Hearing Research 304 (2013) 77–90

Contents lists available at SciVerse ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research paper

Brain dynamics encode the spectrotemporal boundaries of auditory objects

Amanda R. McMullan*, Dillon A. Hambrook, Matthew S. Tata

University of Lethbridge, Department of Neuroscience, 4401 University Drive W., Lethbridge, AB, Canada T1K 3M4

ARTICLE INFO

Article history:
Received 11 March 2013
Received in revised form
14 June 2013
Accepted 24 June 2013
Available online 2 July 2013

ABSTRACT

Perception of objects in the scene around us is effortless and intuitive, yet entails profound computational challenges. Progress has been made in understanding some mechanisms by which the brain encodes the boundaries and surfaces of visual objects. However, in the auditory domain, these mechanisms are poorly understood. We investigated differences between neural responses to spectrotemporal boundaries in the auditory scene. We used iterated rippled noise to create perceptual boundaries with and without energy transients. In contrast to boundaries marked by energy transients, second-order boundaries were characterized by an absence of early components in the event-related potential. First-order energy boundaries triggered a transient evoked gamma-band response and a well-defined P90 component of the event-related potential, whereas second-order boundaries evoked only the later N1 component. Furthermore, the N1 component was delayed when evoked by second-order boundaries and theta-band electroencephalography activity at this latency exhibited significant phase lag for second-order compared to first-order boundaries. We speculate that boundaries defined by sharp energy transients can be registered by early feed-forward mechanisms. By contrast, boundaries defined only by discontinuities at discrete frequency bands require integration across the tonotopic representation of the frequency spectrum and require time-consuming interaction between auditory areas.

1. Introduction

We effortlessly parse complex scenes into objects and background. In vision science, a substantial effort has been made to understand how the visual system is able to group surfaces and edges together into objects. These mechanisms are believed to be modular and hierarchical so that different aspects of visual input can be handled by specialized structures (Goodale and Milner, 1992; Marr, 1982; Zeki et al., 1991). Despite much progress in vision science, the mechanisms by which the auditory scene is analyzed for object boundaries remain unclear. Anatomical and functional evidence regarding the specificity of 'what' and 'where' *auditory* pathways suggests that the general process might be broadly analogous to that of the visual system (Alain et al., 2001b; Kubovy and Van Valkanburg, 2001; Rauschecker and Tian, 2000; Tian et al., 2001). The goal of the two experiments in this study was to identify neural signals associated with boundary or edge detection in the auditory system.

In vision, segmentation of objects and background can occur in response to first-order boundaries such as discontinuities in luminosity or brightness. Higher-order boundaries can be defined in terms of discontinuities in color (Shapley, 1990), texture (Mareschal and Baker, 1998), or motion (Anderson, 1997). In order to investigate first-order and higher-order boundaries in the auditory scene it is important to note that while the visual scene is spatiotopic, the auditory scene is tonotopic. Therefore, auditory boundaries are spectrotemporal in nature (Griffiths and Warren, 2002, 2004) and can be distinguished by discontinuities in the temporal evolution of features. In this sense, first-order auditory boundaries are increments or decrements of sound energy over time and higher-order auditory boundaries are spectrotemporal discontinuities such as those in pitch and interaural timing. For example, auditory boundaries can be defined and studied by temporal discontinuities in dichotic pitch (Johnson et al., 2003; Ungan et al., 1989), mistuned harmonics (Alain et al., 2001a, 2001b, 2002), energy-matched pitch (Krumbholz et al., 2003; Seither-Preisler et al., 2004, 2006b),

---

Abbreviations: FDR, false-discovery rate; IRN, iterated rippled noise; ITPC, inter-trial phase coherence; LORN, lateralized object-related negativity; ORN, object-related negativity; TSE, time-spectral evolution

* Corresponding author. Tel.: +1 403 394 3994; fax: +1 403 329 2775.
E-mail address: amanda.mcmullan@uleth.ca (A.R. McMullan).

0378-5955/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.heares.2013.06.009

motion (Butcher et al., 2011), organized and disorganized sequences of tones (Chait et al., 2008; Chait et al., 2007b), and interaural correlation (Chait et al., 2007a). In each of these examples, however, the distinction is made between the stimulus features that supports boundary detection and the functional anatomy that performs boundary detection. For example, it is not known how the functional anatomy of objects in vision corresponds to the functional anatomy of objects in audition. Non-analogous structures might perform analogous computational tasks.

Electroencephalography (EEG) and magnetoencephalography (MEG) have been useful in identifying brain responses to boundaries in the auditory scene. Robust deflections of the associated event-related potential (ERP) and the analogous event-related field (ERF) in MEG are triggered by the onsets of acoustic stimuli (e.g. Hillyard and Picton, 1978; Shahin et al., 2007). Well-known components of the ERP/ERF are modulated by both sensory and cognitive factors. For example, the N1 component of the ERP is modulated by attention (Hillyard et al., 1973; Näätänen and Picton, 1987) and the P300 components reflects not only cognitive factors, such as novelty of a stimulus, but also psychophysiological context, such as arousal (see Polich and Kok, 1995 for review). Other components of the event-related waveforms offer specific clues about neural mechanisms that represent boundaries in the environment. However, sound onsets typically have energy transients – increments or decrements in total acoustic power. Thus, ERP/ERF responses to first-order auditory boundaries are well studied, but offer little opportunity to distinguish low-level sensory responses from higher order boundary-detection processes.

Alain et al. (2001a), Alain et al. (2002), Dyson and Alain (2004), McDonald and Alain (2005), Snyder et al. (2006) identified a component of the auditory ERP that was associated with the registration of discrete objects in the auditory scene and termed this deflection the object-related negativity (ORN). Similarly, Butcher et al. (2011) identified an ERP component that was evident when auditory objects could be segmented by discontinuous motion trajectory in a virtual auditory scene. This deflection occurred on the side contralateral to the sound location and they accordingly termed this deflection the lateralized object-related negativity (LORN). In a set of investigations using iterated rippled noise (IRN), Krumbholz et al. (2003) reported an MEG response evoked by the transition from background noise to a discrete pitch with no acoustic energy change. They termed this the pitch onset response (POR). An important design feature shared by these lines of research is that perceptual objects were defined by spectrotemporal boundaries rather than energy boundaries. This allowed a distinction to be made between energy-related and other boundary-related responses. For example, the POR is at least partially distinct from a more general energy-onset response (Krumbholz et al., 2003; Seither-Preisler et al., 2004, 2006a).

Previous work on the neural correlates of auditory boundary representation has emphasized ERP/ERFs. The EEG/MEG signal is composed of power in several frequency bands, but the ERP/ERF tends to be dominated by low-frequency power in the theta (4–8 Hz) and alpha (8–12 Hz) bands. Time–frequency decomposition allows for investigation of changes in signal composition at many frequency bands simultaneously. In addition, ERP/ERF waveforms are averages of many trials and only reveal signals with high temporal coherence. However, analysis of the spectrotemporal decomposition of single trials can reveal signals that are not strictly phase-locked to stimuli (Ponjavic-Conte et al., 2013). In Experiment 1 below, we contrast boundaries defined by energy (first-order) with boundaries defined by pitch (second-order), with emphasis on characteristic differences in the time–frequency decomposition of the EEG. In Experiment 2, we considered how lateralization of the stimuli would affect responses to these boundaries.

2. Experiment one procedure

2.1. Participants

Nineteen students from the University of Lethbridge participated in Experiment 1 for course credit. The nature of the experiment was explained to the participants. Written informed consent was completed and the participants reported no hearing or neurological problems. Procedures were in accordance with the Declaration of Helsinki and approved by the University of Lethbridge Human Subjects Review Committee. Trials contaminated with eye artifact (>120 μV) were rejected and data from subjects with less than 50% of original trials were not analyzed, leaving fifteen participants (five male; mean age: 20.1; one left-handed).

2.2. Stimuli

The goal of this experiment was to measure differences in brain responses to different kinds of spectrotemporal boundaries. We refer to boundaries as ‘first order’ if they entailed a broadband energy increment or decrement relative to background noise (Fig. 1a). The energy change in this case is a global change in acoustic energy. Therefore, these first-order stimuli always contained an abrupt discontinuity in energy at all frequency bands. We refer to boundaries as ‘second order’ if they entailed spectrotemporal discontinuities, but the total acoustic energy of the stimulus remained constant (Fig. 1b). These second-order boundaries also contained some spectral discontinuity in amplitude at particular frequencies, however, the overall energy of the stimulus remained constant through time. Meaning that while global acoustic energy remains constant, there are local regions that experience increments while other local regions experience decrements. We used iterated rippled noise (IRN) to create second order boundaries. The IRN is produced using a delay-and-add algorithm developed by Yost (1996). The pitch of the IRN is determined by the reciprocal of the delay time and the salience is determined in part by the number of iterations of the delay. For the first experiment our IRN stimuli had 4 iterations of a time delay of 40 samples, given a 32,000 Hz sample rate, resulting in a 1.25 ms time delay. The perceived pitch of these IRN stimuli was 800 Hz. IRN stimuli were scaled so that the root-mean-square (RMS) amplitude matched that of the temporally adjacent broadband noise over the entire spectrum. We also considered that the transitions between noise and IRN could be marked by changes in loudness despite constant RMS. We applied Moore’s time-varying loudness model (Glasberg and Moore, 2002) and confirmed that the stimuli did not exhibit a loudness modulation at this boundary (Fig. 1c and d). It is important to note that, by definition, boundaries between broadband noise and IRN feature energy increments at some frequencies and decrements at others, which can be seen in Fig. 1e and f for first-order and second-order boundaries, respectively. Fig. 1 also shows the spectral composition of the IRN stimulus (Fig. 1g) and the white noise (Fig. 1h).

Experiment 1 had three stimulus conditions. These varied in the type of spectrotemporal discontinuity used to define a target sound. In the silence-to-IRN condition, IRN stimuli were interleaved with silent intervals. Participants responded when they detected the onsets of IRN (i.e. IRN was the target). In the noise-to-IRN and the IRN-to-noise conditions the IRN stimulus alternated with broadband noise. Participants responded to IRN onsets in the noise-to-IRN condition (i.e. IRN was the target) and they responded to noise onsets in the IRN-to-noise condition (i.e. noise was the target). Stimuli were presented through a speaker (Mackie HR-624mk2 studio monitor) directly in front of the participant. Twenty blocks of each condition were presented in pseudorandom