Hearing Research 348 (2017) 70–77

Contents lists available at ScienceDirect

Hearing Research

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

Research Paper

Indexing cortical entrainment to natural speech at the phonemic level: Methodological considerations for applied research

Giovanni M. Di Liberto [a, *], Edmund C. Lalor [a, b, **]

[a] Trinity College Institute of Neuroscience, School of Engineering, and Trinity Centre for Bioengineering, Trinity College Dublin, Dublin 2, Ireland
[b] Department of Biomedical Engineering and Department of Neuroscience, University of Rochester, Rochester, NY, USA

ARTICLE INFO

Article history:
Received 10 October 2016
Received in revised form
15 February 2017
Accepted 17 February 2017
Available online 27 February 2017

Keywords:
Language impairment
Continuous speech
EEG
Categorical perception
Clinical research
Neuromarker

ABSTRACT

Speech is central to human life. As such, any delay or impairment in receptive speech processing can have a profoundly negative impact on the social and professional life of a person. Thus, being able to assess the integrity of speech processing in different populations is an important goal. Current standardized assessment is mostly based on psychometric measures that do not capture the full extent of a person's speech processing abilities and that are difficult to administer in some subjects groups. A potential alternative to these tests would be to derive "direct", objective measures of speech processing from cortical activity. One such approach was recently introduced and showed that it is possible to use electroencephalography (EEG) to index cortical processing at the level of phonemes from responses to continuous natural speech. However, a large amount of data was required for such analyses. This limits the usefulness of this approach for assessing speech processing in particular cohorts for whom data collection is difficult. Here, we used EEG data from 10 subjects to assess whether measures reflecting phoneme-level processing could be reliably obtained using only 10 min of recording time from each subject. This was done successfully using a generic modeling approach wherein the data from a training group composed of 9 subjects were combined to derive robust predictions of the EEG signal for new subjects. This allowed the derivation of indices of cortical activity at the level of phonemes and the disambiguation of responses to specific phonetic features (e.g., stop, plosive, and nasal consonants) with limited data. This objective approach has the potential to complement psychometric measures of speech processing in a wide variety of subjects.

1. Introduction

Speech is central to human life. Over the past 30 years, neuroscience has provided tremendous insights into the neurobiology of language using brain imaging. As a result, it is now generally understood that speech is processed in a hierarchically organized system of functionally distinct cortical areas (Hickok and Poeppel, 2007; Okada et al., 2010; Peelle et al., 2010; Poeppel, 2014). However, much work remains to be done to elucidate the details of this system, particularly in the context of natural speech. This is an important issue in and of itself. However, it is also important in that there are significant numbers of people worldwide who suffer from some form of speech and language impairments. These can arise as a consequence of developmental disorders (Leonard, 2014) or from a decline in related cortical functions (e.g., through ageing, psychosis, injury; (Kemper and Anagnopoulos, 2008; Mesulam et al., 2014; Ross et al., 2007)). A better understanding of the underlying speech processing network and an ability to identify specific impairment within that network are crucial to developing clinically useful assessments of speech and language in these populations.

Speech and language impairment can disrupt the ability to understand auditory speech and efficiently communicate in a number of ways, which correspond to different symptoms. In this context, standardized assessment of such impairments is usually pursued using a number of behavioral tests (e.g., non-verbal hearing, speech, and language tests; standardized test of intelligence) (Ford and Dahinten, 2005; Gardner et al., 2006; Tomblin et al., 1996). However, these measures are inadequate at capturing the full extent of a person's impairment and should be considered only as one aspect of a comprehensive assessment

* Corresponding author. Trinity Centre for Bioengineering, Trinity College Dublin, Dublin 2, Ireland.
** Corresponding author. Department of Biomedical Engineering and Department of Neuroscience, University of Rochester, Rochester, NY, USA.
E-mail addresses: diliberg@tcd.ie (G.M. Di Liberto), edlalor@tcd.ie (E.C. Lalor).

http://dx.doi.org/10.1016/j.heares.2017.02.015
0378-5955/

process (Flanagan et al., 1997; Mody and Belliveau, 2013). Furthermore, some of these measures cannot be derived for some groups such as infants or participants with reading impairment or no reading skills.

A complementary approach is to "directly" investigate the causes that underpin such conditions, rather than evaluate "indirect" effects on specific behavioral markers. In this sense, neuroimaging provides an opportunity to derive measures directly related to the cortical processing of speech in the human brain. In particular, noninvasive, safe, functional brain measurements (EEG, MEG, fMRI, NIRS) have now been proven feasible for use with both children (starting at birth) and adults (Aslin and Mehler, 2005; Kuhl, 2010; Kuhl et al., 2005; McNealy et al., 2006). Neuroimaging research in speech perception has traditionally focused on neural activation patterns corresponding to the perception of minimal linguistic contrasts (e.g., how we distinguish "cat" from "mat") (Obleser et al., 2007; Peter et al., 2016; Salmelin, 2007), cortical responses cases of syntactic or semantic violation (Kutas and Hillyard, 1980; Lau et al., 2008), and the processing of the low-level acoustics of an incoming sound stimulus (Lakatos et al., 2005; Overath et al., 2008). However, the study of speech comprehension needs to account for how humans process continuous natural speech, which is a task performed efficiently by healthy people in their everyday life and is profoundly different from, for example, the perception of isolated syllables (Bonte et al., 2006).

Recent studies showed an innovative way to investigate continuous speech perception in humans, by indexing how cortical activity tracks the dynamics of that speech. This phenomenon of cortical entrainment has been demonstrated in humans for the amplitude envelope of speech using magnetoencephalography (MEG; Ahissar et al., 2001; Luo and Poeppel, 2007), electroencephalography (EEG; Aiken and Picton, 2008; Lalor and Foxe, 2010), and electrocorticography (ECoG; Nourski et al., 2009). And the effect has been quantified using a cross-correlation analysis between the speech envelope and the recorded neural data (Ahissar et al., 2001; Abrams et al., 2008; Nourski et al., 2009; Millman et al., 2015). However, this approach is ill-suited to the study of naturalistic stimuli (Crosse et al., 2016). The reason for this is that naturalistic stimuli vary in a non-random way and so such stimuli are correlated with time-shifted versions of themselves. This leads to temporal smearing when cross-correlating the stimulus with shifted versions of the response. For this reason, system identification methods based on ridge regression have been recently applied in this context, and were shown to be effective for investigating the cortical processing of natural speech (Machens et al., 2004; Crosse et al., 2016). And, in turn, the ability to use more natural stimuli facilitates the design of more engaging paradigms. These issues, and others, are discussed in several recent reviews on the approaches for and applications of the speech-entrainment phenomenon (Ding and Simon, 2014; Wöstmann et al., 2016; Crosse et al., 2016).

In this context, a recent study (Di Liberto et al., 2015) introduced a framework for disentangling phoneme-level cortical responses from cortical activity elicited by low-level acoustics. Results from this study indicated that low-frequency cortical entrainment to speech features reflects more than a simple acoustic analysis of the stimulus, and that it also reflects phoneme-level processing. Therefore, this framework provides a potential methodology for investigating speech encoding under a variety of conditions and in a variety of cohorts. This could include research on the causes of speech impairments in particular cohorts by deriving direct indices of cortical activity at specific levels of the speech processing hierarchy using non-invasive EEG. However, short experimental times are preferable in applied research (Mirkovic et al., 2015), whereas Di Liberto et al., 2015 used a recording time of 72 min per subject, which may constitute an obstacle when studying particular cohorts (e.g., young children).

Here we introduce a modification to our previously introduced framework that allows for a significant reduction of the experimental time needed to derive such indices of phoneme-level cortical entrainment. Our previous framework involved relating different representations of a speech signal to ongoing EEG. In particular, it involved building a model for each subject that would map a specific speech representation to that subject's own EEG signal. This type of approach has previously been used to "decode" how attention is being deployed in so-called cocktail party environments (O'Sullivan et al., 2015; Mirkovic et al., 2015). The modification we make here follows innovation introduced in these attention decoding studies (O'Sullivan et al., 2015; Mirkovic et al., 2015). Specifically, these authors showed that it was possible to decode attention for an individual subject using a generic model that was built from the data from other subjects. This led to a large reduction in how much data was needed from each subject to perform decoding (Mirkovic et al., 2015). Here, we seek to do something similar in the context of our approach for assessing phoneme-level speech processing. While it is known that not many electrodes are needed for this approach to be effective (by construction; see forward modeling approach in Crosse et al., 2016), the ability to use the framework with small amounts of data from individual subjects is uncertain. To clarify this issue, an extensive analysis was conducted to assess the minimum experimental time needed to detect meaningful cortical responses. The goal of the analysis was to show that it is possible to utilize short data sets across multiple subjects to make inferences about speech processing in individual subjects. Specifically, we aimed to show that we can robustly index phoneme-level processing in the context of natural speech in cases of limited amounts of experimental data.

2. Material and methods

Ten healthy subjects (7 male) aged between 23 and 38 years old participated in the experiment. The study was undertaken in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the School of Psychology at Trinity College Dublin. Each subject provided written informed consent. Subjects reported no history of hearing impairment or neurological disorder.

2.1. Stimuli and experimental procedure

Subjects undertook 28 trials, each of ~155 s in length, where they were presented with an audiobook version of a classic work of fiction read by a male American English speaker. The trials preserved the storyline, with neither repetitions nor discontinuities. All stimuli were presented monophonically at a sampling rate of 44,100 Hz using Sennheiser HD650 headphones and Presentation software from Neurobehavioral Systems (http://www.neurobs.com). Testing was carried out in a dark room and subjects were instructed to maintain visual fixation for the duration of each trial on a crosshair centered on the screen, and to minimize eye blinking and all other motor activities.

2.2. Data acquisition and preprocessing

Electroencephalographic (EEG) data were recorded from 128 scalp electrodes (plus 2 mastoid channels). Data were filtered over the range 0–134 Hz, and digitized with a sampling frequency of 512 Hz using a BioSemi Active Two system. Data were analyzed offline using MATLAB software (The Mathworks Inc.). EEG data were digitally filtered between 1 and 8 Hz using a Chebyshev Type 2 zero-phase filter. In order to reduce the processing time, all EEG