Hearing Research 307 (2014) 111–120

Contents lists available at SciVerse ScienceDirect

Hearing Research

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

Review

Using neuroimaging to understand the cortical mechanisms of auditory selective attention

Adrian K.C. Lee [a,b,*], Eric Larson [a], Ross K. Maddox [a], Barbara G. Shinn-Cunningham [c,d]

[a] Institute for Learning and Brain Sciences, University of Washington, WA 98195, USA
[b] Department of Speech & Hearing Sciences, University of Washington, Seattle, WA 98195, USA
[c] Center for Computational Neuroscience and Neural Technology, Boston University, Boston, MA 02215, USA
[d] Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

ARTICLE INFO

Article history:
Received 13 May 2013
Received in revised form
20 June 2013
Accepted 25 June 2013
Available online 9 July 2013

ABSTRACT

Over the last four decades, a range of different neuroimaging tools have been used to study human auditory attention, spanning from classic event-related potential studies using electroencephalography to modern multimodal imaging approaches (e.g., combining anatomical information based on magnetic resonance imaging with magneto- and electroencephalography). This review begins by exploring the different strengths and limitations inherent to different neuroimaging methods, and then outlines some common behavioral paradigms that have been adopted to study auditory attention. We argue that in order to design a neuroimaging experiment that produces interpretable, unambiguous results, the experimenter must not only have a deep appreciation of the imaging technique employed, but also a sophisticated understanding of perception and behavior. Only with the proper caveats in mind can one begin to infer how the cortex supports a human in solving the "cocktail party" problem.

This article is part of a Special Issue entitled <Human Auditory Neuroimaging>.

1. Introduction

"How do we recognize what one person is saying when others are speaking at the same time?" With this question, E. Colin Cherry defined the "Cocktail Party Problem" six decades ago (Cherry, 1953). Attention often requires a process of selection (Carrasco, 2011). Selection is necessary because there are distinct limits on our capacity to process incoming sensory information, resulting in constant competition between inner goals and external demands (Corbetta et al., 2008). For example, eavesdropping on a particular conversation in a crowded restaurant requires top-down attention, but as soon as a baby starts to cry, this salient stimulus captures our attention automatically, due to bottom-up processing. The fact that sufficiently salient stimuli can break through our attentional focus demonstrates that all sound is processed to some degree, even when not the focus of volitional attention; however, the stimulus that is selected, whether through top-down or bottom-up control, is processed in greater detail, requiring central resources that are limited (Desimone and Duncan, 1995). In order to operate effectively in such environments, one must be able to i) select objects of interest based on their features (e.g., spatial location, pitch) and ii) be flexible in maintaining attention on and switching attention between objects as behavioral priorities and/or acoustic scenes change. In vision research, there is a large body of work documenting the competitive interaction between volitional, top-down control and automatic, bottom-up enhancement of salient stimuli (Knudsen, 2007). However, there are comparatively fewer studies investigating how object-based auditory attention operates in complex acoustic scenes (Shinn-Cunningham, 2008). By utilizing different human neuroimaging techniques, we are beginning to understand the cortical dynamics associated with directing and redirecting auditory attention.

This review begins by providing a brief overview of neuroimaging approaches commonly used in auditory attention studies. Particular emphasis is placed on functional magnetic resonance imaging (fMRI),

Abbreviations: AESPA, auditory evoked spread spectrum analysis; BOLD, blood-oxygenation level dependent; ECD, equivalent current dipole; EEG, electroencephalography; ERF, event-related field; ERP, event-related potential; FEF, frontal eye fields; (f)MRI, (functional) magnetic resonance imaging; HG, Heschl's gyrus; IPS, intraparietal sulcus; NIRS, near-infrared spectroscopy; MEG, magnetoencephalography; MMN, mismatch negativity; RON, reorienting negativity; PET, positron emission tomography; PP, planum polare; PPC, posterior parietal cortex; PT, planum temporale; (r)MFG, (right) middle frontal gyrus; (r)TPJ, (right) temporoparietal junction; STS, superior temporal sulcus; TMS, transcranial magnetic stimulation

* Corresponding author. Institute for Learning and Brain Sciences, University of Washington, 1715 NE Columbia Road, Box 357988, Seattle, WA 98195, USA. Tel.: +1 206 616 0102; fax: +1 206 221 6472.
E-mail address: akclee@uw.edu (A.K.C. Lee).

0378-5955/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.heares.2013.06.010

magnetoencephalography (MEG) and electroencephalography (EEG) because these modalities are currently used more often than other non-invasive imaging techniques, such as positron emission tomography (PET) or near-infrared spectroscopy (NIRS). To facilitate a fuller understanding of the array of neuroimaging studies, we discuss the strengths and limitations of each imaging technique as well as the ways in which the technique employed can influence how the results can be interpreted. We then review evidence that attention modulates cortical responses both in and beyond early auditory cortical areas (for a review of auditory cortex anatomy, see Da Costa et al., 2011; Woods et al., 2010). There are many models to describe auditory attention, including phenomenological models (e.g., Näätänen, 1990), which accounts for attention and automaticity in sensory organization while focusing on human neuroelectric data), behavioral models (e.g., Cowan, 1988), and neurobiological models (e.g., McLachlan and Wilson, 2010). Many processes, from organizing the auditory scene into perceptual objects to dividing attention across multiple talkers in a crowded environment, influence auditory attention. These processes are discussed in a recent comprehensive review (Fritz et al., 2007). Here, we focus on selective attention, which Cherry cites as the key issue in allowing us to communicate in crowded cocktail parties. Moreover, we use as an organizing hypothesis the idea that all forms of selective attention operate on perceptual objects, so in this review we focus on object-based attention (see Shinn-Cunningham, 2008 for review). This also enables us to compare and contrast results with those from the visual attention literature. We conclude by highlighting other important questions in the field of auditory attention and neuroimaging.

2. Methodological approaches

2.1. Spatial and temporal resolution considerations

Magneto- and electroencephalography (MEG, EEG; M/EEG when combined) record extracranial magnetic fields and scalp potentials that are thought to reflect synchronous post-synaptic current flow in large numbers of neurons (Hämäläinen et al., 1993). Both technologies can detect activity on the millisecond time scale characteristic of communication between neurons; the typical sampling frequency ($\sim$1000 Hz) makes it particularly suited to studies of auditory processing, given the importance of temporal information in the auditory modality. There are important differences between MEG and EEG. For example, the skull and scalp distort magnetic fields less than electric fields, so that MEG signals are often more robust than the corresponding EEG signals. MEG is also mainly sensitive to neural sources oriented tangentially to the skull, whereas EEG is sensitive to both radially and tangentially oriented neural sources. When MEG and EEG are used simultaneously, they can provide additional complementary information about the underlying cortical activities (Ahlfors et al., 2010; Goldenholz et al., 2009; Sharon et al., 2007). By using anatomical information obtained from magnetic resonance imaging (MRI) to constrain estimates of neural sources of observed activity, reasonable spatial resolution of cortical source can be achieved (Lee et al., 2012).

Functional MRI is another widely used non-invasive neuroimaging technique. It measures the blood-oxygenation level dependent (BOLD) signal, which reflects local changes in oxygen consumption. This BOLD signal is used as a proxy for neural activity in a particular cortical (or subcortical) location; this assumption is supported by the fact that the BOLD signal correlates strongly with the underlying local field potential in many cases (Ekstrom, 2010; Logothetis, 2008). Compared to M/EEG, fMRI has much better spatial resolution (better by a factor of about 2–3) but poorer temporal resolution (worse by about a factor of 1000, due to the temporal sluggishness of the BOLD signal). Fig. 1 provides a summary of the tradeoffs between spatial and temporal resolution for these neuroimaging approaches.

In designing experiments, the scientific question being asked should inform the choice of which neuroimaging technique to use. For example, due to its superior spatial resolution, fMRI is well suited for a study to tease apart precisely what anatomical regions are engaged in particular tasks (e.g., comparisons of “what”/“where” processing within auditory cortical areas); in contrast, M/EEG can tease apart the dynamics of cortical activity (e.g., to temporally distinguish neural activity associated with top-down control signals before a sound stimulus begins from the signals effecting selective attention when the stimulus is playing). Other factors, apart from considerations of spatial and temporal resolution, also influence both the choice of neuroimaging technique to use and the way to interpret obtained results. These factors are summarized below.

2.2. Other tradeoffs related to attention studies in different techniques

2.2.1. fMRI scanner noise

In order to achieve good spatial and temporal resolution along with high signal-to-noise ratios, MRI scanners need powerful magnetic fields and fast switching of magnetic gradients. When a current is passed through coils inside the MRI scanner to set up these gradients, the resulting Lorentz forces cause them to vibrate, generating acoustic noise that can exceed 110 dB SPL (Counter et al., 2000; Hamaguchi et al., 2011). This scanner noise is part of the auditory scene that a subject hears during an fMRI study (Mathiak et al., 2002). As a result, in auditory paradigms involving attentional manipulation, brain activity will reflect not only activity in response to the controlled auditory stimuli, but also in response to the scanner noise, e.g., inducing involuntary orienting (Novitski et al., 2001). Sparse temporal sampling (Hall et al., 1999), wherein the stimulus is presented during silent periods between imaging acquisition, is commonly used to reduce the influence of scanner noise on the brain activity being measured. However, this technique significantly reduces the number of imaging volumes that can be acquired in a given experiment, which lowers the signal-to-noise ratio compared to continuous scanning (Huang et al., 2012). A sparse sampling strategy also decreases the temporal resolution of the measured signal acquired, making it much more difficult to estimate BOLD time courses. The sparse sampling technique does not eliminate scanner noise; it only controls the timing of the noise. Thus, the scanner noise still interacts with the controlled sound stimuli. For example, in an fMRI streaming experiment using sparse sampling, scanner noise contributed to an abnormal streaming build-up pattern (Cusack, 2005). This is consistent with the observation that auditory attention influences the formation of auditory streams (Cusack et al., 2004). Furthermore, the

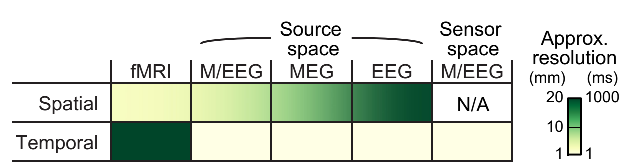

Fig. 1. Approximate spatial resolution and temporal resolution differ dramatically across imaging modalities. While fMRI has excellent spatial resolution (sub-centimeter) compared to M/EEG (around a centimeter), it has comparatively poor temporal resolution (seconds versus milliseconds, respectively). Sensor space analysis is based directly on the field topographical patterns (see Section 2.2.2), while source space analysis seeks to map the topographical patterns to the underlying neural sources analysis using ECD or inverse modeling (see Section 2.2.3).