NeuroImage 126 (2016) 164–172

Contents lists available at ScienceDirect

NeuroImage

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

Neural dynamics of change detection in crowded acoustic scenes

Ediz Sohoglu *, Maria Chait *

UCL Ear Institute, 332 Gray's Inn Road, London WC1X 8EE, UK

ARTICLE INFO

Article history:
Received 20 July 2015
Accepted 22 November 2015
Available online 26 November 2015

Keywords:
Change deafness
Auditory scene analysis
MEG

ABSTRACT

Two key questions concerning change detection in crowded acoustic environments are the extent to which cortical processing is specialized for different forms of acoustic change and when in the time-course of cortical processing neural activity becomes predictive of behavioral outcomes. Here, we address these issues by using magnetoencephalography (MEG) to probe the cortical dynamics of change detection in ongoing acoustic scenes containing as many as ten concurrent sources. Each source was formed of a sequence of tone pips with a unique carrier frequency and temporal modulation pattern, designed to mimic the spectrotemporal structure of natural sounds. Our results show that listeners are more accurate and quicker to detect the appearance (than disappearance) of an auditory source in the ongoing scene. Underpinning this behavioral asymmetry are change-evoked responses differing not only in magnitude and latency, but also in their spatial patterns. We find that even the earliest (~50 ms) cortical response to change is predictive of behavioral outcomes (detection times), consistent with the hypothesized role of local neural transients in supporting change detection.

Introduction

A key aspect of the process by which our brains analyze, represent and make sense of our surroundings is the ability to rapidly detect changes in the ongoing sensory input. The auditory system is hypothesized to play a central role in the brain's change detection network by serving as an 'early warning' device, continually monitoring the acoustic background for potentially relevant events (Demany et al., 2008; Murphy et al., 2013). Although the neural mechanisms by which listeners detect change in simple acoustic patterns have been extensively investigated (e.g. Martin and Boothroyd, 2000; Krumbholz, 2003; Gutschalk et al., 2004; Näätänen et al., 2007; Chait et al., 2008; Grimm et al., 2011; Andreou et al., 2015), how change is detected in crowded acoustic scenes containing multiple concurrent sources remains poorly understood.

Previous neuroimaging studies of auditory change detection suggest that changes in crowded acoustic scenes are successfully encoded by the earliest stages of cortical processing in primary auditory cortex (Puschmann et al., 2013a, 2013b). It is only later stages of processing in non-primary temporal and frontal regions that determine whether listeners report hearing a change (Gregg and Snyder, 2012; Puschmann et al., 2013a, 2013b; Gregg et al., 2014; Snyder et al., 2015). These findings are compatible with the notion that when detecting change, the behavioral outcome depends on the success of higher-level processes that extract object-based perceptual representations from acoustic scenes (Eramudugolla et al., 2005; Backer and Alain, 2012) or that maintain and compare information from prechange and postchange portions of the sensory input (Eramudugolla et al., 2005; Gregg and Samuel, 2008; Pavani and Turatto, 2008).

However, a common feature of previous neuroimaging studies of auditory change detection (Gregg and Snyder, 2012; Puschmann et al., 2013a, 2013b; Gregg et al., 2014) is the use of silent or noise interruptions separating the pre-change and post-change scenes (see also Eramudugolla et al., 2005). Consequently, the extent to which the results might generalize to naturalistic listening situations in which changes occur in an uninterrupted, ongoing scene is unclear (Cervantes Constantino et al., 2012). In particular, it is likely that the scene interruptions masked local neural transients evoked by change, thereby forcing listeners to rely on higher-level processes that encode and compare pre-interruption and post-interruption scene information in a working memory store (see Rensink et al., 1997; Demany et al., 2008). Indeed, in a series of behavioral experiments, Cervantes Constantino et al. (2012) demonstrated that auditory change detection is at least partly reliant on local transients, similar to what has been established for visual change detection (Yantis and Jonides, 1984; Rensink et al., 1997).

A further question concerns the extent to which the neural mechanisms supporting change detection are specialized for different forms of acoustic change (see Molholm et al., 2005). Previous behavioral investigations suggest that listeners are more accurate and quicker to detect a change involving the appearance, as opposed to the disappearance, of an auditory object (Huron, 1989; Pavani and Turatto, 2008; Cervantes Constantino et al., 2012). This perceptual asymmetry may have an origin in known differences in neural responses evoked by the onset versus offset of sound, which include amplitude, latency and spatial distribution

---

* Corresponding authors.
E-mail addresses: e.sohoglu@gmail.com (E. Sohoglu), m.chait@ucl.ac.uk (M. Chait).

http://dx.doi.org/10.1016/j.neuroimage.2015.11.050
1053-8119/

(Hari et al., 1987; Pantev et al., 1996; Phillips et al., 2002; Qin et al., 2007; Pratt et al., 2008; Scholl et al., 2010). However, as previous work on onset vs. offset detection only investigated neural responses to single sounds, and in passive listening situations, the extent to which neural processing is specialized for detecting appearing and disappearing objects in crowded acoustic scenes is unknown.

In the current study, we recorded magnetoencephalography (MEG) brain responses while listeners were presented with artificial acoustic scenes containing as many as ten auditory objects, each formed of a sequence of tone pips with a unique carrier frequency and amplitude modulation pattern. The task for listeners was to detect a change involving the appearance or disappearance of one of those objects within the scene. Our aims were twofold: 1) to characterize neural responses to appearing and disappearing objects in an ongoing, crowded acoustic scene and 2) to determine which stage of neural processing contributes to detection success by relating MEG responses to behavioral outcomes.

Methods

Participants

14 (5 female) right-handed participants aged between 22 and 36 years (mean = 27.8, SD = 3.98) were tested after being informed of the study's procedure, which was approved by the research ethics committee of University College London. All reported normal hearing, normal or corrected-to-normal vision, and had no history of neurological disorders.

Stimuli

Stimuli were 2500–3500 ms duration artificial acoustic 'scenes' populated by four or ten streams of pure-tones designed to model auditory sources (shown in Fig. 1). Each source had a unique carrier frequency (drawn from a pool of fixed values spaced at 2*ERB between 200 and 2800 Hz; Moore and Glasberg, 1983) and temporal structure. Within each object, the durations of the tones (varying uniformly between 22 and 167 ms) and the silent interval between tones (varying between 1 and 167 ms) were chosen independently and then fixed for the duration of the scene. This pattern mimics the regularly modulated temporal properties of many natural sounds. Previous experiments have demonstrated that each stimulus is perceived as a composite 'sound-scape' in which individual objects can be perceptually segregated and selectively attended to, and are therefore good models for listening in natural acoustic scenes (Cervantes Constantino et al., 2012). Importantly, the large spectral separation between neighboring objects (at least 2*ERB) minimizes energetic masking at the peripheral stages of the auditory system (Moore, 1987), thus enabling the investigation of the effects of increasing scene size without the confound of increasing inter-object sensory masking. Signals were synthesized with a sampling rate of 44100 Hz and shaped with a 30 ms raised cosine onset and offset ramp. They were delivered diotically to the subjects' ears with tubephones (EARTONE 3A 10 Ω, Etymotic Research, Inc) inserted into the ear-canal and adjusted to a comfortable listening level.

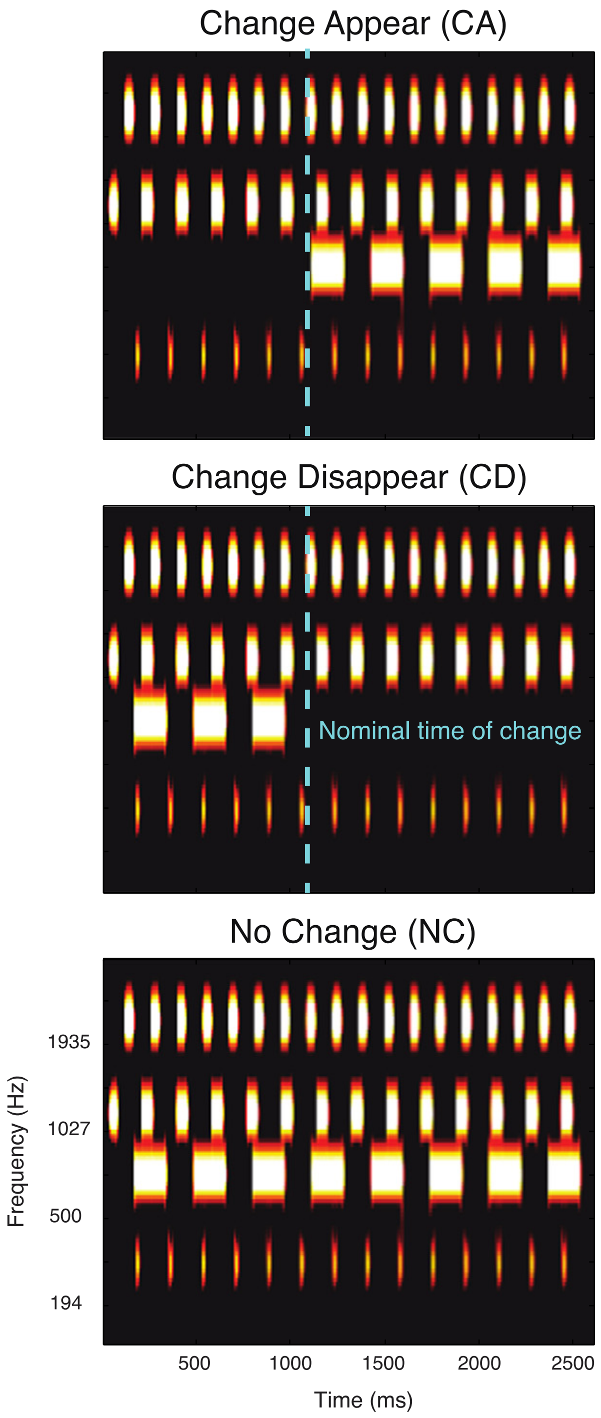

Fig. 1. Example acoustic scenes containing four objects (scene size four condition). The plots represent 'auditory' spectrograms, equally spaced on a scale of ERB-rate (Moore and Glasberg, 1983). Channels are smoothed to obtain a temporal resolution similar to the Equivalent Rectangular Duration (Plack and Moore, 1990). Dashed lines show the nominal change time of appearing and disappearing objects.

As shown in Fig. 1, acoustic scenes in which each object was active throughout the stimulus are referred to as 'No Change' stimuli (NC). In other scenes, either of two types of change could occur partway through the stimulus: one that involved the appearance of a new object into the scene or one that involved the disappearance of an object from the scene. These are referred to as 'Change Appear' (CA) and 'Change Disappear' (CD) stimuli, respectively. Importantly, the other (non-changing objects) in the scene remained active without interruption. The specific configuration of carrier frequencies and temporal modulation patterns varied randomly across scenes. To enable a controlled comparison between conditions, NC, CA and CD stimuli were derived from each configuration of carrier frequencies and modulation patterns, and then presented in random order during the experiment. Fig. 1 shows an example of one such configuration.

The timing of change varied randomly (uniformly distributed between 1000 ms and 2000 ms post scene onset), but with the following constraints: The nominal time of change for CA objects coincided with the onset of the first tone while for CD objects, the nominal time of change was set to the offset of the last tone augmented by the inter-tone interval, i.e. at the expected onset of the next tone, which is the