Available online at www.sciencedirect.com
ScienceDirect
Journal of Physiology - Paris 100 (2006) 154–170
Journal of Physiology Paris
www.elsevier.com/locate/jphysparis

The role of predictive models in the formation of auditory streams

S.L. Denham [a,*], I. Winkler [b]

[a] Centre for Theoretical and Computational Neuroscience, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK
[b] Institute for Psychology, Hungarian Academy of Sciences, Budapest, Hungary

Abstract

Sounds provide us with useful information about our environment which complements that provided by other senses, but also poses specific processing problems. How does the auditory system distentangle sounds from different sound sources? And what is it that allows intermittent sound events from the same source to be associated with each other? Here we review findings from a wide range of studies using the auditory streaming paradigm in order to formulate a unified account of the processes underlying auditory perceptual organization. We present new computational modelling results which replicate responses in primary auditory cortex [Fishman, Y.I., Arezzo, J.C., Steinschneider, M., 2004. Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. J. Acoust. Soc. Am. 116, 1656–1670; Fishman, Y. I., Reser, D. H., Arezzo, J.C., Steinschneider, M., 2001. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear. Res. 151, 167–187] to tone sequences. We also present the results of a perceptual experiment which confirm the bi-stable nature of auditory streaming, and the proposal that the gradual build-up of streaming may be an artefact of averaging across many subjects [Pressnitzer, D., Hupé, J. M., 2006. Temporal dynamics of auditory and visual bi-stability reveal common principles of perceptual organization. Curr. Biol. 16(13), 1351–1357.]. Finally we argue that in order to account for all of the experimental findings, computational models of auditory stream segregation require four basic processing elements; segregation, predictive modelling, competition and adaptation, and that it is the formation of effective predictive models which allows the system to keep track of different sound sources in a complex auditory environment.

Keywords: Perceptual bi-stability; Auditory scene analysis; Auditory streaming; Attention; Predictive modelling

1. Introduction

There are two particular problems faced by the auditory system in processing acoustic signals; firstly, sounds are typically received as a mixture produced by several concurrently active sources, and secondly, they are generally intermittent and cannot be re-examined at will. It is therefore very important for the auditory system to keep track of sound sources in real time by building contextual representations which allow it to form associations between the individual sound events emanating from the same source (finding sound streams) and, on a higher level, between the sound sequences originating from different sources (finding global patterns, such as the music played by an orchestra) (Bregman, 1990). From this point of view it is clear that an important function of the auditory system is to evaluate how well incoming sounds fit within the existing sound streams, because the arrival of a sound that cannot be regarded as a probable continuation of any of the previously registered streams indicates either the presence of a new sound source or a change in the behaviour of an existing one. Such events carry new information and they are known to initiate the updating of the descriptions of active sources (Näätänen and Winkler, 1999; Winkler et al., 1996). The view proposed in this paper is that keeping track of sound sources and detecting new information are closely related functions of the auditory system and that they rely on largely common resources. Moreover, as we will argue, both of these functions require the

[*] Corresponding author. Tel.: +44 1752 232610; fax: +44 1752 233349.
E-mail addresses: sdenham@plymouth.ac.uk (S.L. Denham), iwinkler@cogpsyphy.hu (I. Winkler).

0928-4257/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jphysparis.2006.09.012

formation of predictive models, which can extrapolate from regularities extracted from the preceding auditory input. Predictive models allow the auditory system to rapidly provide possible solutions to the inverse problem (i.e., decomposing the mixture of sounds according to their likely source) as well as to determine whether the incoming signal carries information that could not be predicted from what is "known" about the current auditory environment (i.e., to spot new information).

Although some of these problems have been addressed in psychophysics over many years, e.g., van Noorden (1975), the neural mechanisms underlying all but the simplest of these phenomena are not yet well understood. The current paper combines evidence from a wide range of research, from recordings of multiunit activity through electro-magnetic field potential investigations to behavioural studies of perception, in order to formulate a unified account of the processes underlying auditory perceptual organization. Here we show how primitive sound clustering processes result in the segregation of activity relating to potentially different streams, on the basis of which various representations of auditory regularities or 'models' can be constructed and used to predict the behaviour of future auditory input. This is essentially an 'old-plus-new' strategy (Bregman, 1990), which supports the organization of incoming sounds and the detection of new information at the same time. We will argue that the auditory system produces and simultaneously maintains alternative versions of predictive models, which leads to competing perceptual organizations of the sound input. Alternative organizations vie for dominance, with the "winner" determining perception and non-dominant alternatives being suppressed. However, adaptation eventually weakens the suppression of non-dominant sound organizations, thereby allowing an alternative percept to emerge, if one exists. This results in bi- or even multi-stable perception (Leopold and Logothetis, 1999) (here we consider only the bi-stable case).

The remainder of the paper is organized as follows. Firstly we briefly review perceptual experiments and theoretical proposals regarding the formation of auditory streams. Next we discuss the significance of recent experiments highlighting the bi-stable nature of auditory streaming and the implications of these experiments for theoretical and computational models. We then go on to describe our proposals for a unified theoretical model of perceptual sound organization, providing evidence for each of the component processing stages in the model, and present new empirical data to support the suggestion (Pressnitzer and Hupé, 2006) that perceptual bi-stability can explain the apparent build-up of streaming in auditory streaming experiments (Anstis and Saida, 1985). We also discuss the controversial role of attention in auditory stream segregation. Finally we consider the predictions of this model and the notion that auditory stream segregation is not a simple pre-processing stage but rather part of the active and flexible exploration of the sensory environment (Neisser, 1967).

2. Auditory streaming

Although the auditory input usually contains a mixture of sounds emanating from several concurrently active sources, we are normally able to select and follow distinct streams of sound, such as a tune, someone speaking, the trill of a bird, or the sound of a car passing by. Bregman (1990) proposed the "auditory scene analysis" framework for describing the way in which the composite auditory input is parsed into coherent sound streams. The starting point of this theory is that there is no single method by which sounds originating from different sources can always be disentangled from each other. Therefore, the auditory system employs several sound analysis 'heuristics' in parallel, which reflect the characteristics of natural sounds (Bregman, 1990; Darwin and Carlyon, 1995; Moore and Gockel, 2002). For example, successive sounds emitted by the same source are usually similar to each other and their features change gradually in time. Many of these principles were originally described by Gestalt psychologists, who regarded them as physical rules for grouping elements of the sensory input; see, e.g., Köhler (1947). A clear distinction was made between simultaneous (e.g., common onset or harmonicity) and sequential (e.g., continuity) grouping cues (Bregman, 1990). However, in real-life auditory scenes, there is some interplay between the two (Darwin and Carlyon, 1995). In general, finding out what belongs together within the current auditory input almost always requires information about the previous behaviour of the active sound sources. For example, the notes of a scale played in succession by the same instrument are grouped together by similarity of the timbre of the instrument and by "good continuation" of the pitch, whereas they are separated e.g., from a concurrent stream of speech, which has its own internal regularities.

Bregman (1990) also proposed that the analysis processes of auditory stream segregation can be divided into primitive and schema-based processes. Whereas the former are based on innate capabilities and rely on principles that are valid for most sounds (van Noorden, 1975), the latter require the learning of rules, are influenced by previous experience, and may apply only to a subset of sounds (Bey and McAdams, 2002). For example, neonates can already segregate low and high sounds presented in rapid succession (McAdams and Bertoncini, 1997; Winkler et al., 2003a). In contrast, orchestra conductors can follow the tune carried by a single instrument within a large ensemble, a feat most of us would not be able to duplicate. Given the variety and complexity of natural auditory scenes, it is evident that the various heuristic analysis processes may come up with different solutions as to how the auditory input could be organized. Bregman's two-stage model suggests that at first, primitive sound organization processes automatically segregate the auditory input into notional streams. Then, in a subsequent stage, competition between alternative notional sound organizations is resolved by selecting a dominant organization, which then

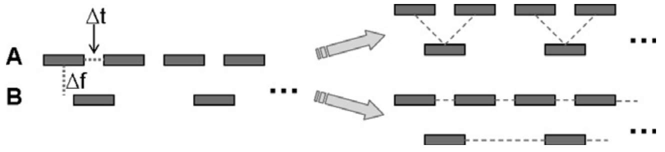

Fig. 1. Cartoon of the auditory streaming paradigm. A sequence of high (A) and low (B) tones presented repeatedly in ABA-groups can be perceived as a single coherent stream with a galloping rhythm (upper right), or as two segregated streams (lower right), each with an isochronous rhythm.

appears in perception. It has been further suggested that whereas competition between alternative organizations can be biased by attention, primitive processes, such as those governing auditory streaming (see below) are attention-independent. However, as will be discussed later, the role of attention in sound organization is not yet fully understood.

The general features of auditory stream segregation have been investigated most extensively via the primitive auditory streaming phenomenon. In the typical streaming paradigm (van Noorden, 1975), a tone sequence of the structure ABA–ABA–ABA– . . . is presented at a fast stimulus rate (A and B denote tones differing from each other in frequency; the "–" sign stands for a silent interval equal to the duration of the B tone; see Fig. 1). When all sounds are grouped together into a single coherent stream, a galloping rhythm is heard. By increasing the frequency separation (Δf) between the A and B tones and/or by shortening the interval between subsequent A tones (the within-stream inter-tone interval,[1] Δt) perception of the sound sequence changes to that of two homogeneous isochronous streams; one consisting of A tones and the other of B's (van Noorden, 1975). In general, there is a trade-off between Δf and Δt in determining the dominant perceptual organization. van Noorden (van Noorden, 1975) identified three separate regions of the $\Delta f - \Delta t$ space with different characteristic perceptual organizations (see Fig. 2). With very low Δf's and long Δt's, all tones are heard as part of a single sound stream and the galloping rhythm is perceived. With slightly larger Δf's and/or shorter Δt's, subjects are able to hear either two separate sound streams or a single integrated stream. Further increasing Δf and/or decreasing Δt results in the perception of two streams becoming the dominant sound organization.

Van Noorden (van Noorden, 1975) used a 'frequency sweep' method in order to map the perceptual dominance regions described above. An alternative way to investigate auditory streaming is to present sequences of fixed-frequency A and B tones and to require subjects to report their perception on a continuous basis. Using this approach we conducted an experiment to investigate bi-stability in auditory streaming. We found, similar to findings in vision (Hupé and Rubin, 2003), that the starting percept is generally that of a single integrated sequence characterized by the galloping rhythm, and that the two-stream percept gradually emerges, with the 'build-up' time related to Δf (and Δt) (Anstis and Saida, 1985), as illustrated in Fig. 3. However, the idea that the auditory system fixes on a single *unchanging* dominant percept is to some extent an artefact of the analysis method used which obscures an important detail about streaming, namely, that it fluctuates between the two possible percepts (Pressnitzer and Hupé, 2005, 2006); the significance of perceptual bi-stability in auditory streaming is discussed in the next section.

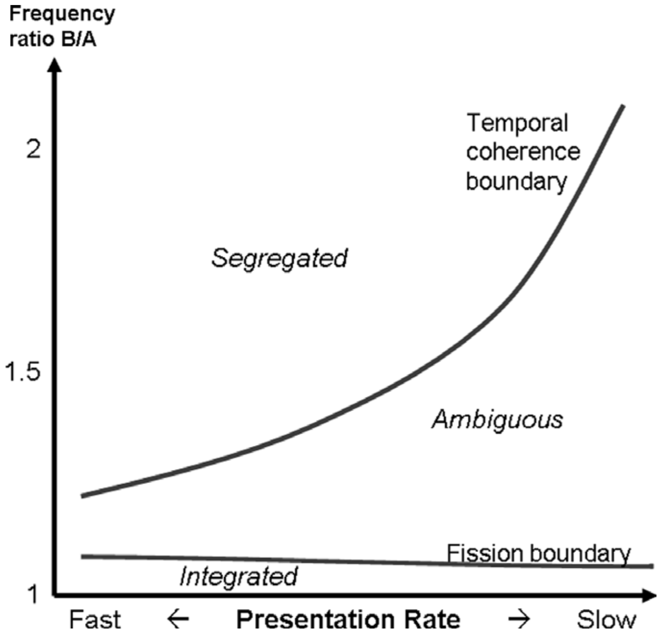

Fig. 2. The dependency of primitive auditory streaming on frequency difference and presentation rate found in human psychophysical experiments using alternating pure tones (Beauvois and Meddis, 1996; van Noorden, 1975). Stimuli in the region of parameter space above the 'temporal coherence boundary' are generally perceived as two segregated streams, and those with parameters in the region below the 'fission boundary' as a single coherent stream. Those falling in the ambiguous region can be perceived in either way, and can be influenced by attention (van Noorden, 1975).

3. Perceptual bi-stability

In vision, bi-stable perception has been studied extensively since it offers an ideal way for investigating correlates of conscious perception, as changes in perceptual awareness can be experienced in the absence of stimulus changes. The key factor is stimulus ambiguity; i.e., there should be more than one plausible alternative perceptual organization. This is often achieved through ambiguous depth cues which are easy to generate in 2D images; e.g., the Necker cube (Necker, 1832), or through binocular rivalry (Helm-

[1] Bregman and his colleagues (Bregman et al., 2000) found that the temporal parameter involved in establishing auditory streaming was the interval separating subsequent tones within a single stream (the A-to-A interval in an ABA-type of tone sequence). In ABA-type sequences, the within-stream inter-tone interval can be set by adjusting the tone durations and/or the interval between the onset of the A and the subsequent B tone (the stimulus onset asynchrony; SOA).

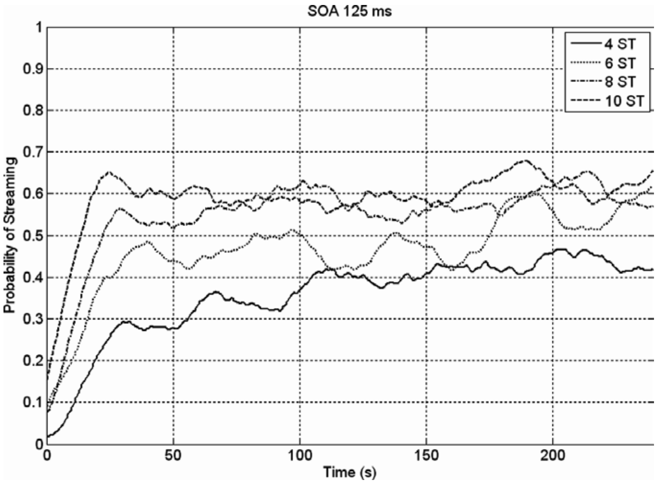

Fig. 3. Build-up of auditory stream segregation. Subjects were presented with 4-minute long trains of the ABA-structure, where A and B were pure tones of 75 ms duration. In separate trains Δ*f* was 4, 6, 8, or 10 semitones (ST) and SOA 75, 125, 175, or 225 ms ($4 \times 4 = 16$ different types of trains). The order of the trains was randomized separately for each subject. Subjects were instructed to keep a key depressed whenever they heard the galloping rhythm, and to release it when they did not. They were asked to mark in this way their perception throughout the trains and not to attempt hearing the sound according to one or another perceptual organization. The figure shows the probability of not hearing the galloping rhythm as a function of time for the 125-ms SOA, averaged across 23 subjects. Different Δ*f*s are shown according to the line styles indicated in the legend. The initial 2 seconds of the trains is not shown since this method does not measure stream segregation before subjects can initially report any pattern (e.g., ABA–). For compatible results, see Fig. 2 in Cusack (2005).

holtz, 1925), where each eye is presented with a different and incompatible image. In both cases clear switches in perception are experienced, even though the stimulus does not change.

A number of recent studies (Cusack, 2005; Gutschalk et al., 2005; Snyder et al., 2006; Winkler et al., 2005) have similarly exploited the bi-stability of auditory stream segregation in investigating the neural correlates of auditory perceptual organization. Bregman's notion that auditory streaming is not a single process is supported by two of the above event-related potential (ERP) studies. Winkler et al. (2005) presented a bi-stable tone sequence, the perception of which switched spontaneously between integrated and streaming organizations; i.e., the Δ*f* and Δ*t* parameters fell in the ambiguous region between the fission and temporal coherence boundaries. Subjects continuously indicated their perception by depressing or releasing a key. Averaging ERP responses, grouped according to the subject's perception, showed an early (50–90 ms) ERP component, which did not change with perception, but varied according to the differences in frequency between the alternating tones. A later (150–200 ms) ERP wave was only elicited when subjects actually perceived the sequence as a single integrated stream, but not when they heard two independent streams. Thus the earlier component appeared to reflect the outcome of an early segregation process, whereas the later one seemed to be related to the emerging dominant organization. Snyder et al. (2006) also found two different ERP effects related to streaming. One set of effects (N1c and P2) correlated with the perception of the sequence (increasing amplitudes with increasing Δ*f*s), but were insensitive to attention. Another effect (N2) correlated with the progress of the build-up of streaming in short ABA-trains and was significantly increased by attention; but was not directly affected by Δ*f*. On the basis of these results, the authors argued for the existence of a separate Δ*f*-based primitive clustering and an attentive build-up process.

By taking advantage of auditory bi-stability, neural activity in a region outside of the auditory cortex has also been found to correlate with the streaming percept (Cusack, 2005). fMRI results pinpointed a region in the intraparietal sulcus, which was differentially activated when subjects perceived two streams. This is interesting because it suggests a correspondence between auditory streams and perceptual 'objects' in the visual modality (Xu and Chun, 2006).

Perceptual bi-stability may best be understood as a means for optimising interpretations of the sensory environment. Clearly a veridical representation is desirable in order to guide appropriate behavioural decisions and ultimately enhance survival prospects. However, as empiricist theories suggest (Helmholtz, 1860/1962), sensory inputs may be inherently ambiguous, so it is important for the perceptual system to explore any plausible alternatives in order to minimise misinterpretations. Consequently it has been proposed that multi-stable perception is a result of the active exploration of the sensory environment (Leopold and Logothetis, 1999), and a fundamental aspect of sensory cognition which supports flexible decision making (Kim et al., 2006).

The key characteristics of visual bi-stability are: exclusivity, the existence of two plausible yet mutually exclusive alternative interpretations of the sensory input; randomness, stochastic switching between percepts such that successive dominance durations are uncorrelated; and inevitability, the finite duration of perceptual dominance; i.e., even when the intention is to hold onto one interpretation, a switch will always eventually occur (Leopold and Logothetis, 1999). Most models of visual bi-stability therefore comprise three essential ingredients (Kim et al., 2006). Exclusivity is achieved through competitive interactions between rival percepts, generally implemented as mutual inhibition or through the creation of network attractors. Stochasticity, often artificially injected, ensures that even though it is constrained by the attractor, the state of the network is never completely stationary. This promotes exploration and results in randomizing the time spent within any attractor state. Finally, adaptation ensures that attractors are only marginally stable, and eventually become unstable allowing the system to explore alternative perceptual states. In a recent elegant study exploring the intrinsic dynamics of this process, Kim et al showed that visual bi-stability exhibits stochastic resonance (at a period of roughly 600 ms) and that, in the biophysical models they analysed, this required adaptation to be stochastic (Kim et al., 2006).

The demonstration that auditory streaming exhibits many of the same bi-stable characteristics as vision (Pressnitzer and Hupé, 2005, 2006), including the properties of exclusivity (Winkler et al., 2006), randomness (see, e.g., the current empirical results discussed in Section 4.4) and inevitability, as well as the reduction of suppression durations rather than extension of dominance durations with increasing salience, suggests that 'auditory scene analysis' is not a simple pre-processing stage but an intrinsic part of the active and flexible perceptual exploration of the acoustic environment. This view is somewhat at odds with Bregman's proposal that following a default starting position of 'coherence', in which all sounds are considered to be part of the same stream, the auditory system gradually accumulates evidence in favour of the segregation of incoming sounds into separate streams (Bregman, 1990). Although auditory bi-stability experiments have found an initial bias towards coherence, this is followed by a situation in which perceptual organization switches randomly between coherence and streaming with no indication of a 'final decision' (Pressnitzer and Hupé, 2005), although depending on the stimulus parameters there is generally a bias towards one or the other organization.

These findings also pose problems for computational models of auditory stream segregation, e.g., Beauvois and Meddis (1996), Hartmann and Johnson (1991), McCabe and Denham (1997), and Wrigley and Brown (2004), none of which have been shown to exhibit bi-stability. In aiming to achieve perceptual exclusivity between 'streaming' and 'coherent' conditions and the gradual build-up of streaming, some computational auditory streaming models (McCabe and Denham, 1997; Wrigley and Brown, 2004) do so by making a single transition from an initial coherent state to the streaming state; although Wrigley's (Wrigley and Brown, 2004) model does include a reset mechanism that restarts the auditory stream segregation process when it is triggered. It may seem that one way to address this deficit would be to extend these models to include some form of adaptation (and stochasticity); however, as competition is essentially between regions representing different frequencies (or by extension different stimulus features (Moore and Gockel, 2002)) the coherent state is unstable, and this is in fact what drives the model responses to move from coherence to streaming in the first place. On the other hand models which are firmly based on an assumption of 'peripheral channelling' as a basis for stream segregation (Beauvois and Meddis, 1996; Hartmann and Johnson, 1991) have been discounted by more recent perceptual evidence for streaming on the basis of other stimulus features (Akeroyd et al., 2005; Moore and Gockel, 2002). Interestingly though, through its inclusion of a stochastic switching mechanism it is possible that the Beauvois and Meddis model could simulate perceptual bi-stability. However, in this model an attractor (corresponding to a dominant frequency channel) is formed through the suppression of all other channels, which seems to be inconsistent with experiments demonstrating *enhanced* responses to deviant stimuli (Näätänen et al., 2001; Ulanovsky et al., 2003). We therefore argue that both theoretical and computational models of auditory stream segregation require some rethinking in order to account for perceptual bi-stability.

4. Proposed unified model of auditory stream segregation

The unified model of auditory stream segregation we propose here builds upon Bregman's influential theory (Bregman, 1990), and takes into account a wide range of experimental evidence; including recordings of multiunit activity, electro-magnetic field potential investigations and behavioural studies of perception, and insights gained from models of visual bi-stable perception (Dayan, 1988; Laing and Chow, 2002; Wilson, 2003). It is also inspired by the view of sensory perception as a generative process of analysis-through-synthesis (Friston, 2005; Neisser, 1967). There are four key aspects to the model: (a) segregation; (b) predictive modelling; (c) competition; and (d) adaptation. Below we discuss the contribution of each of these to the composite perceptual function of auditory stream segregation.

4.1. Segregation

The first process we consider is one which achieves the context-dependent segregation of activity in primary auditory cortex (PAC) in response to successive sound events (Fishman et al., 2004, 2001; Micheyl et al., 2005). Given the ubiquitous assumption that PAC is organized tonotopically, this may seem a rather surprising process to include. However, it is possible that tonotopy in PAC may be a side-effect of organization with respect to some other feature(s) (Schonwiesner et al., 2002), and neurophysiological measurements have shown that even though sub-cortical response fields are generally narrowly tuned, subthreshold receptive fields of cells in PAC are more widely tuned, often in excess of five octaves (Kaur et al., 2004). Such broad tuning is also evident in the spiking responses in PAC.

In experiments in awake monkey, it was found that cells often responded initially to both tones in an alternating ABAB ... sequence, even if only one of them was at the best frequency (BF) of the cell (Fishman et al., 2004, 2001). The differential suppression of responses to the non-BF tone took some time to develop and depended on the context within which the tone was presented. Specifically, the emergence of non-BF tone suppression depended on the presence of an alternating BF tone, and on the presentation rate (Δt) and frequency difference (Δf) between the BF and the non-BF tones in a way which was consistent with the effects of these parameters on streaming perception in humans. These findings led Micheyl et al. (2005) to propose that some higher level process could make a streaming judgement simply on the basis of the relative frequency of neural responses to the two tones, and they showed (using an ABA_ABA_ sequence) that such a 'model' could account for the build-up of streaming typically reported in perceptual experiments. However, in the light of the bi-stability findings described in the previous section, we may question this interpretation, because there is no evidence in the published experimental data for a bimodal distribution of neural activity in PAC that would correspond to switches between the streaming and the integrated percepts (Fishman et al., 2004, 2001; Micheyl et al., 2005). Similarly, the early ERP component measured in the experiments of Winkler et al. (2005) did not correlate with the reported changes in perceptual organization. Thus, short-latency neural responses originating in PAC probably cannot fully explain even the simplest case of auditory streaming. However, it is also clear that some of the characteristics of streaming derive from the response properties observed in PAC.

Here we present some results from a new investigation in which we have found that a neurocomputational model of auditory processing, which includes synaptic depression such as that found in thalamocortical synapses (Thomson and Deuchars, 1994), and which was previously shown to explain many of the temporal response properties measured in PAC (Denham, 2001; Denham and Denham, 2001), can also account for the observed responses to streaming stimuli (Fishman et al., 2004, 2001). Below, it can be seen that the model (summarised in Appendix) exhibits a differential suppression of responses to non-BF tones (Fig. 4), which is similarly dependent upon the presentation rate and frequency difference between the BF and non-BF tones (Fig. 5). Furthermore, the model responses are primarily sensitive to the within-stream inter-tone interval (Fig. 6), as shown in perceptual experiments (Bregman et al., 2000), and later measured in PAC (Fishman et al., 2004).

The segregation of activity in PAC corresponding to different putative streams is consistent with the findings of ERP studies, described previously, of two essentially different ERP components; a sensory one, sensitive to Δf, and generated in PAC (Snyder et al., 2006; Winkler et al., 2005); and a higher-level one, generated in auditory association cortex and associated with more abstract regularities (Opitz et al., 2005). Compatible localization for streaming-related neural activity has also been obtained by fMRI (Deike et al., 2004).

In summary, we have found that the differential suppression of responses in PAC and its dependence on Δf and Δt can be explained by a model which includes synaptic depression such as that found in thalamocortical synapses. This model can also account for the differential responses to low probability sounds where the strength of the response reflects the recent probability of that sound's occurrence (Ulanovsky et al., 2003). Thus this early stage of processing results in the clustering and segregation of activity corresponding to putatively different sound streams. Although primitive clustering was originally thought to be simply a consequence of cochlear processing, there is a clear effect of presentation rate on stream segregation (see Fig. 2), which cannot be explained by peripheral

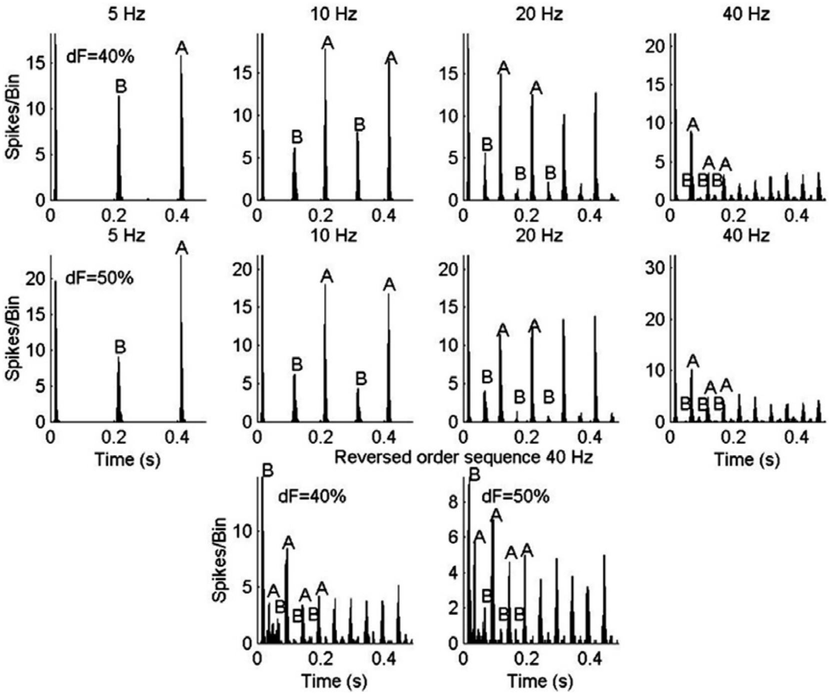

Fig. 4. Post stimulus time histograms of the neural array in region with BF = A for frequency differences of 40 and 50%. The response to B tones evident at 5 Hz decreases with increasing presentation rate until at 40 Hz the responses to B tones are virtually absent. In response to the reversed order sequences, the large response to the initial B tone is followed by the rapid suppression of further B tones responses and a relative increase in responses to A tones. PSTH bin size, 2 ms; A = 1000 Hz.

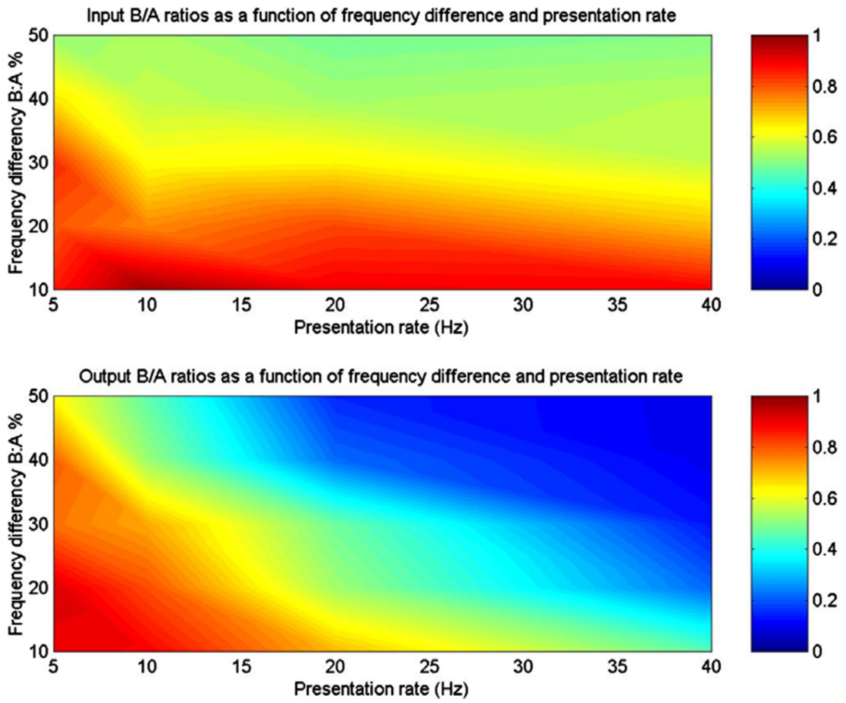

Fig. 5. Response ratios (B/A) as a function of presentation rate and frequency difference, calculated for the population with BF = A (non-BF = B). The colour scale indicates the degree of suppression present from low where B/A is close to one (red), to high, where B/A is close to zero (blue). Response ratios are calculated as the mean steady state response to B tones divided by the mean steady state response to A tones; for A = 1000 Hz, Δf = 10%, 20%, 30% and 40%, and presentation rate = 5, 10, 20, 30 and 40 Hz. Upper plot: response ratios for input stimuli, the result of peripheral processing. This shows sensitivity to frequency difference but not to presentation rate. Lower plot: 'cortical' response ratios are clearly sensitive both to frequency difference and presentation rate; i.e., there is increasing differential suppression with increasing frequency difference and presentation rate (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.).

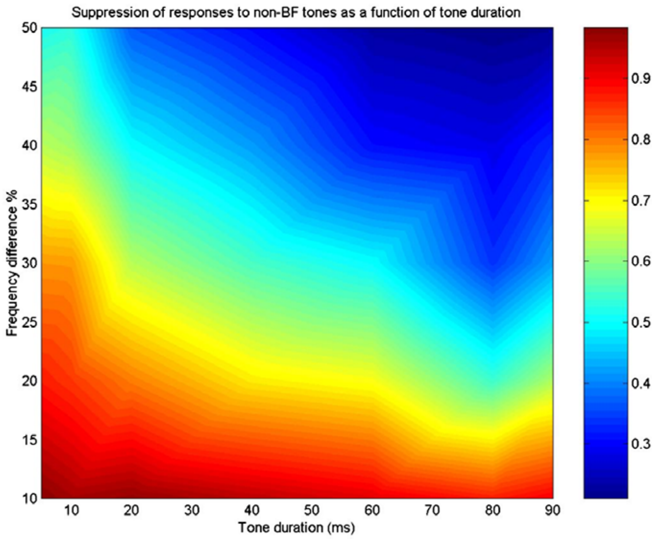

Fig. 6. The degree of suppression as a function of tone duration, or same-frequency offset to onset time which is (100 ms – tone duration). The increasing suppression with increasing duration and corresponding decreasing recovery time is clearly evident, particularly at larger frequency differences. Presentation rate = 10 Hz, A = 1000 Hz. Colour scale indicates B/A ratio, calculated as described in Fig. 5 for tone durations from 10 to 90 ms in 10 ms steps, and Δf's from 10% to 50% in 5% steps (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.).

processing, as well as experimental evidence in support of a PAC origin of at least some aspects of the primitive segregation process (Deike et al., 2004; Snyder et al., 2006; Winkler et al., 2005). While the segregation of activity in the pure tone streaming paradigm is straightforward, in the general case of complex sounds with overlapping and interleaved components, the early primitive clustering stage is likely to be crucial. The success of predictive modelling, which is important for keeping track of streams, depends upon the correct segregation of activity originating from different sound sources into different clusters. Mixing activity from independent, uncorrelated sound sources would unavoidably introduce prediction errors.

4.2. Predictive modelling

Prediction appears to be a fundamental aspect of sensory perception. A large number of experiments have shown that the auditory system uses an 'old-plus-new' strategy (Bregman, 1990). This means that searching for continuation of the existing streams within the new sound input is an important aspect of perceptual organization. Examples of this principle are the illusory auditory continuity experienced when gaps in an otherwise continuous tone are filled by noise sounds (e.g., Darwin, 2005) as well as several other perceptual restoration effects (e.g., Repp, 1992). Finding the continuation of a previously segregated sound stream within the composite auditory input requires the formation of representations describing the regularities detected for this stream, ones which can produce temporal extrapolations. These representations are thus predictive models of the given sound stream.

An important experimental method for investigating predictive models in sensory perception utilizes the mismatch negativity (MMN) ERP component, which is elicited by unpredicted violations of auditory regularities (Näätänen et al., 1978); for recent reviews, see (Näätänen and Winkler, 1999; Picton et al., 2000). The simplest MMN-yielding paradigm is the auditory oddball sequence, in which a repeating sound is occasionally exchanged for a different sound. However, violations of auditory regularities of much more complex nature also trigger the MMN, such as presenting a low tone of low intensity amongst tones complying with the "the higher the frequency the lower the intensity" rule (Paavilainen et al., 2001); for a review of the "intelligent" features of MMN, see (Näätänen et al., 2001; Winkler et al., 1996). Winkler and his colleagues (Winkler et al., 1996) hypothesized that MMN is elicited by differences between the incoming sound and the extrapolations drawn from the regularities extracted from the preceding sound sequence. That is, MMN is based on predictive models established for the given sound sequence (Winkler, 2003; Winkler et al., 1996). Several studies have suggested (Winkler and Czigler, 1998; Winkler et al., 1996, 2001, 2005) that the process reflected by MMN is involved in maintaining such predictive models, updating them when their predictions are not confirmed. A very important feature of MMN is that it is elicited whether or not subjects perform some task related to the sounds (Näätänen, 1990). In fact, the MMN-generating process is not affected by attention unless deviation in a feature occurs in an unattended stream when the same feature is task-relevant in the attended stream (Sussman et al., 2003). Therefore, with some precaution, MMN can be used to test the processing of unattended sounds.

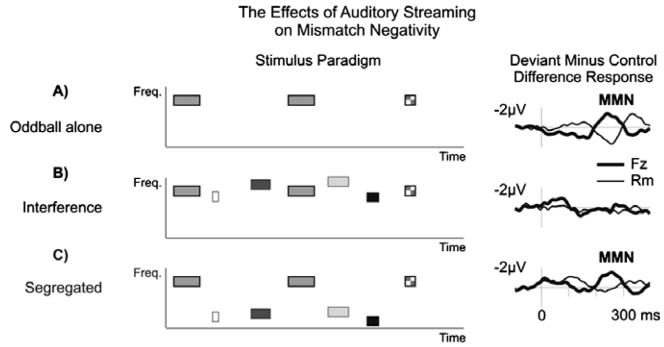

Fig. 7. Schematic illustration of the stimulus sequences (left side). Tones are denoted by rectangles, with the *y*-axis position showing the tone frequency, the width tone duration, and shading sound intensity. Panel (A) presents the "Oddball alone" condition. Infrequent short-duration (deviant) tones are marked with a chequered pattern. MMN was elicited by the deviant tones (right side). Panel (B) presents the "Interference" condition in which two tones were inserted between successive tones of the oddball sequence. Intervening tones had randomly varying duration, frequency (varying in a narrow range centred on the frequency in the oddball sequence), and intensity. No MMN was elicited by the original oddball deviant tones. Panel (C) presents the "Segregated" condition. The frequency range of the intervening tones was set far apart from the frequency in the oddball sequence. MMN was again elicited by the oddball deviant tones. Adapted from Winkler et al. (2003b).

A number of MMN studies have shown that sound organization and MMN elicitation are related because they both depend on what regularities can be detected from a given sequence of sounds (Ritter et al., 2000; Shinozaki et al., 2000; Sussman et al., 2001, 1998, 1999, 2005; Takegata et al., 2005; Winkler et al., 2001, 2003b, 2005, 2003c; Yabe et al., 2001). A simple example of the link between auditory streaming and MMN elicitation has been created using an oddball sequence (Winkler et al., 2003b) in which a repeating tone was occasionally exchanged for another tone with a different duration (Fig. 7A). In this sequence, duration deviants elicited the MMN. Then, two additional tones were introduced between consecutive sounds, whose duration varied over a wide range. In one condition, the frequency of these intervening tones varied in a range surrounding the frequency of the tones in the original oddball sequence (Fig. 7B). In this condition no MMN was elicited by the original duration-deviant tones, because the overall variation of duration in the sequence did not lead to the extraction of a duration rule and, therefore, the original deviants did not violate any regularity. However, when the frequency of the intervening tones varied in a range that was much lower than the common frequency of the original oddball sequence (Fig. 7C), MMN was elicited by the original duration-deviant tones. Because the interference and the streaming conditions did not differ from each other in terms of tone duration, the emergence of MMN in the streaming condition suggested that (1) the tones of the original oddball sequence and the intervening tones were streaming (indeed, the Δf and within-stream inter-tone intervals together fell into the range strongly promoting streaming – see (van Noorden, 1975)) and (2) separate predictive models were set up for the two sound streams, including the extraction of a tone-duration regularity for the original oddball sequence (which was then violated by the duration deviants embedded in this stream, thus eliciting the MMN). Using a version of the intervening-tone paradigm (with intensity instead of duration deviance), the link between MMN elicitation and streaming was confirmed in school-aged children (Sussman et al., 2001) and in adults (Sussman et al., in press, 2005) and the occurrence of streaming was suggested in newborn infants. Furthermore, Ritter and his colleagues (Ritter et al., 2000) provided corroborating evidence that the predictive models underlying MMN generation are stream-specific.

Here we argue that the relationship between MMN and sound organization is a causal one: The predictive models maintained by the MMN-generating process underlie the sequence-based processes of stream formation. That is, MMN is a part of the auditory scene analysis function; it is an indicator that something unexpected has been detected, and is essentially a correlate of prediction errors.

In conclusion, there is ample evidence that the brain forms representations of the regularities detected in continuous sound sequences, and uses such regularities in order to evaluate successive sound events. Unexpected events, which by definition contain information the system does not currently have, trigger processes, which update the regularity representations. In other words, the formation of predictive models is precisely what underlies the formation of auditory streams; i.e., one or more predictive models together define a 'stream'.

4.3. Competition

If the competitive interactions in current computational models of auditory streaming result in a lack of stability for the coherent state, then to paraphrase Logothetis et al. (1996), an important question to address is: 'What is rivalling in auditory stream segregation?' As previously noted, given the variety and complexity of natural auditory scenes, it is likely that there may be more than one plausible solution as to how the auditory input could be organized. According to Bregman (1990), competition between alternative sound organizations is used to resolve perceptual ambiguity, and to decide upon a single coherent interpretation.

The competition between alternative sound organizations must then involve competition between alternative predictive models. Interestingly, the only documented examples of auditory bi-stability are auditory stream segregation (Gutschalk et al., 2005; Pressnitzer and Hupé, 2005) and Warren's verbal transformations (Warren, 1961); both of which involve alternative temporal organizations. Previous ERP studies provide evidence that several alternative regularity representations may be formed to describe even simple sound sequences (Horváth et al., 2001). Regularity representations may describe "local" rules (i.e., relationship between adjacent sounds) as well as global rules (relationships between non-adjacent sounds).

If we consider the typical stimulus sequence used in streaming experiments (see Fig. 1), ABA_ABA_ABA_... then we can see that within this sequence there is perceptual 'evidence' for the following transitions: $A \rightarrow B(\Delta t)$, $B \rightarrow A(\Delta t)$, $A \rightarrow A(2\Delta t)$, $B \rightarrow B(4\Delta t)$. (There is also evidence for higher order transitions, but we ignore them here.) These then comprise the set of expectations or predictions derived from this sound sequence, which form a 'local' ($A \rightarrow B(\Delta t)$, $B \rightarrow A(\Delta t)$) and a 'global' ($A \rightarrow A(2\Delta t)$, $B \rightarrow B(4\Delta t)$) set of transition rules. Horváth et al. (2001) showed that incoming sounds are simultaneously tested against predictions derived from both sets of rules, thus demonstrating that these (and possibly further) predictive models are maintained at the same time in the auditory system. If we suppose that these representations are embodied in the activity of distinct neural populations in cortex, and that there is competition between mutually incompatible 'rules', then two mutually exclusive stable perceptual states emerge in response to the above stimulus sequence; namely those of coherence and streaming. This idea is illustrated in the cartoon in Fig. 8. Competition between mutually exclusive rules in this case amounts to competition between those neural populations, activated by the same current sensory input, which predict a different transition, and between those neural populations which predict the same stimulus, but are triggered by different input. The diagram shows that under this form of competition either the coherent or the streaming state can be stable, since the rules associated with each, separately, do not compete with each other. At the same time, the competitive interactions described above (and shown on the diagram) ensure that the two perceptual states (coherence and streaming) are mutually exclusive.

The finding that global regularities as well as local ones are detected and represented in the brain demonstrates that the length of the temporal windows within which regularities are detected varies quite substantially, as has been suggested by research aimed at finding higher-level auditory features (Nelken et al., 2003) and understanding speech perception (Poeppel, 2003). Processing sound on various time scales in parallel also has a bearing on the way in which perception develops through the stimulus sequence, and helps to explain why the initial percept is one of coherence. There are two reasons then why at the onset of the stimulus subjects generally experience the coherent percept. Firstly, at the onset of a sound sequence there are no predictive models available, and hence essentially no organization of the incoming sounds. Since there is nothing which can cause the sounds to segregate, the default state is coherence (this is essentially Bregman's argument). Sec-

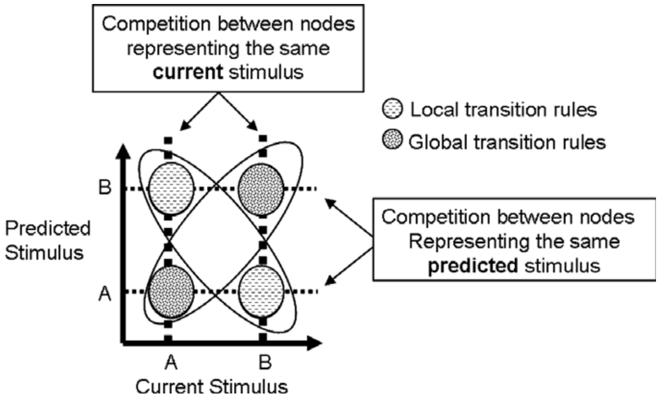
Competition between nodes representing the same **current** stimulus
Local transition rules
Global transition rules
Competition between nodes Representing the same **predicted** stimulus
Predicted Stimulus
B
A
A
B
Current Stimulus

Fig. 8. Cartoon illustrating the competitive interactions between populations encoding different transition rules. The x-axis shows the current inpu, and y-axis the stimulus predicted by local (dashed) and global (dotted) transition rules. Thus encountering current stimulus "A" leads to the prediction of stimulus "B" by local rules (top) and to stimulus "A" by global rules (bottom). At the same time, the stimulus "A" (lower part) is predicted by global rules from a preceding "A" (left) and by local rules from a preceding "B" (right). The two sets of rules (framed separately by elliptic curves) can lead both to a stable perceptual state of coherence (local rules dominant), or of streaming (global rules dominant), but not both at the same time.

ondly, the predictive models which form first are those which extract regularities over shorter time windows, namely the ‘local’ ones. The dominance of local predictive models, in this case also amounts to the perception of coherence. Then, after a while the more global models become established and begin to compete with the local ones. After some time, depending on the stimulus parameters, they may eventually come to win this competition and as a result the streaming percept emerges. However, perceptual organization is rather more flexible and less stable than this account would suggest; as will be discussed in the next section.

4.4. Adaptation

Although Fig. 3 appears to show a build-up of streaming with time consistent with many other experiments, e.g., Anstis and Saida (1985), Cusack (2005) and Micheyl et al. (2005); this is the result of averaging across the responses of many subjects at each point in time and as suggested by Pressnitzer and Hupé (2006) the apparent convergence to a stable perceptual state is really an artefact of this process. If we consider the responses of individual subjects then it is found that in general subjects continually switch back and forth between the two mutually exclusive perceptual states, i.e., coherence and streaming. A typical example of this behaviour is shown in Fig. 9. From this figure it is clear not only that perceptual organization never completely ‘stabilises’, but also that even for an individual subject the rate of perceptual switching can vary widely from one stimulus to the next. The number of perceptual switches for all subjects and experimental conditions is shown in Fig. 10. In Fig. 11, the mean durations of the first seven perceptual phases for all subjects (excluding those with fewer perceptual switches for these conditions), show the correspondence between the first phase duration and frequency difference which explains the differences in ‘build-up’ of streaming with Δf. In addition, these plots make clear the relationship between the final asymptotic probability of streaming and the relative durations of streaming and coherent phases.

This phenomenon of perceptual switching between two mutually exclusive organizations is very similar to that observed in visual experiments (Pressnitzer and Hupé, 2005, 2006) and might therefore be explained in a similar way. Almost all models of visual bi-stability; e.g., Laing and Chow (2002) and Wilson (2003), include adaptation in order to ensure that network attractors are only marginally stable, and that eventually a transition is always made to another perceptual state. In other words adaptation ensures that suppression of the non-dominant perceptual organizational state (i.e., the non-dominant set of predictive models) gradually weakens until eventually a new perceptual state emerges. In the case of auditory stream segregation this means that the percept switches between integrated and streaming percepts. If the conditions strongly favour streaming then this may happen less often and the switch back to streaming may be very rapid. However, in general there is no one fixed stable perceptual organization, and the cycle of dominance followed by

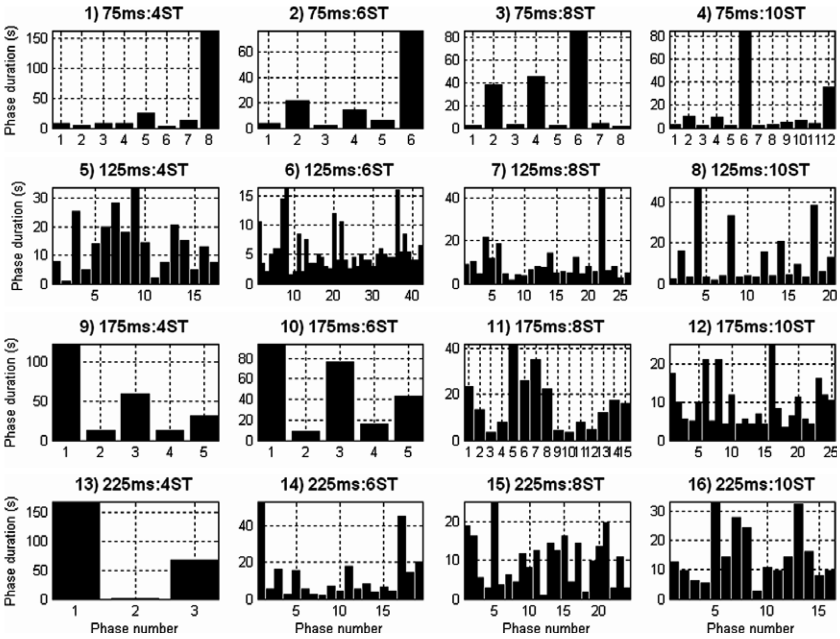

Fig. 9. Duration in seconds of successive perceptual states during each 4 minute stimulus for a typical subject. Each subplot indicates the perceptions reported by the same subject during each condition; indicated above each plot are the experimental condition number, Δt in ms and Δf in semitones. In these plots the first phase, and all subsequent odd numbered phases, correspond to coherence; i.e., the ‘galloping’ percept. All even numbered phases correspond to streaming. As can be seen there is huge variability both in the number of perceptual switches across conditions and in the duration of each perceptual state. Also evident is the lack of stability in perceptual organization.

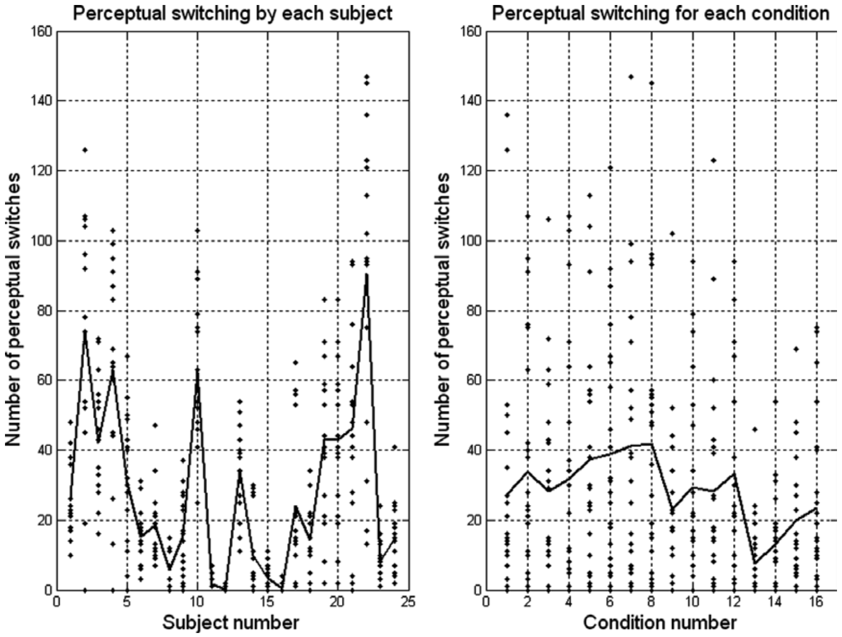

Fig. 10. Perceptual switching. Left plot shows the number of perceptual switches reported by subjects in each stimulus condition plotted against subject number. The mean for each subject is indicated by the solid line. Right plot shows the number of perceptual switches for all subjects plotted against stimulus condition, numbered as in Fig. 9. From this it is clear that even for conditions which are less ambiguous, i.e., the 4 semitone (1, 5, 9 and 13) and 10 semitone (4, 8, 12 and 16) conditions, there is still a marked tendency for perceptual bi-stability.

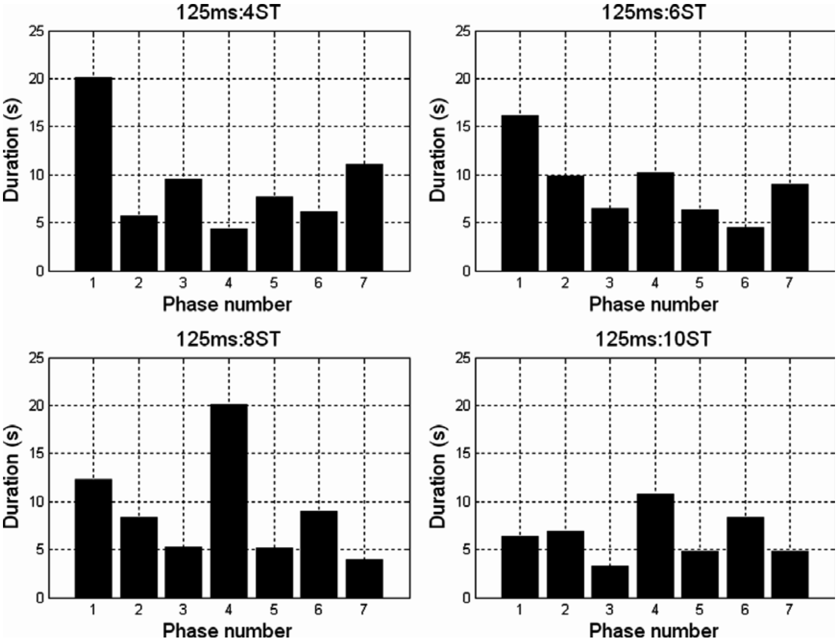

Fig. 11. Average durations of the first seven phases of perceptual organization for the 125 ms SOA conditions, also used for Fig. 3. As in Fig. 9, odd numbered phases correspond to coherence and even numbered phases correspond to streaming. From these plots it can be seen that there is a clear reduction in the first phase duration with increasing frequency difference, which would give rise to a faster 'build-up' of streaming if averaged across subjects. In addition, the relative duration of subsequent integrated and streaming phases tends to stabilise, and the proportion of time spent in integrated phases reduces with increasing frequency difference.

adaptation and perceptual switching then continues for the duration of the stimulus sequence. This is evident in all of the conditions we tested; e.g., consider the number of perceptual switches in the 10 semitone Δf conditions (numbers 4, 8, 12 and 16 in Fig. 10, right plot) for which a streaming percept is expected to be relatively stable.

It will be interesting to explore whether the recently reported finding of stochastic resonance in binocular rivalry (Kim et al., 2006) can also be demonstrated in auditory perception, since this is informative with respect to the dynamics of the system. The resonant period of approximately 600 ms they reported is significant in auditory per-

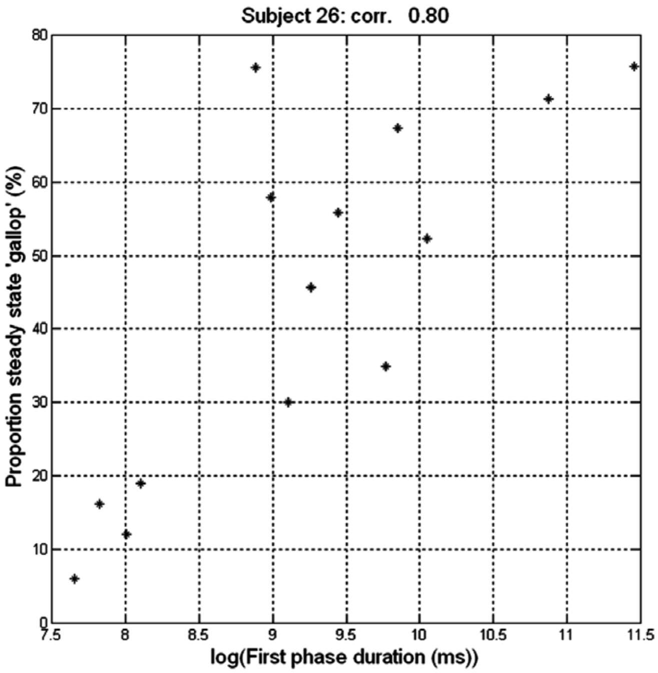

Fig. 12. Correlation between the log first percept duration and overall proportion of integration during the remainder of the stimulus sequence for one subject. This is very similar to results reported for visual plaid experiments (Hupé and Rubin, 2003).

ception, consistent with preferred musical tempi (Iwanaga and Tsukamoto, 1998) and mean consonant + vowel durations found in many languages (Ramus et al., 1999). Furthermore, it was shown that in order to generate stochastic resonance in the visual models investigated, it was necessary for adaptation to be stochastic (Kim et al., 2006). Therefore, this may provide useful guidance for models of auditory perceptual organization.

Interestingly, in visual plaid motion experiments it was found that the duration of the first coherent period, is strongly correlated with the final relative durations of coherence and dominance (Hupé and Rubin, 2003). This makes sense in terms of streaming too; a fast switch to streaming indicates a strong bias towards the streaming percept and hence rather brief subsequent periods of coherence (integration), while a slow initial switching to streaming indicates a weak bias towards streaming and hence far longer durations of coherence. Although we found a smaller correlation in our experiments over all subjects, some individual subjects showed a high correlation between first phase and final relative durations; e.g., see Fig. 12.

5. The role of attention in auditory stream segregation

Bregman (1990) suggested that in the two extreme parameter ranges (below the fission or above the temporal coherence boundary) perception is fully determined by stimulation factors. That is, in these parameter ranges sound organization would be attention-independent. In contrast, Jones (1976) and Jones et al. (1981) suggested that streaming represents the failure of switching attention with sufficient speed between sounds of highly different pitch. It is difficult to test the effects of attention on sound organization by behavioural methods, which usually require the subject to indicate his/her perception or to solve a task involving the test sounds. In support of the attention-independence of auditory streaming, Jones and his colleagues (Jones et al., 1999; Jones and Macken, 1995) found that performance in retaining a list of items in memory is less disturbed by task-irrelevant sound sequences presented during the retention interval, when the sequence was segregated by frequency into two (separately) repetitive streams of sound than when the sequence was organized as a single stream containing two or more sounds of different frequencies. Because subjects were instructed to ignore the irrelevant sounds, the authors interpreted these results as suggesting that auditory streaming occurs without attention. Furthermore, as was already mentioned, frequency-based streaming is probably an innate function of the human auditory system (McAdams and Bertoncini, 1997; Winkler et al., 2003a). However, stronger tests of attention dependence can be provided by physiological measures, which can be recorded even when subjects are instructed to ignore the sounds and to perform a primary task unrelated to the test sounds. In this regard the MMN ERP component has been found to be a useful tool.

Most MMN studies have used the so-called passive condition, in which subjects read a book or watch a movie and are instructed to ignore the experimental sounds, for testing the relationship between auditory streaming and MMN; e.g., Ritter et al. (2000), Sussman et al. (1999), Winkler et al. (2003c), Yabe et al. (2001). Although the results of these studies are compatible with the notion that auditory streaming can occur without attention, they did not provide a strong test of this hypothesis, because subjects could covertly and/or intermittently attend the sounds and it has been shown that MMN is correlated with sound organization also when this organization requires attention (Sussman et al., 2002; Winkler et al., 2003b). Winkler et al. (2003b) tested MMN elicitation using the streaming paradigm while subjects performed a visual n-back task. With two different difficulty levels of the n-back task (1- and 3-back) no MMN difference was found in the streaming condition, in which MMN elicitation could only occur if the intervening tones and the original oddball sequence were segregated into separate streams (see Section 4.2 and Fig. 7). On this basis, Winkler et al. concluded that maintaining separate streams does not require focused attention. Somewhat contradictory evidence has been obtained by studies in which subjects attended one sound stream, while ignoring two other sets of sounds. Whereas Winkler et al. (2003c) found that the unattended sounds were also segregated into separate streams, Sussman et al. (2005) found that only the attended stream was created, whereas the rest of the sounds formed an undiscriminated background. The latter finding is compatible with the results of a behavioural study using a similar design and stimuli (Brochard et al., 1999). There are two differences between Winkler et al.'s (2003c) paradigm and the other two studies, which may explain the contradictory results. Firstly, the sound stream attended by subjects in Winkler et al.'s paradigm was the sound of a movie, which they watched, whereas in the other two studies, subjects performed a difficult detection task in the attended tone sequence. Thus it is possible that Winkler et al.'s subjects intermittently attended the unattended sound sequences, segregating them this way. Secondly, the three sound streams presented in Winkler et al.'s study were qualitatively different (movie sound, street noise, series of footsteps), whereas in the other two studies, three (or more) streams of pure tones were delivered to subjects. Cusack et al. (2004) suggested that auditory stream segregation may be hierarchical with qualitatively different streams being possibly segregated without attention, whereas attention would be required to segregate from each other sounds with generally similar make-up. Alternatively, if attention acts by biasing the competition between competing perceptual organizations, e.g., Deco and Zihl (2001), it may be that this causes a distortion in the network attractor landscape which can cause some attractors to disappear. In summary, it appears that when no auditory stream is voluntarily selected, at least two, and possibly more sound streams may be maintained without attention. On the other hand, when attention is directed to one stream, multiple other streams may or may not be segregated.

Whereas continuously focussed attention does not have dramatic effects on auditory stream segregation, it has been shown that switching attention can have a marked effect on the 'build-up' of streams. Carlyon et al. (2001) presented short (21 s) trains of sounds having the ABA-structure to the left ear of subjects. The Δf frequency separation between the A and B tones varied from trial to trial between 4 and 10 semitones. In the base condition, subjects were instructed to press one button when they heard the galloping rhythm and another when they did not. By plotting the average score representing perception of one vs. two streams as a function of time within the trains, it was found that the perception of two streams gradually increased during the trains. The increase was sharper and the final perception more uniformly two streams with larger as opposed to smaller Δfs (see the current perceptual streaming results: Fig. 3). In the main experimental condition, during the first 10 s of each train subjects were instructed to perform a difficult discrimination task between two types of noise sounds presented to their right ear, which prevented them from attending the tones presented to their left ear. After the first 10 s, however, the noise sounds stopped and subjects were again asked to judge whether they heard one or two streams in their left ear. Again, plotting the scores representing perception as a function of time it was found that this function resembled the initial 10 s, rather then the second 10 s of the similar curves of the base condition. This was interpreted as suggesting that that no build-up of streaming occurs in an unattended sequence of sounds, at least when another sound sequence is attended. Cusack et al. (2004) extended these findings by showing that switching attention from a tone sequence for only 5 s to a concurrent noise sequence reset sound organization for the tone sequence. That is, even though two separate streams had already emerged within the tone sequence, after switching to a concurrent noise stream for 5 s subjects again heard a single integrated stream and it took a few seconds for two streams to emerge again in their perception.

In contrast, Sussman et al. (in press) recently found that the build-up of streaming can also occur for a sequence of tones when attention is strongly focused on a separate noise stream. These authors presented short (ca. 3.6 s) trains of the intervening-tone paradigm (two variable sounds inserted between successive tones of a simple oddball sequence; see Fig. 7B and C), separated by ca. 4 s of silence. The Δf was, in separate trains, either 1 or 8 semitones (termed "Near"– see Fig. 7B – and "Far" –see Fig. 7C – trains, respectively) and the absolute frequency was varied from train to train to prevent carryover between trains. An intensity deviant was placed either at the 4^{th} or the 10th position in the oddball sequence, thus testing the early and later phase of the build-up process. Subjects attended a continuous stream of noise delivered by a loudspeaker placed in front of them, while the tones of the experimental sequences were presented by two loudspeakers placed symmetrically on each side and somewhat behind them. Subjects performed a difficult detection task, with targets, slight changes of the noise intensity, appearing with random intervals (square distribution, 0.5–30 s, including the inter-train intervals). MMN was only elicited by 10th-position deviants in the Far trains. In a control oddball-only condition (see Fig. 7A), 4th-position deviants also elicited the MMN. These results suggested that streaming was built up by the end of the Far trains without attention being focused on the tones.

The picture emerging from the experiments reviewed above is that, whereas attention is not required for the build-up of streaming, *switching* attention generates some sort of reset. Taking into account the results of studies showing that streaming occurs when attention is not directed to the sounds, it seems likely that the resetting of streaming for a subset of the auditory input occurs when attention is directed towards that subset of sounds, but not when attention is directed away from them (for a similar conclusion, see (Cusack et al., 2004)). In the light of the arguments and perceptual results presented here, we suggest that the corresponding predictive models are reset by switches in attention, and that since the 'local' models reform first, a 'build-up' period once again occurs. The prediction is that a (possibly prolonged) coherent phase occurs for all subjects immediately after attentional switching, but that the early clustering is not affected.

6. Implications for models of auditory streaming

The role of expectations in resolving perceptual ambiguities underlies the 'generative' modelling approach which has been used to explain binocular rivalry (Dayan, 1988), and to formulate a general theory of cortical function (Friston, 2005). In this framework, each level in the sensory hierarchy imposes expectations on lower levels which help to constrain and guide their processing. In addition, lateral connections decorrelate responses at each level. Within such a hierarchy, incoming sensory signals generate activity which is passed to higher levels only to the extent that they are not predicted by prior expectations. In accordance with this framework, the formation of predictive models is an inherent aspect of processing at each level of the sensory hierarchy, and also explains why MMN generators may be localised to different parts of auditory cortex (Alho, 1995), depending on the particular feature (and the corresponding part of the sensory hierarchy) which causes the prediction error.

We suggest that it is just such a framework which could usefully form the basis for a comprehensive and unified model of auditory stream segregation. In summary, key processes within such a model include: (a) the segregation or clustering of activity corresponding to putatively different sound sources; (b) the generation of predictive models at all levels of the processing hierarchy through the extraction of regularities found within different clusters; (c) competition between mutually exclusive models, with attentional effects mediated through the biasing of this competition; (d) stochastic adaptation causing a weakening of the suppression of alternative models, and the eventual emergence of an alternative perceptual organization.

Acknowledgements

The authors thank Dr. János Horváth and Kinga Gyimesi for designing the stimulus paradigm and Kinga Gyimesi and Gabriella Pálfy for conducting the perceptual streaming experiment. This work is supported by European Research Area Specific Targeted Project EmCAP (IST-FP6-013123).

Appendix

The dynamical behaviour of the synaptic model is determined by a system of three coupled differential equations:

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= z(t) - \alpha \cdot x(t)\\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \beta \cdot w(t) - z(t)\\ \frac{\mathrm{d}w}{\mathrm{d}t} &= \alpha \cdot x(t) - \beta \cdot w(t)\end{aligned} \tag{1}$$

where $x(t)$ is the amount of *effective* resource, and could be interpreted as the activated neurotransmitter within the synaptic cleft; $y(t)$ is the amount of *available* resource or free neurotransmitter in the synapse, and $w(t)$ is the amount of *inactive* resource, neurotransmitter being reprocessed. In the model all of these are considered as a proportion of the total synaptic resource, and hence always sum to 1. The constant β determines the rate at which the inactive resource $w(t)$ is returned to the pool of available resource on a continuing basis, and α represents the rate at which effective resource becomes inactive again subsequent to being activated.

Synaptic transmission is a stochastic process, postsynaptic EPSPs vary in amplitude; there is an increasingly high probability of failure in transmission at depressing synapses with increasing stimulus duration and the probability of failure is inversely related to the failure of the previous pre-synaptic spike to elicit an EPSP (Galarreta and Hestrin, 1998). To account for these aspects the input to the synaptic model, $z(t)$, is defined as follows:

$$z(t) = I(t) \cdot f[g, y(t)] \tag{2}$$

where $I(t)$ represents the occurrence of a pre-synaptic action potential and is set equal to one at the time of arrival of the pre-synaptic action potential and otherwise is set equal to 0. In this model both the probability of successful transmission and the amount of transmitter actually released is a probabilistic function, $f[g, y(t)]$, of the transmitter available for release $y(t)$, and the instantaneous efficacy of the synapse, g, which takes a value in the range zero to one.

$$\begin{aligned}f[g, y(t)] &= (p_{\text{event}} > r) \cdot y(t) \cdot r_{\text{n}}\\ p_{\text{event}} &= (1-g)^{y(t)}\end{aligned} \tag{3}$$

where p_{event} is the probability of a successful transmission and is a function of the available transmitter, $y(t)$, and the efficacy of the synapse, g; r is a uniform random variable in the range 0–0.25, and r_{n} is a normal random variable, with zero mean and standard deviation 0.25.

The EPSP at the synapse, $e(t)$, is computed from $x(t)$ in (1) using the following equation for the passive membrane mechanism (Tsodyks and Markram, 1997):

$$\tau_{\text{EPSP}} \cdot \frac{\mathrm{d}e}{\mathrm{d}t} = \gamma \cdot x(t) - e(t) \tag{4}$$

The neurone model used is described by the following system of equations, which has been adapted from a model in McGregor (1989):

$$\begin{aligned}&\tau_{\text{E}} \frac{\mathrm{d}E}{\mathrm{d}t} = -E(t) + V(t) + G_{\text{K}}(t) \cdot [E_{\text{K}} - E(t)]\\ &s(t) = 1 \quad \text{if } E(t) \geqslant \theta(t) \quad \text{else } s(t) = 0\\ &\tau_{G_{\text{K}}} \frac{\mathrm{d}G_{\text{K}}}{\mathrm{d}t} = -G_{\text{K}}(t) + \eta \cdot s(t)\\ &\tau_{\theta} \frac{\mathrm{d}\theta}{\mathrm{d}t} = -(\theta(t) - \theta_0) + s(t)\end{aligned} \tag{5}$$

where $E(t)$ is the variation of the neurone's membrane potential relative to its resting potential, $V(t)$ is the driving input found by summing all the synaptic EPSPs, $G_K(t)$ is the potassium conductance, divided by the sum of all the voltage-dependent ionic membrane conductances, E_K is the potassium equilibrium potential of the membrane relative to the membrane resting potential, $\theta(t)$ is the firing threshold potential, θ_0 is the resting threshold, $s(t)$ is the variable which denotes firing of the cell, τ_E, τ_{EPSP}, τ_θ, and τ_{GK} are time constants, and γ, χ and η are constant parameters.

In this system of equations, $s(t)$ is set to 1 to signal the occurrence of an action potential, i.e., $E(t)$ reaching a value above the firing threshold $\theta(t)$; otherwise $s(t)$ is zero. Eq. (5, line 4) is introduced to provide a refractory period. It allows representation of an absolute period and a relative period. For the first few milliseconds after firing the value of $\theta(t)$ becomes very large, preventing any further firing. As $\theta(t)$ decays between spikes, the threshold for firing decreases with time elapsed since the last spike. A further spike can occur therefore in this period if the value of $E(t)$ is sufficiently large. When $s(t)$ is zero, the potassium conductance term $G_K(t)$ decays to zero via Eq. (5). When $s(t) = 1$, the value of G_K is increased instantaneously by an amount η, and then decays again. The action potentials generated when the cell fires are not explicitly modelled, and the spiking variable $s(t)$ is used as the output from the model. In the simulations the following values were used for the constants in the model: $\alpha = 125$, $\beta = 8$, $g = 0.7$, $\gamma = 6$, $\tau_{EPSP} = .007$, $\tau_E = .005$, $\tau_{GK} = .01$, $\tau_\theta = .001$, $E_K = -10$, $\theta_0 = 10$, $\eta = 100$.

In order to compare the behaviour of the model with the experiments described it is desirable to use actual sound stimuli. For this reason the Development System for Auditory Modelling,[2] was used to generate signals characteristically found in auditory nerve fibre recordings in response to acoustic stimuli. The output from the peripheral model was reprocessed to ensure that the firing rates remained below about 100 Hz by enforcing a reasonable refractory period. This was achieved by assigning 30 auditory nerve fibres to each frequency channel, and only transmitting spikes when typically more than ca. 8 spikes were coincident within a 1 ms bin. Clearly this ignores the computations which occur in the rest of the sub-cortical auditory system. However, there does appear to be a fast veridical auditory pathway that transmits signals from the periphery to the cortex with relatively little alteration. While recognising that this simplification may result in a poor approximation of actual thalamic relay cell activity, it has the benefit of making the simulations tractable.

In their experiments (Fishman et al., 2004, 2001) obtained population responses from auditory cortex of awake monkeys to alternating pure tone stimuli. Multiunit activity (MUA) was measured using multi-contact recording electrodes placed in the thalamocortical recipient layers of PAC and provided evidence of net changes in summed action potential activity in neural ensembles of about 50-100 μm in diameter. In addition PSTH responses were constructed from a cluster analysis of the outputs of the recording electrodes. Stimuli were generated on the basis of the best frequency (BF) at the site of a recording electrode. For each site, the frequency of the A tones was set to the BF and the frequency of the B tones was set in the range of 10–50% either side of A. Stimuli were presented at rates of 5, 10, 20 and 40 Hz.

To simulate the response of the model in a similar experimental paradigm, the A tones were set to 1000 Hz and B to $A + \Delta f * A/100$. Inputs were generated using the peripheral processing described above. For this experiment 100 bandpass filter channels with centre frequency ranging from 500 to 2500 Hz on the ERB scale (Glasberg and Moore, 1990), were used. This resulted in 100, tonotopically organized spike trains which were used as input to the neural array, which consisted of 100 neurones, each with 30 synapses. Synaptic connectivity was generated by assigning to each neurone a nominal best frequency, and randomly selecting input connections from a Gaussian distribution centred on this frequency. The spread of the distribution was chosen so that the response to both A and B tones at the first presentation was similar at the region with BF = A, as found in PAC (Fishman et al., 2001). For comparison with multiunit activity the output spike train from each neurone was integrated with time constant $\tau_A = 5$ ms, and summed to give a population response (MUA_A) in the region with BF = A, where region A was defined as those neurones with BF's in the range A ± 10%; similarly for B.

$$\begin{aligned} \tau_A \frac{dA}{dt} &= s(t) - A(t) \\ MUA_A &= \sum_{i \varepsilon \, \mathrm{region}(A)} A_i(t) \end{aligned} \tag{6}$$

References

Akeroyd, M.A., Carlyon, R.P., Deeks, J.M., 2005. Can dichotic pitches form two streams? J. Acoust. Soc. Am. 118, 977–981.

Alho, K., 1995. Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. Ear Hear. 16, 38–51.

Anstis, S., Saida, S., 1985. Adaptation to auditory streaming of frequency-modulated tones. J. Exp. Psychol.: Hum. Percept. Perform. 11, 257–271.

Beauvois, M.W., Meddis, R., 1996. Computer simulation of auditory stream segregation in alternating-tone sequences. J. Acoust. Soc. Am. 99, 2270–2280.

Bey, C., McAdams, S., 2002. Schema-based processing in auditory scene analysis. Percept. Psychophys. 64, 844–854.

Bregman, A.S., 1990. Auditory Scene Analysis.

Bregman, A.S., Ahad, P.A., Crum, P.A., O'Reilly, J., 2000. Effects of time intervals and tone durations on auditory stream segregation. Percept. Psychophys. 62, 626–636.

Brochard, R., Drake, C., Botte, M.C., McAdams, S., 1999. Perceptual organization of complex auditory sequences: effect of number of simultaneous subsequences and frequency separation. J. Exp. Psychol. Hum. Percept. Perform. 25, 1742–1759.

[2] DSAM: Development Software for Auditory Modelling, a library of compiled C routines for auditory modelling. This software is publically available from http://www.essex.ac.uk/psychology/hearinglab/.

Carlyon, R.P., Cusack, R., Foxton, J.M., Robertson, I.H., 2001. Effects of attention and unilateral neglect on auditory stream segregation. J. Exp. Psychol. Hum. Percept. Perform. 27, 115–127.

Cusack, R., 2005. The intraparietal sulcus and perceptual organization. J. Cogn. Neurosci. 17, 641–651.

Cusack, R., Deeks, J., Aikman, G., Carlyon, R.P., 2004. Effects of location, frequency region, and time course of selective attention on auditory scene analysis. J. Exp. Psychol. Hum. Percept. Perform. 30, 643–656.

Darwin, C.J., Carlyon, R.P., 1995. Auditory grouping. In: Moore, B.C.J. (Ed.), Handbook of Perception and Cognition, Hearing, vol. 6. Academic Press, Orlando, FL, pp. 387–424.

Darwin, C.J., 2005. Simultaneous grouping and auditory continuity. Percept. Psychophys. 67, 1384–1390.

Dayan, P., 1988. A hierarchical model of visual rivalry. Neural Comput. 10, 1119–1136.

Deco, G., Zihl, J., 2001. Top-down selective visual attention: a neurodynamical approach. Visual Cogn. 8, 119–140.

Deike, S., Gaschler-Markefski, B., Brechmann, A., Scheich, H., 2004. Auditory stream segregation relying on timbre involves left auditory cortex. Neuroreport 15, 1511–1514.

Denham, S., 2001. Cortical synaptic depression and auditory perception. In: Greenberg, S., Slaney, M. (Eds.), Computational Models of Auditory Function. IOS Press, Amsterdam.

Denham, S., Denham, M., 2001. An investigation into the role of cortical synaptic depression in auditory processing. In: Wermter, S., Austin, J., Willshaw, D. (Eds.), Lecture Notes in Artificial Intelligence. Springer, pp. 494–506.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M., 2001. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear. Res. 151, 167–187.

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

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 815–836.

Galarreta, M., Hestrin, S., 1998. Frequency-dependent synaptic depression and the balance of excitation and inhibition in the neocortex. Nat. Neurosci. 1, 587–594.

Glasberg, B.R., Moore, B.C., 1990. Derivation of auditory filter shapes from notched-noise data. Hear. Res. 47, 103–138.

Gutschalk, A., Micheyl, C., Melcher, J.R., Rupp, A., Scherg, M., Oxenham, A.J., 2005. Neuromagnetic correlates of streaming in human auditory cortex. J. Neurosci. 25, 5382–5388.

Hartmann, W.M., Johnson, D.H., 1991. Stream segregation and peripheral channelling. Music Percept. 9, 155–184.

Helmholtz, H., 1860/1962. Handbuch der Physiologischen Optik (English translation). Dover, New York.

Helmholtz, H., 1925. Treatise on Physiological Optics. Columbia University Press for the Optical Society of America.

Horváth, J., Czigler, I., Sussman, E., Winkler, I., 2001. Simultaneously active pre-attentive representations of local and global rules for sound sequences in the human brain. Brain Res. Cogn. Brain Res. 12, 131–144.

Hupé, J.M., Rubin, N., 2003. The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. Vision Res. 43, 531–548.

Iwanaga, M., Tsukamoto, M., 1998. Preference for musical tempo involving systematic variations of presented tempi for known and unknown musical excerpts. Percept. Mot. Skills 86, 31–41.

Jones, M.R., 1976. Time, our lost dimension: toward a new theory of perception, attention, and memory. Psychol. Rev. 83, 323–355.

Jones, D.M., Macken, W.J., 1995. Organizational factors in the effect of irrelevant speech: the role of spatial location and timing. Mem. Cogn. 23, 192–200.

Jones, M.R., Kidd, G., Wetzel, R., 1981. Evidence for rhythmic attention. J. Exp. Psychol. Hum. Percept. Perform. 7, 1059–1073.

Jones, D., Alford, D., Bridges, A., 1999. Organizational factors in selective attention: the interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. J. Exp. Psychol.: Learn., Mem. Cogn. 25, 464–473.

Kaur, S., Lazar, R., Metherate, R., 2004. Intracortical pathways determine breadth of subthreshold frequency receptive fields in primary auditory cortex. J. Neurophysiol. 91, 2551–2567.

Kim, Y.J., Grabowecky, M., Suzuki, S., 2006. Stochastic resonance in binocular rivalry. Vision Res. 46, 392–406.

Köhler, W., 1947. Gestalt Psychology. Liveright, New York.

Laing, C.R., Chow, C.C., 2002. A spiking neuron model for binocular rivalry. J. Comput. Neurosci. 12, 39–53.

Leopold, D.A., Logothetis, N.K., 1999. Multistable phenomena: changing views in perception. Trends Cogn. Sci. 3, 254–264.

Logothetis, N.K., Leopold, D.A., Sheinberg, D.L., 1996. What is rivalling during binocular rivalry? Nature 380, 621–624.

McAdams, S., Bertoncini, J., 1997. Organization and discrimination of repeating sound sequences by newborn infants. J. Acoust. Soc. Am. 102, 2945–2953.

McCabe, S.L., Denham, M., 1997. A model of auditory streaming. J. Acoust. Soc. Am. 101, 1611–1621.

McGregor, R.J., 1989. Neural and Brain Modelling. Academic Press.

Micheyl, C., Tian, B., Carlyon, R.P., Rauschecker, J.P., 2005. Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48, 139–148.

Moore, B.C., Gockel, H., 2002. Factors influencing sequential stream segregation. ACTA Acust. United Acust., 88.

Näätänen, R., 1990. The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. Behav. Brain Sci. 13, 201–288.

Näätänen, R., Winkler, I., 1999. The concept of auditory stimulus representation in cognitive neuroscience. Psychol. Bull. 125, 826–859.

Näätänen, R., Gaillard, A.W., Mantysalo, S., 1978. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol. (Amst.) 42, 313–329.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., Winkler, I., 2001. "Primitive intelligence" in the auditory cortex. Trends Neurosci. 24, 283–288.

Necker, L.A., 1832. Observations on some remarkable optical phenomena seen in Switzerland; and on an optical phenomenon which occurs on viewing a figure of a crystal or geometrical solid. Lond. Edinburgh Philos. Mag. J. Sci. 1, 329–337.

Neisser, U., 1967. Cognitive Psychology. Appleton-Century-Crofts, New York.

Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., Farkas, D., 2003. Primary auditory cortex of cats: feature detection or something else? Biol. Cybernet.

Opitz, B., Schroger, E., von Cramon, D.Y., 2005. Sensory and cognitive mechanisms for preattentive change detection in auditory cortex. Eur J. Neurosci. 21, 531–535.

Paavilainen, P., Simola, J., Jaramillo, M., Naatanen, R., Winkler, I., 2001. Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). Psychophysiology 38, 359–365.

Picton, T.W., Alain, C., Otten, L., Ritter, W., Achim, A., 2000. Mismatch negativity: different water in the same river. Audiol. Neurootol. 5, 111–139.

Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time. Speech Commun. 41, 245–255.

Pressnitzer, D., Hupé, J.M., 2005. Is auditory streaming a bistable percept? Paper Presented at: Forum Acusticum, Budapest.

Pressnitzer, D., Hupé, J.M., 2006. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. Curr. Biol. 16 (13), 1351–1357.

Ramus, F., Nespor, M., Mehler, J., 1999. Correlates of linguistic rhythm in the speech signal. Cognition 73, 265–292.

Repp, B.H., 1992. Perceptual restoration of a “missing speech sound: auditory induction or illusion? Percept. Psychophys. 51, 14–32.

Ritter, W., Sussman, E., Molholm, S., 2000. Evidence that the mismatch negativity system works on the basis of objects. Neuroreport 11, 61–63.

Schonwiesner, M., von Cramon, D.Y., Rubsamen, R., 2002. Is it tonotopy after all? Neuroimage 17, 1144–1161.

Shinozaki, N., Yabe, H., Sato, Y., Sutoh, T., Hiruma, T., Nashida, T., Kaneko, S., 2000. Mismatch negativity (MMN) reveals sound grouping in the human brain. Neuroreport 11, 1597–1601.

Snyder, J.S., Alain, C., Picton, T.W., 2006. Effects of attention on neuroelectric correlates of auditory stream segregation. J. Cogn. Neurosci. 18, 1–13.

Sussman, E., Ritter, W., Vaughan Jr., H.G., 1998. Attention affects the organization of auditory input associated with the mismatch negativity system. Brain Res. 789, 130–138.

Sussman, E., Ritter, W., Vaughan Jr., H.G., 1999. An investigation of the auditory streaming effect using event-related brain potentials. Psychophysiology 36, 22–34.

Sussman, E., Ceponiene, R., Shestakova, A., Naatanen, R., Winkler, I., 2001. Auditory stream segregation processes operate similarly in school-aged children and adults. Hear. Res. 153, 108–114.

Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., Näätänen, R., 2002. Top-down effects can modify the initially stimulus-driven auditory organization. Brain Res. Cogn. Brain Res. 13, 393–405.

Sussman, E., Winkler, I., Wang, W., 2003. MMN and attention: competition for deviance detection. Psychophysiology 40, 430–435.

Sussman, E.S., Bregman, A.S., Wang, W.J., Khan, F.J., 2005. Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. Cogn. Affect. Behav. Neurosci. 5, 93–110.

Sussman, E., Horvath, J., Winkler, I., Orr, M., in press. The role of attention in the formation of auditory streams. Percept. Psychophys.

Takegata, R., Roggia, S.M., Winkler, I., 2005. Effects of temporal grouping on the memory representation of inter-tone relationships. Biol. Psychol. 68, 41–60.

Thomson, A.M., Deuchars, J., 1994. Temporal and spatial properties of local circuits in neocortex. Trends Neurosci. 17, 119–126.

Tsodyks, M.V., Markram, H., 1997. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. Proc. Natl. Acad. Sci. USA 94, 719–723.

Ulanovsky, N., Las, L., Nelken, I., 2003. Processing of low-probability sounds by cortical neurons. Nat. Neurosci. 6, 391–398.

van Noorden, L.P.A.S., 1975. Temporal coherence in the perception of tone sequences. Ph.D. Thesis, Eindhoven.

Warren, R.M., 1961. Illusory changes in repeated words: differences between young adults and the aged. Am. J. Psychol. 74, 506–516.

Wilson, H.R., 2003. Computational evidence for a rivalry hierarchy in vision. Proc. Natl. Acad. Sci. USA 100, 14499–14503.

Winkler, I., 2003. Change detection in complex auditory environment: beyond the oddball paradigm. In: Polich, J. (Ed.), Detection of Change: Event-related Potential and fMRI Findings. Kluwer Academic Publishers, Boston, pp. 61–81.

Winkler, I., Czigler, I., 1998. Mismatch negativity: deviance detection or the maintenance of the ‘standard’. Neuroreport 9, 3809–3813.

Winkler, I., Karmos, G., Näätänen, R., 1996. Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. Brain Res. 742, 239–252.

Winkler, I., Schroger, E., Cowan, N., 2001. The role of large-scale memory organization in the mismatch negativity event-related brain potential. J. Cogn. Neurosci. 13, 59–71.

Winkler, I., Kushnerenko, E., Horvath, J., Ceponiene, R., Fellman, V., Huotilainen, M., Näätänen, R., Sussman, E., 2003a. Newborn infants can organize the auditory world. Proc. Natl. Acad. Sci. USA 100, 1182–1185.

Winkler, I., Sussman, E., Tervaniemi, M., Horvath, J., Ritter, W., Naatanen, R., 2003b. Preattentive auditory context effects. Cogn. Affect. Behav. Neurosci. 3, 57–77.

Winkler, I., Teder-Salejarvi, W.A., Horvath, J., Naatanen, R., Sussman, E., 2003c. Human auditory cortex tracks task-irrelevant sound sources. Neuroreport 14, 2053–2056.

Winkler, I., Takegata, R., Sussman, E., 2005. Event-related brain potentials reveal multiple stages in the perceptual organization of sound. Brain Res. Cogn. Brain Res. 25, 291–299.

Winkler, I., van Zuijen, T.L., Sussman, E., Horvath, J., Naatanen, R., 2006. Object representation in the human auditory system. Eur J. Neurosci. 24, 625–634.

Wrigley, S.N., Brown, G.J., 2004. A computational model of auditory selective attention. IEEE Trans. Neural Networks Special Issue Temp. Coding Neural Inform. Process. 15, 1151–1163.

Xu, Y., Chun, M.M., 2006. Dissociable neural mechanisms supporting visual short-term memory for objects. Nature 440, 91–95.

Yabe, H., Winkler, I., Czigler, I., Koyama, S., Kakigi, R., Sutoh, T., Hiruma, T., Kaneko, S., 2001. Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. Brain Res. 897, 222–227.