NeuroImage 150 (2017) 344–357

Contents lists available at ScienceDirect

NeuroImage

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

Characterization of neural entrainment to speech with and without slow spectral energy fluctuations in laminar recordings in monkey A1

Benedikt Zoefel[a,b,c,*], Jordi Costa-Faidella[c,d,e], Peter Lakatos[c,f], Charles E. Schroeder[c,g], Rufin VanRullen[a,b]

[a] Université Paul Sabatier, Toulouse, France
[b] Centre de Recherche Cerveau et Cognition (CerCo), CNRS, UMR5549, Pavillon Baudot CHU Purpan, BP 25202, 31052 Toulouse Cedex, France
[c] Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, United States
[d] Institute of Neurosciences, University of Barcelona, Barcelona, Catalonia 08035, Spain
[e] Brainlab – Cognitive Neuroscience Research Group, Department of Clinical Psychology and Psychobiology, University of Barcelona, Barcelona, Catalonia 08035, Spain
[f] Department of Psychiatry, New York University School of Medicine, New York, NY, United States
[g] Departments of Neurosurgery and Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, United States

ARTICLE INFO

Keywords:
Neural oscillations
Phase
Entrainment
Speech
A1
Monkey
Phase-amplitude coupling

ABSTRACT

Neural entrainment, the alignment between neural oscillations and rhythmic stimulation, is omnipresent in current theories of speech processing – nevertheless, the underlying neural mechanisms are still largely unknown. Here, we hypothesized that laminar recordings in non-human primates provide us with important insight into these mechanisms, in particular with respect to processing in cortical layers. We presented one monkey with human everyday speech sounds and recorded neural (as current-source density, CSD) oscillations in primary auditory cortex (A1). We observed that the high-excitability phase of neural oscillations was only aligned with those spectral components of speech the recording site was tuned to; the opposite, low-excitability phase was aligned with other spectral components. As low- and high-frequency components in speech alternate, this finding might reflect a particularly efficient way of stimulus processing that includes the preparation of the relevant neuronal populations to the upcoming input. Moreover, presenting speech/noise sounds without systematic fluctuations in amplitude and spectral content and their time-reversed versions, we found significant entrainment in all conditions and cortical layers. When compared with everyday speech, the entrainment in the speech/noise conditions was characterized by a change in the phase relation between neural signal and stimulus and the low-frequency neural phase was dominantly coupled to activity in a lower gamma-band. These results show that neural entrainment in response to speech without slow fluctuations in spectral energy includes a process with specific characteristics that is presumably preserved across species.

Introduction

Many stimuli in the auditory environment – such as speech sounds – are rhythmic and alternate between important and less relevant events. Brain activity can be rhythmic as well: Neural oscillations reflect changes between high and low excitability phases of neuronal populations (Buzsáki and Draguhn, 2004). It has been proposed that these oscillations can be seen as alternations between open and closed “windows of opportunity” (more or less optimal moments) for input to be processed (Jensen et al., 2012). Thus, it is a reasonable assumption that the auditory system tries to align its oscillations with external rhythms (Calderone et al., 2014; Lakatos et al., 2008). Indeed, the alignment between neural oscillations and speech has been associated with an improved speech comprehension (Ahissar et al., 2001; Luo and Poeppel, 2007; Park et al., 2015). The precise neural mechanisms underlying this phenomenon are currently debated (see Zoefel and VanRullen, 2015c, for discussion): Examples include a periodic adjustment of the phase of neural oscillations (McAuley, 1995), potentially via phase-reset (Schroeder and Lakatos, 2009), or the involvement of a balanced relation between excitatory and inhibitory neurons (Fries, 2005). In the present work, we therefore use the neutral term “neural entrainment” and define it as the alignment of an internal oscillatory (e.g., electrophysiological) signal to an external rhythm (e.g., speech).

Besides very few intracranial studies (Fontolan et al., 2014; Nourski

---

* Correspondence to: Medical Research Council, Cognition and Brain Sciences Unit, 15 Chaucer Rd, Cambridge CB2 7EF, United Kingdom.
E-mail address: benedikt.zoefel@mrc-cbu.cam.ac.uk (B. Zoefel).

http://dx.doi.org/10.1016/j.neuroimage.2017.02.014
Received 30 September 2016; Accepted 6 February 2017

et al., 2009), reports of neural entrainment are based on surface electrophysiological recordings with relatively low spatial resolution (using, e.g., electro-/magnetoencephalogram, EEG/MEG; e.g., Ahissar et al., 2001; Baltzell et al., 2016; Crosse et al., 2016, 2015; Di Liberto et al., 2015; Ding and Simon, 2013, 2012; Ding et al., 2016; Doelling et al., 2014; Gross et al., 2013; Horton et al., 2013; Kayser et al., 2015; Luo and Poeppel, 2007; Millman et al., 2015; Peelle et al., 2013; Zoefel and VanRullen, 2015b). The characterization of neural entrainment to speech sounds with respect to laminar processing in auditory cortex would thus represent a step forward in our understanding of the brain's processing of speech and, more generally, of rhythmic input. For this purpose, recordings in monkey auditory cortex are an important tool (Rauschecker and Scott, 2009): First, monkeys and humans are genetically closely related. Second, in contrast to humans, laminar profiles can be easily recorded in monkeys. Third, entrainment of neural oscillations has been demonstrated repeatedly in monkeys (Lakatos et al., 2005b, 2008, 2013), indicating that humans and monkeys share a common mechanism of adaptation to rhythmicity.

We are aware of only one other study[1] showing neural responses in monkey A1 during the presentation of common human speech (Steinschneider et al., 2013). However, this study was limited to the presentation of words and focused on neural responses to phonemes. In the current study, we presented long (five one-minute) sequences of everyday speech and measured the entrainment of neural oscillations in the different cortical layers of A1. A1 is tonotopically organized (Merzenich and Brugge, 1973) and its spectral tuning determines how oscillations entrain to sound input: For instance, it has been shown that a given stimulus does not only affect the phase of neural oscillations in the area of A1 tuned to the stimulus frequency (defined as “best frequency” region, or BF region), but also in the rest of auditory cortex – by aligning the high- and low-excitability phase of oscillations in BF- and non-BF regions with the expected sound onset, respectively (Lakatos et al., 2013; O’Connell et al., 2011). However, these results were reported using trains of pure tones as stimuli; we were therefore interested in how oscillations entrain to spectrally complex sounds such as speech, and how the tonotopic organization of A1 affects this mechanism.

Although everyday speech is an interesting stimulus, it contains large fluctuations in amplitude and spectral content (Fig. 1A, top; Fig. 1B, left) – any observed entrainment can thus be “biased”, as it cannot be ruled out that it entails a passive “following” of these fluctuations at very early levels of auditory processing (e.g., a “ringing” or “frequency-following response” (FFR) of the cochlea; Dau, 2003; VanRullen et al., 2014). Recently, we reported the construction of speech/noise stimuli without systematic fluctuations in amplitude and spectral content (or “spectral energy”), i.e. features processed at the cochlear level[2] (Fig. 1A, bottom; Fig. 1B, right; Zoefel and VanRullen, 2015a). Remaining features (including but not restricted to phonetic information) and intelligibility were conserved; we can therefore assume that neural entrainment in response to these constructed stimuli is produced at a level located beyond the earliest level of the auditory hierarchy.

In previous work, we reported that entrainment of EEG oscillations persists in the absence of systematic fluctuations in amplitude and spectral content; this entrainment, however, does not depend on linguistic features, as it is not disrupted when the speech/noise sound is reversed (Zoefel and VanRullen, 2015b; summarized in Zoefel and VanRullen, 2015c). Based on this finding, we hypothesized that we can find entrainment in response to our constructed stimuli in monkey A1 as well. Therefore, in addition to everyday speech, we presented these speech/noise sounds and measured the entrainment of neural oscillations to them. We were thus able to characterize neural entrainment to speech, both with (i.e. to everyday speech) and without slow spectral energy fluctuations (i.e. to our constructed speech/noise stimuli).

Materials and methods

Subjects

In the present study, we analyzed the electrophysiological data recorded during nine penetrations of area A1 of the auditory cortex of one female rhesus macaque weighing ~9 kg, who had been prepared surgically for chronic awake electrophysiological recordings. Before surgery, the animal was adapted to a custom-fitted primate chair and to the recording chamber. All procedures were approved in advance by the Animal Care and Use Committee of the Nathan Kline Institute.

Surgery

Preparation of the subject for chronic awake intracortical recording was performed using aseptic techniques, under general anesthesia, as described previously (Schroeder et al., 1998). The tissue overlying the calvarium was resected and appropriate portions of the cranium were removed. The neocortex and overlying dura were left intact. To provide access to the brain and to promote an orderly pattern of sampling across the surface of the auditory areas, plastic recording chambers (Crist Instrument) were positioned normal to the cortical surface of the superior temporal plane for orthogonal penetration of area A1, as determined by preimplant MRI. Together with socketed Plexiglas bars (to permit painless head restraint), they were secured to the skull with orthopedic screws and embedded in dental acrylic. A recovery time of 6 weeks was allowed before we began data collection.

Electrophysiology

During the experiments, the animal sat in a primate chair in a dark, isolated, electrically shielded, sound-attenuated chamber with head fixed in position, and was monitored with infrared cameras. Neuroelectric activity was obtained using linear array multicontact electrodes (23 contacts, 100 um intercontact spacing, Plexon). The multielectrodes were inserted acutely through guide tube grid inserts, lowered through the dura into the brain, and positioned such that the electrode channels would span all layers of the cortex, which was determined by inspecting the laminar response profile to binaural broadband noise bursts.

Neuroelectric signals were impedance matched with a preamplifier (10X gain, bandpass dc 10 kHz) situated on the electrode, and after further amplification (500X) they were recorded continuously with a 0.01–8000 Hz bandpass digitized with a sampling rate of 20 kHz and precision of 16 bits using custom-made software in Labview. The signal was split into the local field potential (LFP; 0.1–300 Hz) and multiunit activity (MUA; 300–5000 Hz) range by zero phase shift digital filtering. Signals were downsampled to 2000 Hz and LFP data were notch-filtered between 59 and 61 Hz to remove electrical noise (transition bandwidth 2 Hz). MUA data were also rectified to improve the estimation of firing of the local neuronal ensemble (Legatt et al., 1980). One-dimensional CSD profiles were calculated from LFP profiles using a three-point formula for the calculation of the second spatial derivative of voltage (Freeman and Nicholson, 1975). The advantage of CSD profiles is that they are less affected by volume conduction than the LFP, and they also provide a more direct index of the location, direction, and density of the net transmembrane current flow (Mitzdorf, 1985; Schroeder et al., 1998).

At the beginning of each experimental session, after refining the electrode position in the neocortex, we established the BF of the

---

[1] In the work by Kayser et al., 2009, the investigation of neural codes in monkey auditory cortex included the presentation of human speech, but results were generalized across a range of natural sounds.

[2] Note that our constructed stimuli differ from common amplitude-modulated (AM) or frequency-modulated (FM) stimuli in that the latter only control for changes in amplitude *or* frequency but not in both.