NeuroImage 85 (2014) 761–768

Contents lists available at ScienceDirect

NeuroImage

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

Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing

Keith B. Doelling [a], Luc H. Arnal [a], Oded Ghitza [c], David Poeppel [a,b,*]

[a] Department of Psychology, New York University, USA
[b] NYUAD Institute, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, UAE
[c] Department of Biomedical Engineering, Boston University, USA

ARTICLE INFO

Article history:
Accepted 7 June 2013
Available online 19 June 2013

Keywords:
Neural oscillation
Auditory cortex
Speech
Acoustic edge
MEG
Perceptual parsing

ABSTRACT

A growing body of research suggests that intrinsic neuronal slow (<10 Hz) oscillations in auditory cortex appear to track incoming speech and other spectro-temporally complex auditory signals. Within this framework, several recent studies have identified critical-band temporal envelopes as the specific acoustic feature being reflected by the phase of these oscillations. However, how this alignment between speech acoustics and neural oscillations might underpin intelligibility is unclear. Here we test the hypothesis that the 'sharpness' of temporal fluctuations in the critical band envelope acts as a temporal cue to speech syllabic rate, driving delta–theta rhythms to track the stimulus and facilitate intelligibility. We interpret our findings as evidence that sharp events in the stimulus cause cortical rhythms to re-align and parse the stimulus into syllable-sized chunks for further decoding. Using magnetoencephalographic recordings, we show that by removing temporal fluctuations that occur at the syllabic rate, envelope-tracking activity is reduced. By artificially reinstating these temporal fluctuations, envelope-tracking activity is regained. These changes in tracking correlate with intelligibility of the stimulus. Together, the results suggest that the sharpness of fluctuations in the stimulus, as reflected in the cochlear output, drive oscillatory activity to track and entrain to the stimulus, at its syllabic rate. This process likely facilitates parsing of the stimulus into meaningful chunks appropriate for subsequent decoding, enhancing perception and intelligibility.

Introduction

Because auditory signals unfold over time, at multiple scales, the process of decoding input sounds to link them to meaningful objects or concepts requires integrating sensory information over time. In speech perception, this temporal integration must occur in at least two (and arguably more) distinct timescales which relate to syllabic-level (~200 ms or ~5 Hz) and phonemic-level (~25 ms or ~40 Hz) information. Several models have suggested that this type of multi-time resolution analysis and integration could be performed in auditory cortex using neuronal oscillations – corresponding to these two temporal windows of integration (~5 Hz, theta; ~40 Hz, gamma) – to parse the sound input at these separate timescales (Ghitza, 2011; Poeppel, 2003). It is hypothesized, in particular, that the phase of the slow oscillation (nested with gamma) locks to the syllabic rhythm to optimally decode and integrate syllabic and phonemic speech features (Giraud and Poeppel, 2012).

In this magnetoencephalography (MEG) study, we focus on the role of the longer temporal window, most readily corresponding to delta–theta oscillations, to gain a better mechanistic understanding of how neuronal activity in this band might underpin auditory perception and speech comprehension. Recently, much research has focused on slow neural oscillations and their relationship to auditory stimuli (Cogan and Poeppel, 2011; Ding and Simon, 2009; Howard and Poeppel, 2010, 2012; Luo and Poeppel, 2007, 2012; Peelle et al., 2013). In addition, the relevance of low–modulation frequency oscillations to multi-sensory perception has been demonstrated, for example in naturalistic scenes or the well-studied cocktail party scenario (Kerlin et al., 2010; Luo et al., 2010; Zion Golumbic et al., 2013). There is an emerging consensus that the phase of slow oscillations precisely tracks the stimulus acoustics. However whether this stimulus–response alignment across time is necessary for speech comprehension remains debated (Howard and Poeppel, 2010; versus Luo and Poeppel, 2007; Peelle et al., 2013). One hypothesis is that cortical delta–theta oscillations track the critical band envelopes of the stimulus — a feature which carries crucial cues regarding segmental and syllabic information (Rosen, 1992)[1]. Despite the

Abbreviations: CALM, categorization and learning module; CACoh, cerebro-acoustic coherence; MEG, magnetoencephalography.

* Corresponding author at: Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA.
E-mail address: david.poeppel@nyu.edu (D. Poeppel).

[1] Note the distinction between the temporal amplitude envelope of the (full-band) stimulus, on the one hand, and the auditory critical band envelopes (i.e., at the cochlear output), on the other (Ghitza et al., 2013).

1053-8119/$ – see front matter
http://dx.doi.org/10.1016/j.neuroimage.2013.06.035

body of research showing this oscillation tracking the envelope, it remains unclear which aspects of the stimulus drive this response. One plausible hypothesis generated from the Giraud and Poeppel (2012) model suggests that it is the onsets of syllables that produce temporal fluctuations, which entrain slow neural oscillations at the syllabic rate. Here, we test this hypothesis by filtering these fluctuations in very particular ways and analyzing the effect on oscillatory entrainment. As such, the principal goal of this study is to understand more clearly the mechanisms of slow oscillation envelope tracking and, in particular, to uncover aspects in the temporal domain of the stimulus that drive this neuronal activity.

It has recently been demonstrated that theta envelope tracking of speech is enhanced by stimulus intelligibility (Peelle and Davis, 2012; Peelle et al., 2013), while earlier work showed similar neural phase-locking for sentences played backwards (no intelligibility) and forwards (Howard and Poeppel, 2010). Thus the question of whether the linguistic content of the stimuli induces a top-down 'amplification' of the oscillation-based envelope-tracking mechanism is debated. As a result, a secondary goal of this study is to investigate how envelope tracking relates to intelligibility and to understand its putative function in the broader context of speech perception.

This neurophysiological experiment builds on a recent behavioral study that manipulated the temporal acoustic features of speech to delineate the role of low frequency (syllabic) cues in speech intelligibility (Ghitza, 2012). Artificially removing exactly those temporal fluctuations in the critical band envelopes that relate to the syllabic rate (2–9 Hz) significantly reduces the intelligibility of the degraded speech. However, when brief noise bursts are added to the degraded stimulus precisely where the 'acoustic landmarks'[2] of the original *would* have been, the error rate drops by about 50%. The interpretation proposed to explain this psychophysical effect is that removing these cues disrupts the ability of cortical delta–theta oscillations to track the stimulus envelope. While removing slow fluctuations from the stimulus reduced intelligibility, reinstating temporal cues artificially by using transient edges at landmark positions enhanced intelligibility.

We hypothesize that temporal cues that reflect the syllabic rate are at the origin of the envelope-tracking phenomenon, which in turn constitutes a crucial condition for continuous speech to be intelligible. Specifically, we propose that acoustic landmarks entrain intrinsic cortical oscillations to permit the extraction of temporal primitives and subsequently finer grained speech features in a decoding stage. This quasi-periodicity generates the envelope tracking behavior, which could have the capacity to parse the stimulus into syllable-size representations.

Materials and methods

Participants

16 right-handed participants (9 females; mean age 23 years, range 18–31) took part in the experiment after providing informed consent and received compensation for their participation. Handedness was determined using the Edinburgh Handedness Inventory (Oldfield, 1971). All participants were self-reported as having normal hearing and no neurological deficits. One participant was removed because he did not input his behavioral ratings as instructed. Another was removed due to too much noise in the MEG data. Consequently, the data from a total of 14 participants were analyzed. The study was approved by the local Institutional Review Board (New York University's Committee on Activities Involving Human Subjects).

[2] By 'acoustic landmarks' we refer to vocalic landmarks, or glide landmarks, or acoustically abrupt landmarks (sometimes termed 'acoustic edges'). See Stevens (2002).

Stimuli

Twenty stimuli, spoken strings of seven digits, were chosen from a set initially used in a behavioral study (see Ghitza, 2012). These stimuli were filtered into sixteen critical bands logarithmically spaced between 230 and 3800 Hz; the Hilbert envelope of each was manipulated into one of five conditions (described below) and then combined with a noise carrier with bandwidth equal to that of the critical band before being linearly summed across critical bands (see Fig. 1A). Each stimulus was between 2 and 3 s in duration (sampling rate 11 kHz). The 100 stimuli were presented four times to each participant in pseudo-randomized order.

Envelope alterations

In the *Control* condition (Fig. 1B), each critical band envelope was low-pass filtered at 10 Hz. These stimuli are an adaptation of stimuli used by Drullman et al. (1994) and are known to be highly intelligible. The *No*θ condition (Fig. 1C) consisted of critical band envelopes with a band-stop filter from 2 to 9 Hz. Effectively, this removes all temporal cues in the envelope that relate to the syllabic rate of the stimulus. In the *Ch*θ condition (Fig. 1D), the peaks in each critical band envelope are replaced with peaks of uniform height and shape. This, essentially, distils the stimulus down to only the temporal cues relating to syllabic rate. It removes most acoustic–phonetic information and leaves only information pertaining to the peak amplitude of each syllable in each critical band. The *No*θ + *Ch*θ condition is the linear sum of the Noθ and Chθ conditions creating a stimulus in which the natural syllabic temporal fluctuations in the Control condition are replaced by the artificial Chθ. The *No*θ + *Glb*θ is the same as the Noθ + Chθ conditions save for the Chθ peak picking operation which is done on the whole broadband envelope. This is an extreme version of the Noθ + Chθ removing all acoustic phonetic information and leaving only a noise burst at the peak amplitude of each syllable.

Task

The stimuli were delivered diotically via MEG-compatible tubephones (E-A-RTONE 3A 50 Ω, Etymotic Research) attached to E-A-RLINK foam plugs inserted into the ear canal and presented at normal conversational sound levels (~72 dB SPL).

For each trial, participants listened to one stimulus and were asked to rate the stimulus in terms of its intelligibility on a scale from 1 (poor) to 3 (good). In the original behavioral study, participants were asked to repeat the last four digits of each stimulus. This was not so viable in the MEG setup as the head movements associated with speech production can create noise as well as change the orientation of the participants' head during the experiment. Trials for each condition were randomly interleaved and each stimulus was presented four times. Mean Intertrial Interval (ITI) was 1 s with a standard deviation of .3 s. Scores for each stimulus are averaged across repetitions per subject. The aim of the psychophysics was to collect behavioral data during scanning that were compatible and comparable to the data published by Ghitza (2012).

Recording

MEG recording

Neuromagnetic signals were measured using a 157-channel whole-head axial gradiometer system (KIT, Kanazawa Institute of Technology, Japan). Five electromagnetic coils were attached to a participant's head to monitor head position during MEG recording. The locations of the coils were determined with respect to three anatomical landmarks (nasion, left and right preauricular points) on