the processing of the neural activity $\psi(x,t)$, but rather realized by the neural activation in a different non-tonotopically organized cortical area. The latter activation may represent the onset of synchronization or the emergence of an activation pattern. This is the basis of stream classification models.

2.1. Stream competition

An input signal $p(x,t) : \mathbf{R}^2 \to \mathbf{R}$ codes the spectral components in the space x and the temporal component in the time t. It is introduced into a neural network with n subnetworks and its corresponding activations $\psi(x,t) = (\psi_1(x,t), \psi_2(x,t), \ldots, \psi_n(x,t))$, where $\psi_i(x,t) : \mathbf{R}^2 \to \mathbf{R}^m$. The index m denotes the dimension of an individual network node at location x.

Each subnetwork represents a stream which is in competition with the other streams. The coupling between the streams is of "winner-takes-all" type for a given frequency, or equivalently, location x. This type of coupling is generally accomplished by local excitation and global inhibition [41,39,49]. As a consequence, only one stream will emerge as the winner in the competition for an activation at a particular location x, though there may be transients during which two streams have activations at the same location x. See Fig. 3, which illustrates the model structure for two subnetworks, that is, two streams. The generic equations governing the dynamics of stream competition are of the following form:

$$\dot{\psi}_i(x,t) = G_i(\psi_i(x,t)) - \sum_j C_{ij}\psi_j(x,t) + Mp(x,t) \tag{1}$$

where G_i is a nonlinear functional with spatial operators describing the local dynamics at each network node, $C_{ij} > 0$ is a constant and quantifies the coupling between individual streams and M is a constant $m \times 1$ matrix introducing the input uniformly to each node. The competition among streams is established entirely through the inhibitory coupling matrix C_{ij}. The streams compete for the energy in the input signals at the same location x and will mutually inhibit each other, resulting in the exclusive activation of one stream and deactivation of all other

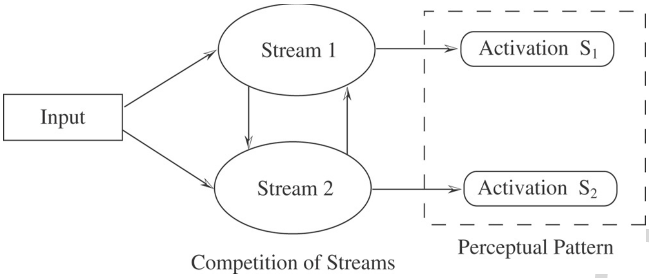

Fig. 3. Stream competition model structure.

streams. As a consequence, an activity distribution $v_i(x)$ emerges for the i-th stream with amplitude $s_i(t)$ across the neural space x. All the information about the perceptual pattern is contained in the set of streams $\{s_i(t)v_i(x)\}$ being activated.

The models suggested by Grossberg and colleagues [41,43] and by McCabe and Denham [39] can both be written in the form of Eq. (1). Obviously, there is a large amount of detail implemented in these models to make them function for realistic stimuli. In particular, the ARTSTREAM model of Grossberg and colleagues is a sophisticated model incorporating mechanisms for pitch formation and symmetry breaking of the inhibitory coupling among the streams to induce a stream formation hierarchy. These technical aspects are crucial for proper functioning but not conceptually.

2.2. Stream classification

This approach to stream segregation assumes a separate component with the activity $y(t) = (\cdots y_i(t) \cdots)$ which is responsible for the integration (or classification) of the spatiotemporally dispersed input signal $\psi(x,t)$. The activation of the i-th component, $y_i(t)$, represents the formation of the percept i and may be realized in two ways: (1) by means of a separate area or network corresponding to an informational convergence zone; or (2) by means of a network of areas that can be synchronized as a function of the input signal. In both cases, $y_i(t)$ is activated (via an amplitude increase in the first case or onset of synchrony in the second case) when the i-th percept is integrated. However, in contrast to the stream competition models, processing of spectral and temporal properties of the input and percept formation do not occur at the same site. More specifically, the auditory input signal does not feed directly into the integrating network $y(t)$, but is first spatiotemporally dispersed in an intermediate and tonotopically organized neural network with activity $\psi(x,t)$. Its dynamics is described via:

$$\dot{\psi}(x,t) = G(\psi(x,t)) + p(x,t) \tag{2}$$

where, as before, G is a nonlinear functional with spatial operators describing the local dynamics at each network node and $p(x,t)$ is the input introduced uniformly to each node. A measure $I_i(t)$ acting on $\psi(x,t)$ is then passed on to $y_i(t)$

$$\dot{y}_i(t) = g_i(y_i) - \sum_{j \neq i} c_{ij}(y_i - y_j) + I_i(t) \tag{3}$$

where g_i is a nonlinear function denoting the intrinsic dynamics of $y_i(t)$ and $c_{ij}(y_i - y_j)$ is the coupling with the nonlinear, typically sigmoidal, function c_{ij}. The measure $I_i(t) = I_i(\psi(x,t))$ is a functional which is introduced to the i-th component with activity $y_i(t)$. Note that $y_i(t)$ in a real network is not limited to the activation of a single node, but may equivalently represent an activity pattern of a network via a simple linear transformation. In this approach, the neural activation $y_i(t)$ corresponds to the strength of a perceptual pattern. By definition, two perceptual patterns cannot be present at the same time and hence the system described in Eq. (3) justifies a winner-takes-all dynamics. Here the competition is between perceptual patterns, not streams, which allows the co-existence

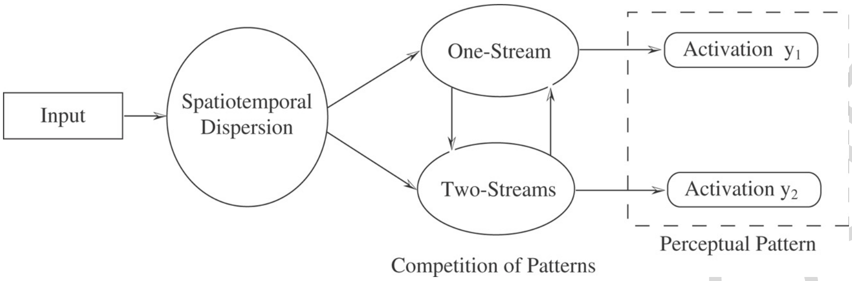

Fig. 4. Stream classification model structure.

of two streams with the same pitch consistent with previous experimental findings by Singh [44]. See Fig. 4 for an illustration of this model structure.

3. Specific realization of a stream classification model

In the following paragraphs we develop the details of a stream classification model. For proof of concept, we will establish both networks, the dispersive neural field $\psi(x,t)$ and the network of competing patterns $y_i(t)$, in its simplest mathematical form, but still sufficiently rich to reproduce a majority of well-known streaming phenomena.

The dynamics of the neural field $\psi(x,t)$ is defined as a spatiotemporal wave equation following the lines of Jirsa and Haken [32,33], but certainly not limited to this particular choice (see [49–52,31] for related approaches). Here a continuous sheet of excitatory and inhibitory neural ensembles is connected over short and long distances via nerve cell fibers represented by an exponentially decaying connectivity function, which relates the neurons' activity at different locations. Time delays in neural cross talk are caused by the time needed for the neural activity to propagate from one location to another. The neural activity is generated by the nonlinear sum of the incoming signals which are weighted by the connection strengths. The nonlinearity characterizes the response function of a neural ensemble and is typically sigmoidal in character. The neural field activity has the ability to show spatiotemporal pattern formation in terms of travelling wave fronts, standing waves, and more complicated spatiotemporal dynamics. Under the assumption of a translationally invariant connectivity function, the dynamics of the neural field, $\psi(x,t)$, is governed by a partial differential equation [32,33,53] as follows:

$$\begin{aligned}&\ddot{\psi}+(\omega_0^2-\nu^2\Delta)\psi+2\omega\dot{\psi}=a_e\left(\omega_0^2+\omega_0\frac{\partial}{\partial t}\right)\cdot S(\psi+p)\\&\psi:\mathbf{R}^2\to\mathbf{R}\quad p:\mathbf{R}^2\to\mathbf{R}.\end{aligned}\tag{4}$$

Here n dots denote the n-th derivative with respect to time, Δ is the Laplacian. a_e, ω_0, and ν are constants, $S:\mathbf{R}\to\mathbf{R}$ is a sigmoid function, and $p:\mathbf{R}^2\to\mathbf{R}$ is the external input or stimulus to the neural sheet. Periodic boundary conditions, $\psi(0,t)=\psi(L,t)$, $t\geq 0$, are used. Other boundary conditions provide identical results, because the neural field activations are local compared to the entire length L of the cortical sheet. Initial conditions are chosen to be random. The parameters of the neural field equation are physiologically interpreted as $a_e=\alpha\mu\rho_e$, where α is the coefficient of the membrane impulse response, μ is the neural membrane time constant, and ρ_e is the space constant of connectivity. The parameter ω_0 is defined as $\omega_0=\nu/\sigma$, where ν is the corticocortical signal propagation velocity and σ is the long range connectivity of fibers. The spatiotemporal characteristics of the auditory input are represented by $p(x,t)$. The sigmoid function S represents the firing rate of a neural ensemble and is defined as: $S(x)=2/(1+\mathrm{e}^{-x})-1$.

The second network is established in terms of the variables $y_i(t)$. In the case of the competition of two streams, there will be only the two perceptual patterns of one-stream with the strength $y_1(t)$ and the pattern of two-streams

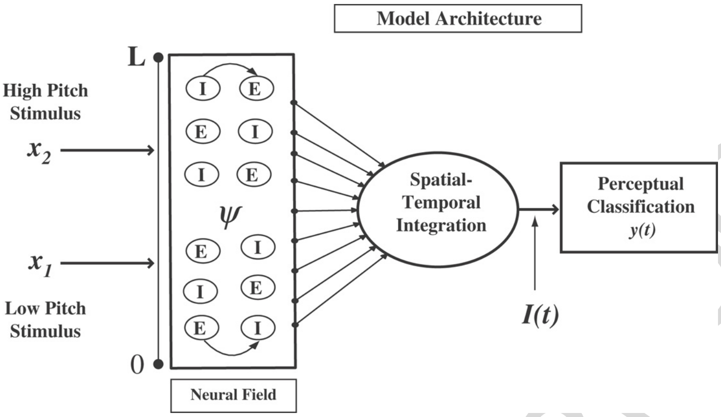

Fig. 5. Model architecture. Stimuli in the form of high and low tone sequences are provided to a one-dimensional neural field $\psi(x,t)$. The resulting neural activity is spatiotemporally integrated and transmitted as the signal $I(t)$ to a classification system $y(t)$.

with the strength $y_2(t)$. Due to the mutual exclusion of these patterns, their difference, $y = y_1 - y_2$, will either be $y = y_1 > 0$ or $y = y_2 < 0$ once the dynamics has become stationary. Then the dynamics of $y(t)$ may be written as:

$$\dot{y}(t) = \dot{y}_1(t) - \dot{y}_2(t) = g(y_1, y_2) - c(y) + I(t) \tag{5}$$

where $g(y_1, y_2) = g_1(y_1) - g_2(y_2)$, $I(t) = I_1 - I_2$ and $c(y) = c_{12}(y) - c_{21}(-y)$.

For simplicity, we choose the dynamics of a node, that is g_1 and g_2, to be linear. We expand the sigmoidal coupling function $c(y) = c_0 + c_1 y + c_2 y^2 + c_3 y^3 + \cdots$ and write the Eq. (5) now in a closed form:

$$\dot{y} = \varepsilon y - y^3 - I_0 + I(t) \tag{6}$$

where $c_3 = 1$ and ε is a constant that captures all linear contributions. I_0 contains all constant contributions, including the one stemming from a transformation of the quadratic term $c_2 y^2$. The functional $I(t)$ is now specified as:

$$I(t) = \int_0^L h(\psi(x,t))\mathrm{d}x \tag{7}$$

where Ω is a neural activity threshold and h is a rectified function:

$$h(n) = \begin{cases} 0, & n \leq \Omega \\ n, & n > \Omega. \end{cases} \tag{8}$$

The Eqs. (4), (6) and (7) define the dynamics of a stream classification model in one of its simplest forms. Fig. 5 illustrates the architecture of the model. The neural field is illustrated by the rectangular box showing the neural activity $\psi(x,t)$ composed of inhibitory and excitatory neurons. The input $p(x,t)$ is provided at locations x_i via the Gaussian localization function $\mathrm{e}^{-(x-x_i)^2/\delta_i}$ with width $\sqrt{\delta_i}$. The explicit model parameters used in the simulations are given in Table 1.

4. Results

4.1. van Noorden diagram

To understand van Noorden's bifurcation diagram as shown in Fig. 2, 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

Table 1
Table of model parameter values

Parameter	Value
δ_1	0.025
δ_2	0.025
v	0.2
σ	0.25
ω_0	$\frac{v}{\sigma}$
L	6
Ω	0.02
ε	0.6
I_0	3

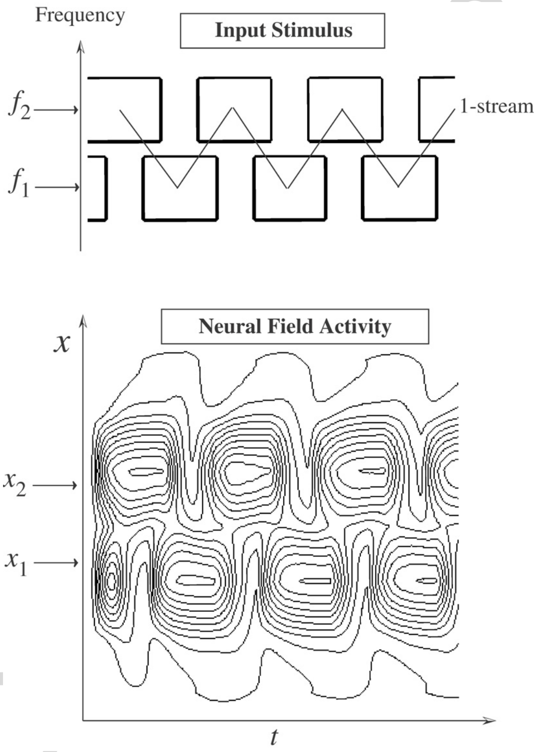

Fig. 6. The stimulus sequences (top) and its resulting neural field dynamics (bottom). The percept of one stream.

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 ψ 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. Figs. 6 through 8 plot the contour lines of neural field activity over space x and time t for three situations, that is, one integrated stream (Fig. 6), the bistable regime (Fig. 7) and two separate streams (Fig. 8).

The final state reached by the second system defined in Eq. (6) with activity y will depend on $I(t)$ and its own intrinsic dynamics. In Fig. 9 the flow (time derivative) $\dot{y}$ is plotted as a function of y. The intersections of the flow with the horizontal axis define the fixed points of y. The curve of the flow is shifted up or down depending on

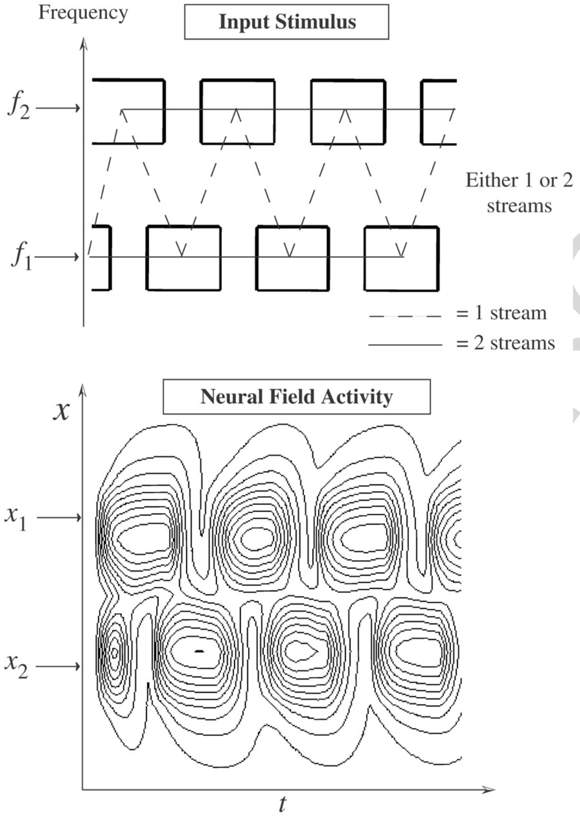

Fig. 7. The stimulus sequences (top) and its resulting neural field dynamics (bottom). Bistable regime.

$I(t)$, creating either one positive or one negative fixed point. 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 Fig. 10 for several different initial conditions of the activity y. After a transient time the activity becomes stationary, displaying three possible scenarios (see Fig. 10 from top to bottom): one stream only, or the bistable situation, in which either one integrated stream or two separate streams may be perceived, or finally two streams only. This simulation highlights the model's character as a classification model.

For each choice of Δf and IOI, the model Eqs. (4) and (6) are numerically solved and their stationary states determined. The results are plotted in the two-dimensional parameter space in Fig. 11. The temporal coherence boundary (TCB) and the fission boundary (FB) are reproduced in a manner that corresponds strongly to van Noorden's [2] results including a bistable region. Note that the exact experimental numerical values at which the boundaries occur vary from subject to subject and depend on the experimental methods employed [1].

4.2. Crossing scales phenomenon

When two interleaved rising and falling tone sequences as shown in Fig. 12 are presented, then human subjects report these to be either crossing or bouncing perceptually [54,1]. This phenomenon is known as the "crossing scales phenomenon". In particular, the bouncing percept is more often reported than the crossing percept. Similar phenomena have also been reported in the visual domain [55]. Fig. 12 (top) shows the interleaved rising and falling tone sequences which serve as input to the neural field. Note that these input sequences change across space

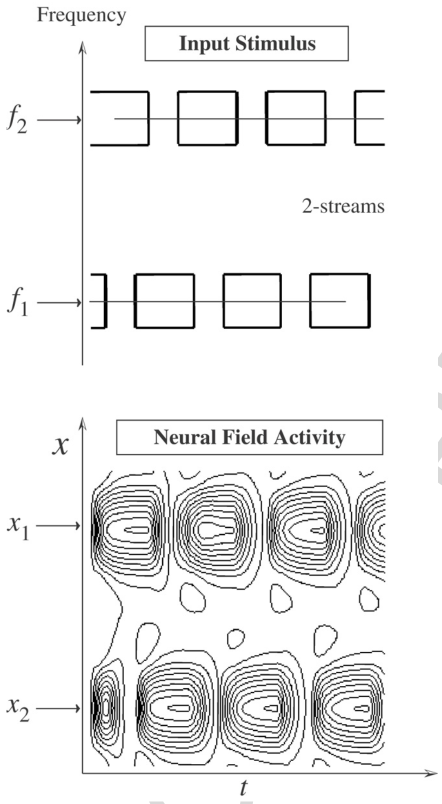

Fig. 8. The stimulus sequences (top) and its resulting neural field dynamics (bottom). The percept of two streams.

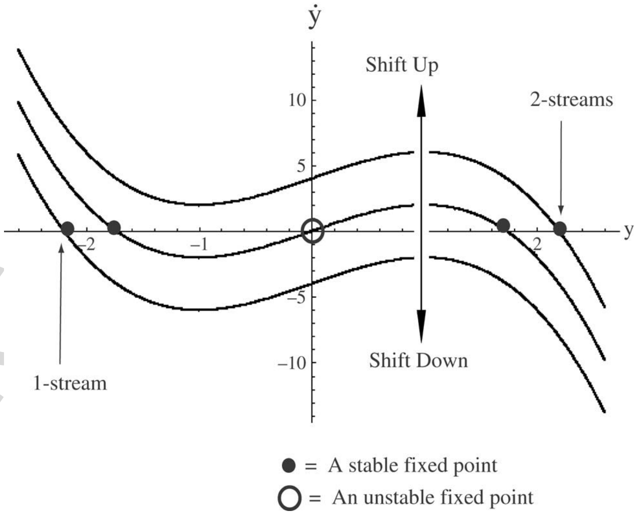

Fig. 9. The flow $\dot{y}$ of the classification system (plotted as a function of y) is shifted up or down in dependence of $I(t)$. The intersections with the horizontal axis characterizes stable (full circles) and unstable (empty circle) fixed points of the system. Left (negative) fixed point: one stream. Right (positive) fixed point: two streams. Three fixed points (two stable, one unstable): bistable regime.

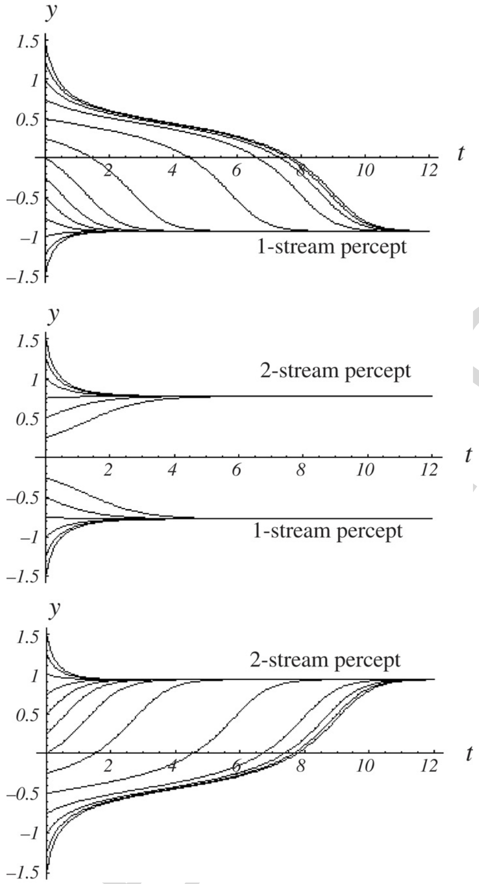

Fig. 10. For multiple initial conditions the time series of $y(t)$ are plotted for the three regimes, one stream only (top), bistable (middle) and two streams only (bottom).

(i.e., frequency in this case) and time by construction and can be characterized by the angle θ that they form. It is computationally more convenient to describe the dynamics of the neural field as a function of the IOI and θ because the latter captures the dispersive properties of the dynamics more naturally in this particular experimental design. But also note that when θ remains constant while varying Δf and the IOI, then the resulting perceptual dynamics will not change. It is implied that the angle θ is the more relevant perceptual control parameter for the "bouncing percept" phenomenon. Fig. 12b shows the resulting dynamics of the neural field under the influence of this input. The corresponding dynamics of the second system, the classification system, is illustrated in Fig. 12c for the phenomenon of the "crossing percept". Following the presentation of the input, the perceptual pattern analogous to the perception of two streams emerges. As the successive input tones get closer in frequency, the positive fixed point destabilizes and a transition occurs towards the negative fixed point analogous to the perception of a single stream. As successive tones continue to diverge in frequency, the system returns towards the positive fixed point. Fig. 12c illustrates the successful formation of the sequence from two-streams to one-stream and back to two-streams. We identify this sequence with the perception of crossover. Fig. 12d illustrates the complementary phenomenon of the bouncing percept, which is represented by the failure of $y(t)$ to change its sign, i.e., the percept remains that of two-streams. After all, during bouncing, perception is always of two non-interacting streams. The transition from the positive to the negative fixed point is not guaranteed to occur, but depends on the characteristics of the input

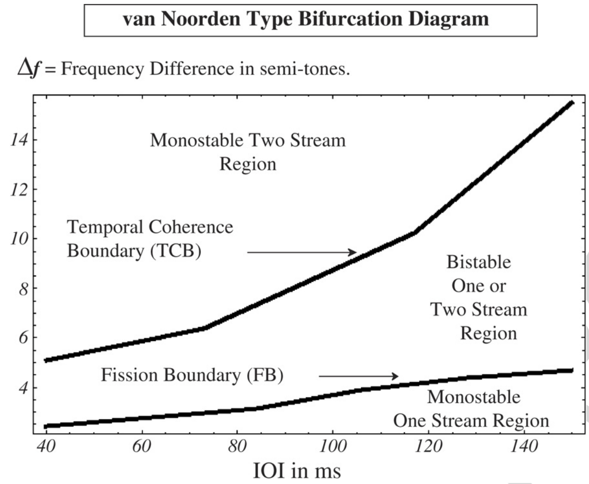

Fig. 11. A partial diagram similar to van Noorden's bifurcation diagram is obtained numerically from the model and compares favorably with the experimental results (see Fig. 2).

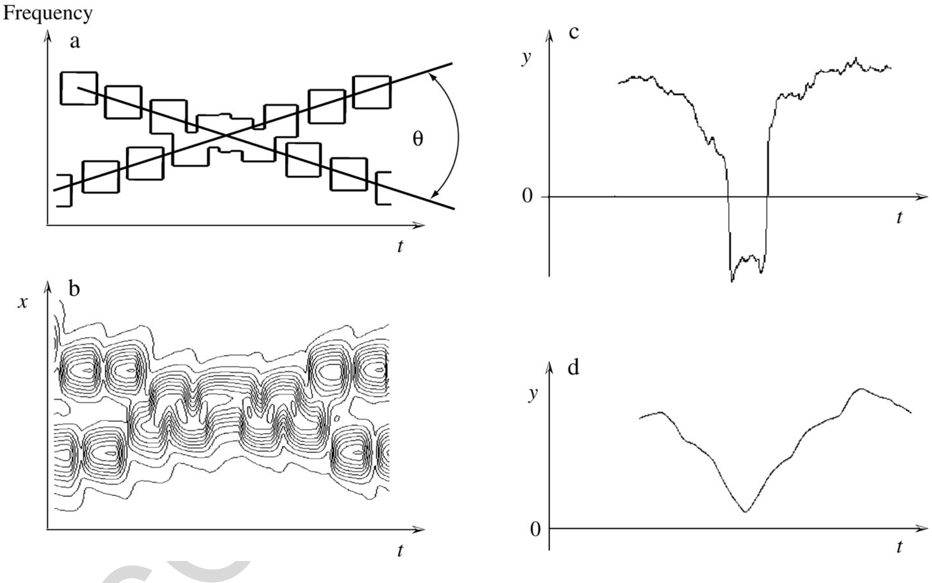

Fig. 12. The input tone sequences used to form a percept of *crossover* are shown in a. The resulting contours of neural field activity are plotted in b, together with the final time series of $y(t)$ shown in c. In this particular case, the classification system $y(t)$ does traverse from the positive (two streams) to the negative (one stream) fixed point and back. This trajectory is identified with the percept of crossover. In the case of the *bouncing percept* the time series of $y(t)$ will not cross the x-axis, as shown in d.

tone sequences (frequency–time relationships, timbre [1], etc.) and the intrinsic dynamics of the two systems and their interactions. Note that the model also predicts a potential hysteresis effect in that the transition from one fixed point to the other only occurs when either of the boundaries in the parameter space is crossed. These novel predictions and the dependence of the dynamics on the angle θ are open to be tested experimentally.

Applying the above identifications of dynamic behavior to stream perception, a bifurcation diagram (see Fig. 13) can be constructed contrasting the IOI of the tone sequences versus the angle θ that they form at the crossover

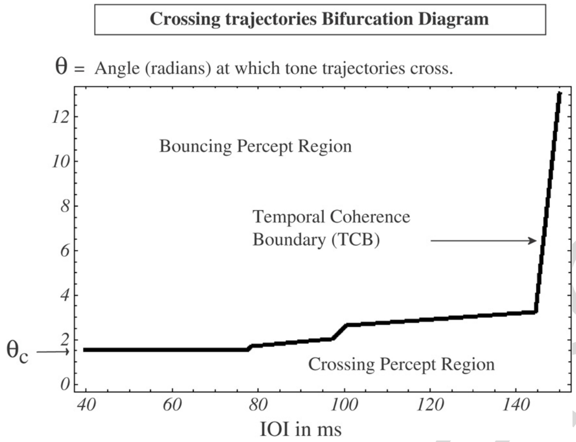

Fig. 13. A bifurcation diagram obtained numerically from the model using crossing tone sequences. The diagram is obtained by varying the IOI and the angle θ formed by the crossing sequences. Note that below a critical value θ_c of the angle θ there is always crossover, as defined in the main text.

point. The control parameters are the angle θ and IOI. All other parameters are fixed, including the tone duration. The result predicts two regions separated by a temporal coherence boundary (TCB). Only below the TCB can crossover percepts occur. Above the TCB, bouncing percepts always occur. Note that for a critical value θ_c of the angle θ in Fig. 13, there is no possibility of a bouncing percept below this value. This is consistent with the experimental and theoretical bifurcation diagrams shown in Figs. 2 and 11. A bouncing percept always occurs whenever the frequency separation Δf is large enough given relatively small IOI values. The actual existence of an experimentally derived bifurcation diagram that is topologically equivalent to 13 needs to be demonstrated to show the validity of our claims.

Based on the same principles we can predict a complementary phenomenon to bouncing and crossover. Following the conventions in Fig. 12, Fig. 14 shows the input tone sequence (top), the resulting dynamics of the neural field (middle), and the time series of the classification system (bottom). In this case, the interleaved input tone sequences initially have a small frequency separation which diverges over time and then returns to the original configuration. The corresponding progressive stabilization and destabilization from one-stream to two-streams and back again to one-stream, we identify with the percept of “splitting and merging”.

4.3. Two streams with the same pitch

4.3.1. Theoretical treatment

When simple tone sequences of alternating amplitude tones (see Fig. 15) are presented to human subjects the percept of two streams can still be perceived even though the frequencies of the alternating tones can be the same or relatively close in frequency ([44] and Section 4.3.2 (subsection 2 of this section “experimental proof of existence”)). From a biologicalperspective, an important reason for considering such tone sequences is that it has been suggested that temporal modulations of acoustic stimuli “are a critical stimulus attribute that assists us in the detection, discrimination, identification, parsing, and localization of acoustic sources and that this wide ranging role is reflected in dedicated physiological properties at different anatomical levels” ([56], p. 542). Thus, there may be a topographic ordering of neurons according to amplitude modulated tuning.

The stream classification model we have presented is highly motivated by the relationship between cortical activity induced by a stimulus and the resulting perception. Thus, we expect the model to capture some of the possibilities for streaming in the context of alternating amplitude tones. The main reason for such expectation is that the most significant factor affecting the neural field dynamics implicit in the model is the magnitude and

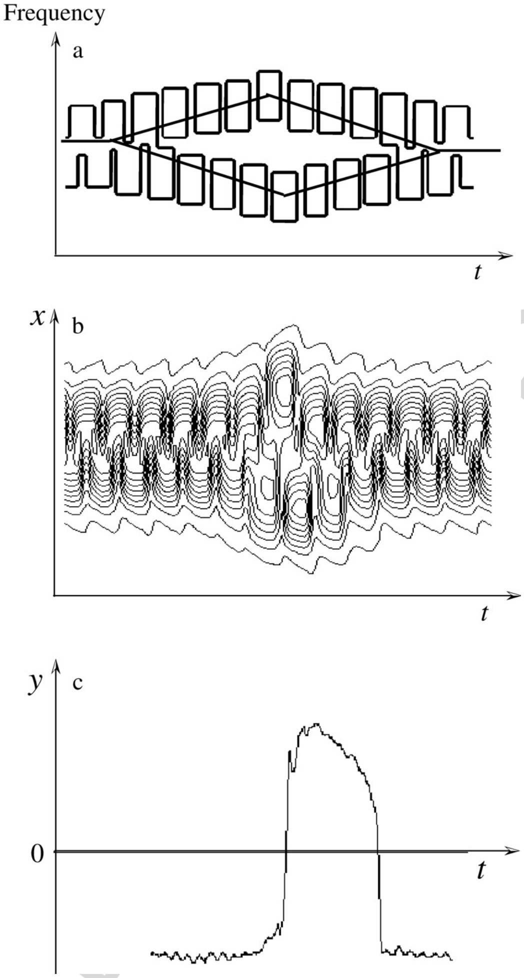

Fig. 14. An illustration of an input pattern is shown in a which results in an inverted crossover or separation effect. b shows the contours of the resulting neural field activity and c the time series of the classification system $y(t)$. The time series $y(t)$ converges from the negative (one stream) fixed point to the positive (two streams) fixed point and back again.

distribution of "stimulus energy" it receives. This essentially translates to the relationship implicit in the model that the higher the amplitude (or magnitude) of a stimulus the larger the contribution it makes to the resulting neural field activity. That relationship has a direct impact on the functional $I(t)$ (Eq. (7)), which in turn affects the dynamics of the classification system $y(t)$ (Eq. (6)). Simply put, the model is sensitive to stimulus magnitude changes.

To test the model's predictive possibilities, simulations were run using sequences of simple alternating amplitude tones, as illustrated in Fig. 15. In each sequence, adjacent tones were set to have a constant amplitude ratio (AR). Across sequences, the amplitude of one of the adjacent tones is kept constant but the amplitude of the intervening tone varies by a constant factor. To account for backward and forward masking, a small silent gap of constant duration was introduced between tones.

The numerical simulations predict the possibility for perception of two streams. Also, the simulations show that one stream and a bistable region reminiscent (in topological structure only) of the van Noorden diagram are also possible. For the case of zero frequency separation ($\Delta f = 0$) between the tones and an amplitude ratio of two

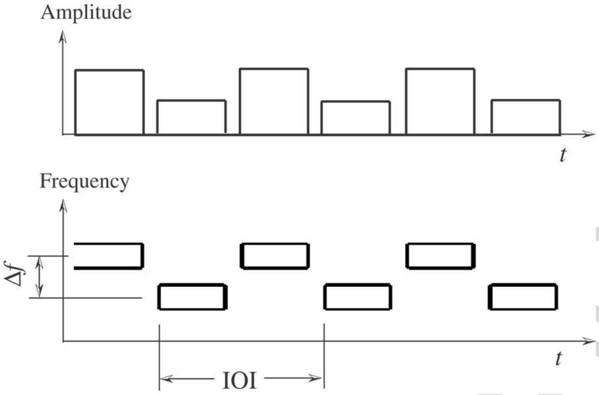

Fig. 15. This figure illustrates the structure of *alternating amplitude simple tone sequences*. In this illustration all tones have nearly the same frequency and adjacent tones have a constant amplitude ratio AR.

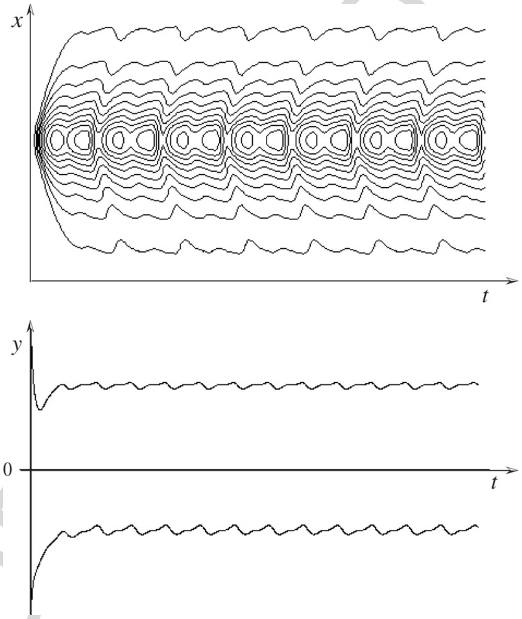

Fig. 16. Contour plot of the neural field activity under the influence of a stimulus consisting of simple alternating amplitude tone sequences. The corresponding time series for the classification system y are shown below. They indicate a bistable percept of one or two streams.

(AR = Fixed Amplitude/Variable Amplitude $= 1/0.5 = 2$) we obtained Fig. 16, showing the resulting dynamics of the neural field (top figure) and the corresponding dynamics (time series) of the classification system y for a negative and positive initial condition (bottom figure).

The scenario depicted by Fig. 16 shows the bistable possibility of one or two stream(s) as indicated by the dynamics of the classification system y. That is, y has both positive and negative attractors. Fig. 17 shows bifurcation diagrams for AR vs. IOI for some values of Δf.

The thick lines indicate a region where the time series of the classification system y oscillates between one stream and two streams, that is, $y(t)$ goes from positive to negative and back again repetitively. The reason for

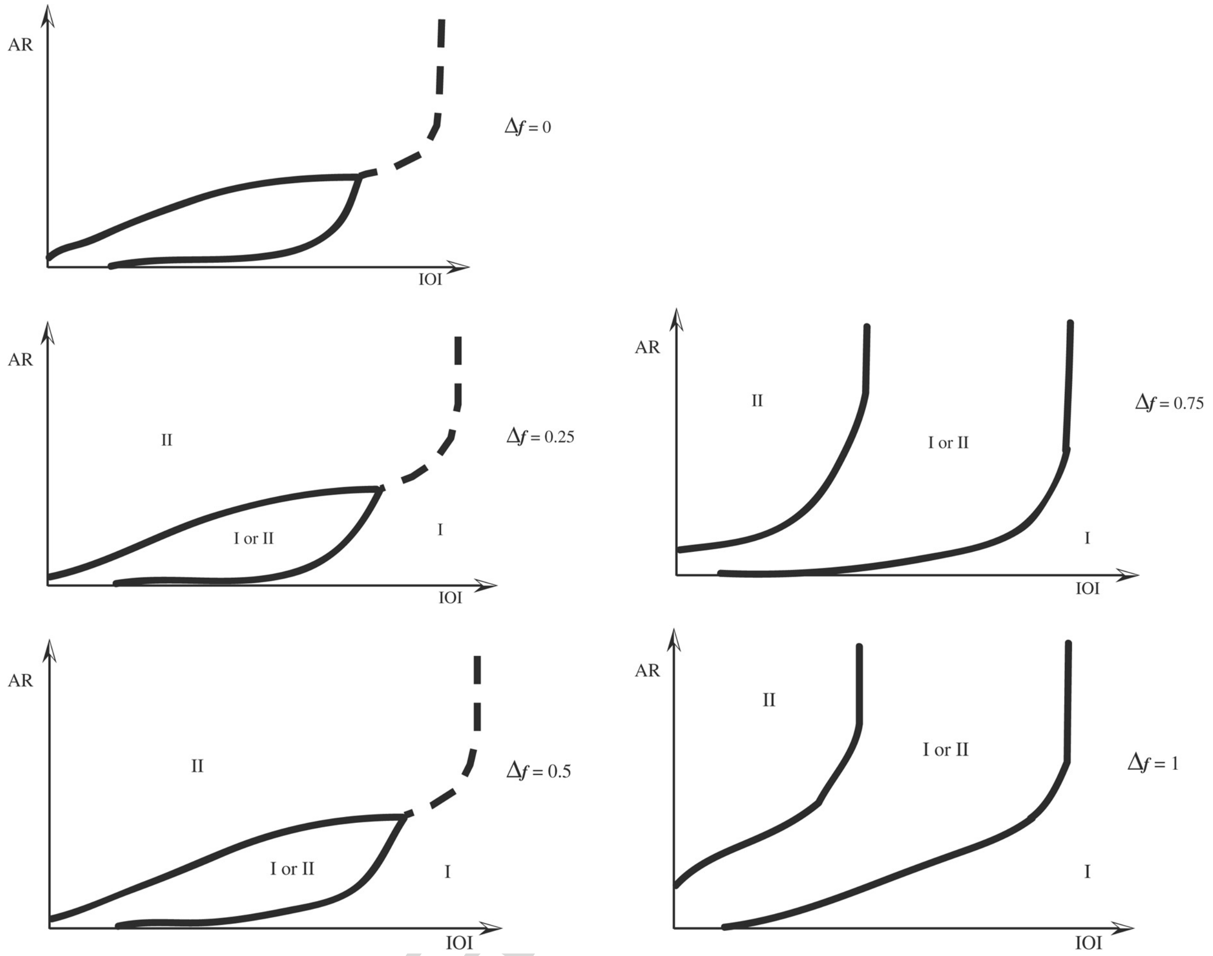

Fig. 17. The ratio of two amplitudes (AR) is plotted as a function of the inter onset interval (IOI) for five values of the frequency separation Δf. The graphs represent the corresponding bifurcation diagram for each fixed set of parameters.

this is that as the dynamics evolve the high amplitude tones in the sequences introduce enough energy to drive the system into the two-stream region while the low amplitude tones only contribute enough energy to take the system to the one-stream region. It should also be noted that this effect depends on the intrinsic time scale of the system which must be sufficiently fast. Fig. 17 also shows that there is a bifurcation along the Δf dimension with the bifurcation diagrams for AR vs. IOI becoming more closer, in the topological sense, to the van Noorden diagram as Δf increases. The meaning of this can be interpreted as theoretical evidence for the existence of a (nonlinear) relationship between frequency and amplitude in streaming and possibly in auditory perceptual phenomena in general.

4.3.2. Experimental proof of existence

Here we describe an experiment demonstrating that the streaming phenomenon can occur with sinusoidal stimuli that differ only in amplitude. Although streaming has been demonstrated in amplitude modulated stimulus sequences comprised of complex tones (e.g., [56]), here we examine perception of simple sinusoidal tones of constant frequency but with systematic amplitude modulation.

Experimental methods

A. Subjects

Eleven subjects participated. All had normal hearing according to self-report and were naive to the purpose of the experiment.

B. Stimuli

Stimulus tones were always 600 Hz sine waves of 100 ms duration. Amplitude was linearly ramped over the first and last 10 ms, with the central 80 ms at constant amplitude. In the first condition, each tone sequence consisted of 20 pairs of tones T1–T2 with an IOI of 110 ms between all tones. The only parameter varied across tone sequences was the amplitude ratio (AR) between adjacent tones T1 and T2. AR was kept constant within 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). The choice of a 10 ms gap between successive tones of a stimulus sequence was based on the constraint that the streaming phenomenon should be perceivable for some range of AR values even if there is some amount of forward and backward masking occurring between adjacent tones.

In a second, control condition, each trial consisted of three short tone sequences presented with a 500 ms IOI between sequences. Each sequence was either (1) a T1–T2 sequence with AR value of 0.2 (intertone $IOI = 110$ ms), or (2) a sequence with T1 being the reference amplitude and intertone $IOI = 220$ ms, i.e., there is no T2. Triad orders were sequences (a) 1, 2, 1; (b) 2, 1, 1; (c) 1, 2, 2; and (d) 2, 1, 2. Sequence 1 was comprised of 12 tones whereas sequence 2 was comprised of 6 tones, with double the IOI. This condition was to verify that subjects in fact heard the lowest amplitude tones.

C. Procedure

In the experimental condition, pairs of tone sequences were presented 5 times to subjects in a random order, with 500 ms between members of each sequence pair. 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. In total, there were 230 sequence pairs. Order of sequence presentation for each pair was counterbalanced. The subjects' task was to judge which of the two sequences in a pair sounded slower. Tone sequences of the type used here that are perceived as two streams are also perceived to be slower than sequences that are perceived as a single stream. Subjects responded by pressing one of three labelled keys on the numeric keypad of the keyboard. 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 were perceived as equal in rate.

Subjects were told that there were no right or wrong answers and that our interest was purely in how they perceived the tone sequences. They were given 10 min rest breaks after the 115th trial.

The control condition 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 indicated their responses by pressing one of two labelled keys on the numeric keypad of the keyboard. Pressing "1" indicated 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.

D. Results and discussion

Responses from the control condition were examined to verify that all subjects were able to hear the lowest amplitude tone reliably. Seven of the subjects responded with 100% accuracy, 2 subjects responded with 95% accuracy, 1 subject with 90% accuracy, and 1 subject with 85% accuracy. Thus, subjects correctly noted the slower sequence, indicating that the low amplitude tone was perceived and contributed to the perception of a faster rate sequence.

Fig. 18 shows the means and standard deviations of the responses across subjects. 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. This pattern indicates that streaming has occurred in the first

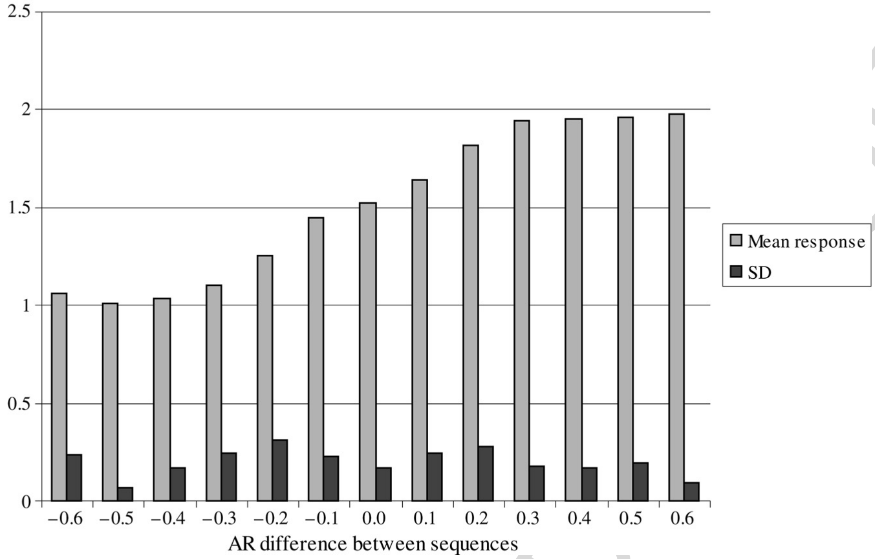

Fig. 18. Alternating amplitude tones streaming: experimental results. Means and standard deviations of the responses across subjects. The x-axis (bar labels) represents stimulus sequence pairs excluding order of presentation.

sequence but not in the second. If streaming occurred in both sequences, subjects would respond that both are of equal rate (denoted on the y-axis by 1.5 in Fig. 18).

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.

Note the lack of an order effect: AR differences of −0.6, for example, are equivalent to AR differences of 0.6, but with the sequences presented in the opposite order. The lack of an order effect was confirmed statistically by a two-way repeated measures ANOVA with order and AR difference as factors, $F(1, 5) = 2.67$, $p > 0.1$. The interaction of order and AR difference was similarly not significant, $F(5, 5) = 1.88$, $p > 0.1$.

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. However, the prediction that a bistable one-or-two stream scenario is possible was not assessed by this paradigm.

4.4. *Timing implications of the model*

Experimental results have demonstrated that auditory streaming becomes more pronounced when the rate of alternation of the input tone sequences increases [1]. There are various time intervals that affect the rate of alternation. Most experiments do not vary these time intervals independently of one another, so there is no direct indication which of these time intervals is responsible for the effects of rate on auditory streaming. In our model the intervals that affect streaming are offset to onset of successive tones within and between streams (B and D in Fig. 19). Interestingly, Gestalt theorists also considered these intervals to be important, because they determine temporal proximity. In contrast, Jones [26], for example, considers onset to onset of successive tones between and within streams (A and C in Fig. 19) as the most relevant intervals, because they determine the regularity or rhythmicity of the sequence. As stated earlier, the mechanism of our model is based on the dispersive properties

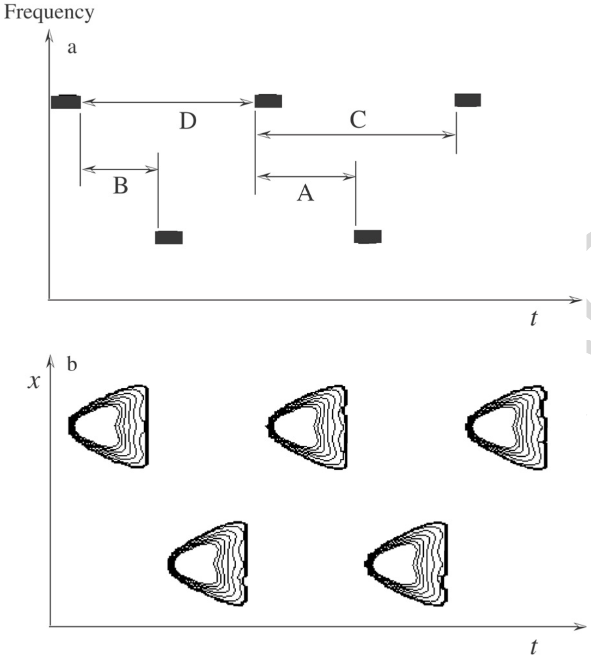

Fig. 19. The definition of the time intervals present in a common streaming experiment are illustrated in a: A = Onset-to-onset time interval between temporally adjacent tones. B = Silent gap between temporally adjacent tones. C = Onset-to-onset time interval between tones in the same frequency range. D = Offset-to-onset time interval between tones in the same frequency range. b shows the induced neural field activity for the above input tone sequence and illustrates the decay time of the neural field after offset of a stimulus tone. Intervals B and D are the most significant for the streaming model presented.

of the neural field, or more precisely on the spatial and temporal decay rates of neural processing. These decay rates account for the grouping of tones based on their proximity in time and frequency. In particular, as tones get closer to each other along the time axis by shortening the time interval D, the likelihood that consecutive tones will be grouped into one stream increases, because of the finite decay time for neural activity to spread as illustrated in Fig. 19. As a consequence, the relevant time interval of our model is D within one stream and is B between streams.

5. Discussion

There is evidence from neurophysiology that cortical and subcortical components are involved in the process of integration of environmental signals beyond the primary sensory areas. In particular, it appears that the recruitment of these higher areas is largely non-specific to the modalities involved, which implies a general mechanism to be in place for the integration of signals to a coherent percept. For instance, 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 [57]. Other researchers have found similar networks to be involved in the spatiotemporal integration of visual cues during collision judgments [58]. In the more complex scenario of speech perception, Hickock and Poeppel [59] argued that the left inferior parietal cortex is also involved in the integration of auditory-motor processes. Inspired by the large literature on integration phenomena across sensory modalities, here we explored the possibility of a system composed of a primary sensory component and a secondary component, that performs integration of external input signals. In particular for the auditory domain, we implemented the special characteristic of the primary sensory system, its tonotopic organization. The secondary system does not have such organization, which is in consensus with the notion that this component is largely independent of the details of the sensory modalities with an integrative function that transcends modalities. We limited our discussion of example input sequences (Sections 4.1 through 4.3) to pure tones or, equivalently, to pitches. For real-world stimuli involving more complex spectral structures, our model will require a preprocessing which identifies the pitch of a tone sequence, perhaps along the lines of the ARTSTREAM model [40,41,43] in which a SPINET model [60] performs an initial processing of the acoustic input sequence. We expect such a preprocessing stage to enable a stream classifier as presented here to reproduce phenomena with more complex spectral properties such as the continuity illusion [61]. The second limitation of the current stream classifier model is the absence of feedback from the second to the first system. Such feedback is well-known to exist in real neurophysiological systems and is involved in the mediation of attention and expectation. For the purpose of this manuscript, it is not necessary to include feedback pathways, but it is very intuitive to suggest that feedback from the second system will bias the neural field propagation in the first system and hence bias the overall dynamics of the stream classifier.

Percepts that depend on more complex temporal properties of the input sequences such as the bouncing and crossing percept can be easily reproduced by a stream classifier model. This is in sharp contrast to extant stream competition models that have difficulty reproducing crossing percepts. The extreme case is an angle of zero spanned by the two interleaved crossing stimulus trajectories. As long as there is another quantity which makes these stimulus trajectories distinct, then two streams with the same pitch may coexist perceptually. Thus, we have shown both empirically and theoretically using a stream classifier that manipulation of the amplitude ratio of successive pure tones can generate a percept of stream segregation. This effect is in sharp contrast to the nature of stream competition models.

We understand that the stream competition models have, first, some kind of intuitive appeal in terms of the notion of competition of streams, and second, have successfully reproduced many phenomena in the literature using real-world stimuli. However, neurophysiological mechanisms for segregation and integration have started to receive much attention. As we have shown here, an architecture suggested by the neurophysiological literature on convergence zones can also reproduce some of the basic streaming phenomena. It is not unlikely that a combination of these two mechanisms is present in the real biological system, in which the more peripheral systems are rather stream-competition-like and the more cortical systems are more stream-classification-like.

References

[1] A.S. Bregman, Auditory Scene Analysis: The Perceptual Organization of Sound, MIT Press, 1990.

[2] L.P.A.S. van Noorden, Temporal coherence in the perception of tone sequences, Ph.D. Thesis, Eindhoven University of Technology, The Netherlands, 1975.

[3] G.A. Miller, G.A. Heise, The trill threshold, The Journal of the Acoustical Society of America 22 (5) (1950) 637–638.

[4] A.S. Bregman, Auditory streaming: Competition among alternative organizations, Perception and Psychophysics 23 (5) (1978) 391–398.

[5] D. Broadbent, Perception and Communication, Pergamon Press, 1958.

[6] U. Neisser, Cognitive Psychology, Appleton-Century-Croft, 1967.

[7] R.P. Carlyon, How the brain separates sounds, TRENDS in Cognitive Sciences 8 (10) (2004) 465–471.

[8] R. Cusack, J. Deeks, G. Aikman, R.P. Carlyon, Effects of location, frequency region, and time course of selective attention on auditory scene analysis, Journal of Experimental Psychology: Human Perception and Performance 30 (4) (2004) 643–656.

[9] J.R. Lackner, L.M. Goldstein, Primary auditory segregation of repeated consonant-vowel sequences, The Journal of the Acoustical Society of America 56 (5) (1974) 1651–1652.

[10] M. Warren, Richard, P. Warren, Roslyn, Auditory illusions and confusions 223 (1970) 30–36.

[11] P. Helenius, K. Uutela, R. Hari, Auditory stream segregation in dyslexic adults, Brain 122 (1999) 907–913.

[12] J. Vliegen, A.J. Oxenham, Sequential stream segregation in the absence of spectral cues, The Journal of the Acoustical Society of America 105 (1) (1999) 339–346.

[13] N. Grimault, S.P. Bacon, C. Micheyl, Auditory stream segregation on the basis of amplitude-modulation rate, The Journal of the Acoustical Society of America 111 (3) (2002) 1340–1348.

[14] A. Izumi, Auditory stream segregation in Japanese monkeys, Cognition 82 (2002) B113–B122.

[15] C.F. Moss, A. Surlykke, Auditory scene analysis by echolocation in bats, The Journal of the Acoustical Society of America 110 (4) (2001) 2207–2226.

[16] M. Wertheimer, Gestalt theory, Social Research 11, Translation of lecture at the Kant Society, Berlin.

[17] K. Koffka, Principles of Gestalt Psychology, Brace and Company, Inc., Harcourt, 1935.

[18] K. Köler, Gestalt Psychology, Liveright, 1947.

[19] G.W. Hartmann, Gestalt Psychology, The Ronald Press Company, 1935.

[20] J.R.P. Michael Kubovy, Perceptual Organization, Lawrence Erlbaum Associates, Inc., Publishers, 1981.

[21] A. Gobar, Philosophic Foundations of Genetic Psychology and Gestalt Psychology, Martinus Nijhoff, The Hague, The Netherlands, 1968.

[22] I.E. Gordon, Theories of Visual Perception, John Wiley & Sons, 1989.

[23] K. Lewin, Principles of Topological Psychology, McGraw-Hill Book Company, 1936.

[24] W.L. Idson, D.W. Massaro, Cross-octave masking of single tones and musical sequences: The effects of structure on auditory recognition, Perception and Psychophysics 19 (1976) 155–175.

[25] G.P. Singh, A.S. Bregman, The influence of different timbre attributes on the perceptual segregation of complex tone sequences, The Journal of the Acoustical Society of America 102 (4) (1997) 1943–1952.

[26] M. Jones, Time, our lost dimension: Toward a new theory of perception, attention, and memory, Psychological Review 83 (5) (1976) 323–355.

[27] R.E. Remez, P.E. Rubin, S.M. Berns, J.S. Pardo, J.M. Lang, On the perceptual organization of speech, Psychological Review 101 (1) (1994) 129–156.

[28] A.K. Engel, P. Fries, W. Singer, Dynamic predictions: Oscillations and synchrony in topdown processing, Nature Reviews: Neuroscience 2 (2001) 704–716.

[29] G.A. Calvert, Crossmodal processing in the human brain: Insights from functional neuroimaging studies, Cerebral Cortex 11 (2001) 1110–1123.

[30] S. Amari, A mathematical theory of self-organizing nerve systems, Biomathematics: Current Status and Future Perspectives, North-Holland, 1982, pp. 159–177.

[31] H.R. Wilson, J.D. Cowan, A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue, Kybernetik 13 (2) (1973) 55–80.

[32] V.K. Jirsa, H. Haken, Field theory of electromagnetic brain activity, Physical Review Letters 77 (1996) 960–963.

[33] V.K. Jirsa, H. Haken, Derivation of a field equation of brain activity, Journal of Biological Physics 22 (1996) 101–112.

[34] J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proceedings of the National Academy of Sciences — Biology 79 (8) (1982) 2554–2558.

[35] J.J. Hopfield, Neurons with graded response have collective computational properties like those of 2-state neurons, Proceedings of the National Academy of Sciences — Biology 81 (10) (1984) 3088–3092.

[36] G.A. Carpenter, S. Grossberg, A massively parallel architecture for a self-organizing neural pattern recognition machine, Computer Vision, Graphics, and Image Processing 37 (1987) 54–115.

[37] W.J. Freeman, Pattern recognition and associative memory as dynamical processes in a synergetic system i, ii, Biological Cybernetics 60 (476) (1988) 17–22. 107–109 (Erratum).

[38] M.W. Beauvois, Computer simulation of auditory stream segregation in alternating-tone sequences, The Journal of the Acoustical Society of America 99 (4) (1996) 2270–2280.

[39] S.L. McCabe, M.J. Denham, A model of auditory streaming, The Journal of the Acoustical Society of America 101 (3) (1997) 1611–1621.

[40] K.K. Govindarajan, S. Grossberg, L.L. Wyse, M.A. Cohen, A neural network model of auditory scene analysis and source segregation, Technical Report CAS/CNS-TR-94-039, Boston University, USA, 1994.

[41] S. Grossberg, Pitch-based streaming in auditory perception, Technical Report CAS/CNS-TR-96-007, Boston University, USA, 1996.

[42] S. Grossberg, Pitch-based streaming in auditory perception, in: Musical Networks: Parallel Distributed Perception and Performance, MIT Press, 1999, pp. 117–140.

[43] S. Grossberg, K.K. Govindarajan, L.L. Wysec, M.A. Cohen, The link between brain learning, attention, and consciousness, Neural Networks 17 (2004) 511–536.

[44] G.P. Singh, Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre? The Journal of the Acoustical Society of America 82 (3) (1987) 886–899.

[45] B. Tuller, F. Almonte, V.K. Jirsa, E. Large, A dynamic model of auditory stream segregation, in: 44th Annual Meeting of the Psychonomic Society, Vancouver, B.C. November 6–9, 2003, Psychonomic Society, 2003.

[46] S. Grossberg, The link between brain learning, attention, and consciousness, Consciousness and Cognition 8 (1999) 1–44.

[47] S.N. Wrigley, G.J. Brown, A computational model of auditory selective attention, IEEE Transactions on Neural Networks 15 (5) (2004) 1151–1163.

[48] D. Wang, Primitive auditory segregation based on oscillatory correlation, Cognitive Science 20 (1996) 409–456.

[49] S. Amari, Dynamics of pattern formation in lateral-inhibition type neural fields, Biological Cybernetics 27 (1977) 77–87.

[50] P. Nunez, The brain wave equation: A model for the *EEG*, Mathematical Biosciences 21 (1974) 279–297.

[51] P.A. Robinson, C.J. Rennie, J.J. Wright, Propagation and stability of waves of electrical activity in the cerebral cortex, Physical Review E 56 (1) (1997) 826–840.

[52] H.R. Wilson, J.D. Cowan, Excitatory and inhibitory interactions in localized populations of model neurons, Biophysical Journal 12 (1) (1972) 1–24.

[53] V.K. Jirsa, H. Haken, A derivation of a macroscopic field theory of the brain from the quasi-microscopic neural dynamics, Physica D 99 (1997) 503–526.

[54] Y. Tougas, A.S. Bregman, The crossing of auditory streams, Journal of Experimental Psychology: Human Perception and Performance 11 (1985) 788–798.

[55] S. Shimojo, C. Scheier, R. Nijhawan, L. Shams, Y. Kamitani, K. Watanabe, Beyond perceptual modality: Auditory effects on visual perception, Acoust. Sci. & Tech. 22 (2) (2001) 61–67.

[56] P.X. Joris, C.E. Schreiner, A. Rees, Neural processing of amplitude-modulated sounds, Physiological Reviews 84 (2004) 541–577.

[57] K.O. Bushara, J. Grafman, M. Hallett, Neural correlates of auditoryvisual stimulus onset asynchrony detection, The Journal of Neuroscience 21 (1) (2001) 300–304.

[58] A. Assmus, J.C. Marshall, A. Ritzl, J. Noth, K. Zilles, G.R. Fink, Left inferior parietal cortex integrates time and space during collision judgments, NeuroImage 20 (2003) S82–S88.

[59] G. Hickok, D. Poeppel, Towards a functional neuroanatomy of speech perception, Trends in Cognitive Sciences 4 (4) (2000) 131–138.

[60] M.A. Cohen, S. Grossberg, L.L. Wyse, A spectral network model of pitch perception, The Journal of the Acoustical Society of America 98 (2) (1995) 862–879.

[61] G.A. Miller, J.C.R. Licklider, Intelligibility of interrupted speech, The Journal of the Acoustical Society of America 22 (1950) 167–173.