Brain & Language 126 (2013) 29–48

Contents lists available at SciVerse ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

The organization and physiology of the auditory thalamus and its role in processing acoustic features important for speech perception

Edward L. Bartlett *

Dept. of Biological Sciences, Purdue University, United States
Weldon School of Biomedical Engineering, Purdue University, United States

ARTICLE INFO

Keywords:
Inferior colliculus
Thalamocortical
Corticothalamic
Amplitude modulation
Calbindin
Monotonic
Vocalizations
Marmoset
Thalamic reticular nucleus

ABSTRACT

The auditory thalamus, or medial geniculate body (MGB), is the primary sensory input to auditory cortex. Therefore, it plays a critical role in the complex auditory processing necessary for robust speech perception. This review will describe the functional organization of the thalamus as it relates to processing acoustic features important for speech perception, focusing on thalamic nuclei that relate to auditory representations of language sounds. The MGB can be divided into three main subdivisions, the ventral, dorsal, and medial subdivisions, each with different connectivity, auditory response properties, neuronal properties, and synaptic properties. Together, the MGB subdivisions actively and dynamically shape complex auditory processing and form ongoing communication loops with auditory cortex and subcortical structures.

1. Introduction

Much of the research focus on the sensory and cognitive aspects of language centers on cortical activities. The cerebral cortex, however, is inextricably linked to the collection of forebrain nuclei known as the thalamus. Thalamocortical pathways dictate the sensory and higher-order representations in cortex while corticothalamic pathways generate dynamic, context-dependent changes in thalamic responsiveness to form an iterative signaling loop. This review will describe the functional organization of the auditory thalamus as it relates to the representation of sound features that are relevant for speech perception.

The thalamus is a collection of nuclei whose main and best-studied projections are to the cerebral cortex, comprising projections to all areas of cortex. The main auditory-responsive portion of the thalamus is called the medial geniculate body (MGB), and it is the information bottleneck for neural representations of sounds being sent to auditory cortex.

Whereas early views of the thalamus were that it served as a simple 'relay' or 'gateway' to the cortex, numerous studies have demonstrated that thalamic neurons transform their inputs en route to their cortical or subcortical targets (Hubel & Wiesel, 1961; Sherman & Guillery, 2002). As we will show, the MGB actively and dynamically shapes the auditory representations that reach the cerebral cortex. Rather than acting as a simple conduit for incoming auditory representations, the MGB acts like a funhouse mirror in the sense that it can filter and distort incoming inputs to enhance representation and perception of acoustic features for use by the auditory cortex (AC).

2. Acoustic characteristics of the speech signal relevant for auditory processing and receptive language

Human speech and animal vocalizations have complex sound characteristics that must be represented accurately by neural populations in order to recognize, understand, and converse with individuals in a variety of sound backgrounds and in the presence of multiple speakers. For speech, this includes not only the ability to identify the words that are spoken, but also the identity and gender of the speaker as well as the emotional content of their speech. Unlike the tones often used to probe auditory function in the lab, the speech spectrum is distributed over multiple octaves ranging from approximately 0.1 to 10 kHz. Although there are spectral peaks in the speech spectrum, such as in formants, there is also significant energy in non-peak frequencies.

One way that the speech signal can be parsed by auditory neuroscientists for understanding neural representations is to decompose the acoustic signal into components defined by different frequency ranges (Rosen, 1992). With appropriate spectral resolution in quiet conditions, most speech fluctuations important for intelligibility occur at <10 Hz (Elliott & Theunissen, 2009). Under a variety of conditions of spectral degradation, envelope modulations of up to 50 Hz are able to produce an adequate

* Address: 206 S. Jischke Dr., West Lafayette, IN 47907, United States. Fax: +1 765 496 1912.
E-mail address: ebartle@purdue.edu

0093-934X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bandl.2013.03.003

level of speech intelligibility (Elliott & Theunissen, 2009; Shannon, et al., 1995). Thus, these modulation frequencies in the 1–10 Hz and 10–50 Hz range, collectively called the sound envelope, are important for speech intelligibility. However, for gender, emotion, and speaker identification, higher frequencies are needed. Higher frequency modulations in the range of 50–500 Hz are referred to as stimulus periodicity or temporal fine structure (Rosen, 1992). This range also overlaps with the fundamental frequencies of speaker's voices. Higher frequencies in speech signal contribute to speaker identity and emotional content and speech perception in noise (Rosen, 1992). What separates these three frequency regions (<50 Hz, 50–500 Hz, and >500 Hz) from a neural standpoint are the auditory nuclei that are able to encode the modulations in a stimulus-synchronized manner. All stations in the ascending auditory pathway, including the auditory cortex, are able to encode stimulus envelope. Temporal fine structure cues (>50 Hz) are generally not represented by stimulus-synchronized responses in the auditory cortex, but often are in the auditory thalamus and inferior colliculus (Bartlett & Wang, 2011; Krishna & Semple, 2000; Rouiller, de Ribaupierre, Toros-Morel, & de Ribaupierre, 1981). Finally, while auditory nerve and some cochlear nucleus fibers can synchronize to stimulus modulations or carriers up to 5 kHz, these high frequency fluctuations are encoded as changes in firing rates at higher levels of the auditory system (Joris, Schreiner, & Rees, 2004).

In addition to temporal processing, the ability to parse the acoustic speech signal into different frequency bands is critical for speech comprehension. This is a job that is performed well by a properly functioning cochlea and then subsequently sharpened in the auditory thalamus and cortex of primates (Bartlett et al., 2011; Bitterman, Mukamel, Malach, Fried, & Nelken, 2008). Whereas adequate speech comprehension can occur with as few as 4 logarithmically spaced frequency channels (Shannon, et al., 1995), comprehension improves in normal hearing listeners with up to 20 channels (Baskent & Shannon, 2006; Friesen, Shannon, Baskent, & Wang, 2001). Furthermore, gender identification and music recognition/appreciation increase with increasing spectral resolution of the auditory signal (Elliott & Theunissen, 2009; Shannon, 2005). Complicating the matter even further is that speech may be understood over a wide range of intensities, from whispers to shouts (40–60 dB dynamic range).

Therefore, for the difficult perceptual tasks of segregating and representing speech signals for comprehension, speaker identification, and emotional content in a variety of backgrounds and over a wide range of sound levels, the auditory system must be able to:

1. Represent and segregate carrier frequencies with high resolution.
2. Represent temporal modulations up to 500 Hz, especially those $\leqslant$50 Hz.
3. Maintain neural representations over a large range of sound levels.

The MGB is heavily involved in shaping these representations in primates and other mammals. The following section will demonstrate the organization of the MGB, how the MGB transforms neural representations of auditory inputs, how representations are shaped by contextual and non-sensory factors, and how the cellular machinery used by MGB neurons contributes to the neural representations.

3. Overview of the anatomical and functional organization of the MGB

The MGB can be divided into three broad subdivisions, whose organization and properties are summarized in Table 1 and whose spatial arrangement can be seen in Fig. 1, going from rostral to caudal, using the marmoset as an example. The ventral division of the MGB (MGV) is the "core" subdivision for the rapid transmission of auditory information that is sharply tuned for frequency and able to respond to fast temporal modulations of sounds. For primates, this pathway also includes the anterodorsal nucleus of MGB (MGAD). These regions project to core regions in auditory cortex (AC), such as primary auditory cortex (A1), and stain heavily for parvalbumin, cytochrome oxidase and acetylcholinesterase (Cruikshank et al., 2001; de la Mothe, Blumell, Kajikawa, & Hackett, 2006; Hashikawa, Rausell, Molinari, & Jones, 1991; Jones, 2003). The dorsal division of the MGB (MGD) is a slower, more integrative

Table 1
Summary of properties of MGB subdivisions.

	MGV	MGD	MGM
Main cell type	Tufted	Stellate	Stellate/magnocellular
Tonotopic organization?	Yes	No	Yes, weaker than MGV
Calcium-binding proteins	Parvalbumin +++, Calbindin +,++, Cytochrome oxidase +++	Calbindin +++, Parvalbumin + Cytochrome oxidase+	Calbindin ++,+++, Parvalbumin +,++, Cytochrome oxidase ++
Main IC input	IC-central nucleus	IC-dorsal cortex	IC-external cortex
Major non-IC inputs (besides cortex)	Neuromodulators (acetylcholine, serotonin, etc.)	Lateral tegmentum, sagulum neuromodulators	Spinal cord, vestibular nuclei, hypothalamus, superior colliculus, neuromodulators
Cortical target	A1, non-A1 core, layers 3/4	Belt and parabelt, weak core, layers 3/4 and some layers 1 and 6	Core, belt, and parabelt, all layers,
Subcortical targets	None	Amygdala	Amygdala, striatum, IC
Cortical feedback	Layer 6, core	Layer 6, belt, layer 5, core and belt	Layer 6, core, belt
Tone frequency tuning	Narrow, usually single-peaked	Broad, inhibited, variable, multi-peaked	Heterogeneous, narrow and broad
Tone latency	Short	Longer	Heterogeneous, short and long
AM response	Synchronized to rapid modulation frequencies	Often non-synchronized or synchronized to low modulation frequencies	Heterogeneous, some synchronized to very rapid modulation frequencies
Receptive field plasticity	Sharply tuned near BF, short duration	Shift in BF to CS tone, lasts $\geqslant$24 h	Shift in BF to CS tone, lasts $\geqslant$24 h
Modulation of responses by non-auditory input	No	Yes, visual	Yes, visual and somatosensory
Modulation of responses by reward	No	Yes	Yes
IC EPSP properties	Large, stronger depression, and small, weak depression or facilitation	Small, weak depression or facilitation	Small, weak depression or facilitation
IC inhibition	$GABA_A$ and $GABA_B$	$GABA_A$ and $GABA_B$	$GABA_A$ only

Abbreviations: MGV – medial geniculate, ventral division. MGD – medial geniculate, dorsal division. Includes posterodorsal division of primates. MGM – medial geniculate, medial division.