Brain and Language 184 (2018) 32–42

Contents lists available at ScienceDirect

Brain and Language

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

The time-course of cortical responses to speech revealed by fast optical imaging

Joseph C. Toscano[a,b,*], Nathaniel D. Anderson[b,c], Monica Fabiani[b,c], Gabriele Gratton[b,c], Susan M. Garnsey[b,c]

[a] Department of Psychological & Brain Sciences, Villanova University, United States
[b] Beckman Institute for Advanced Science & Technology, University of Illinois at Urbana-Champaign, United States
[c] Department of Psychology, University of Illinois at Urbana-Champaign, United States

ARTICLE INFO

Keywords:
Speech perception
Phonological categorization
Spoken language processing
Optical imaging
Event-related potentials

ABSTRACT

Recent work has sought to describe the time-course of spoken word recognition, from initial acoustic cue encoding through lexical activation, and identify cortical areas involved in each stage of analysis. However, existing methods are limited in either temporal or spatial resolution, and as a result, have only provided partial answers to the question of how listeners encode acoustic information in speech. We present data from an experiment using a novel neuroimaging method, fast optical imaging, to directly assess the time-course of speech perception, providing non-invasive measurement of speech sound representations, localized to specific cortical areas. We find that listeners encode speech in terms of continuous acoustic cues at early stages of processing (ca. 96 ms post-stimulus onset), and begin activating phonological category representations rapidly (ca. 144 ms post-stimulus). Moreover, cue-based representations are widespread in the brain and overlap in time with graded category-based representations, suggesting that spoken word recognition involves simultaneous activation of both continuous acoustic cues and phonological categories.

1. Introduction

A long-standing issue in language processing concerns the nature of representations used by the brain to perceive speech. Debate continues about whether perception is based on continuous acoustic cues (Pisoni & Tash, 1974; Toscano, McMurray, Dennhardt, & Luck, 2010), discrete phonemes (Liberman, Harris, Hoffman, & Griffith, 1957; Chang et al., 2010), or other representations (e.g., auditory contrasts, Diehl, Lotto, & Holt, 2004; articulatory gestures, Viswanathan, Fowler, & Magnuson, 2009). For example, in English, the discrete phonemic categories of /b/ and /p/ are distinguished by several continuous acoustic cues, including voice onset time (VOT[1]; Lisker & Abramson, 1964; Fig. 1A). In order to accurately recognize speech, listeners must map these cues from the speech signal onto phoneme categories. The question addressed here concerns when and where cue- and category-level representations are each used during spoken language processing.

This issue is central to an early hypothesis about speech perception, known as *categorical perception* (Liberman et al., 1957)—the idea that listeners only perceive speech in terms of discrete units, assessed behaviorally by showing that listeners' category boundaries along an acoustic continuum (obtained in an identification/labeling task) align with the peak of their discrimination function for those sounds. Categorical perception was proposed, in part, to explain human listeners' remarkable ability to recognize speech sounds, which seems to differ from their ability to recognize other sounds and from the auditory abilities of non-human animals. However, the debate over the original categorical perception hypothesis is largely settled: Listeners are indeed sensitive to graded differences within phonetic categories, contrary to early proposals (Pisoni & Tash, 1974; Massaro & Cohen, 1983; Miller, 1997; McMurray, Tanenhaus, & Aslin, 2002; Toscano et al., 2010). Moreover, sensitivity to continuous cues is critical for accurate speech perception, as it allows listeners to overcome contextual variability (McMurray & Jongman, 2011). Thus, at minimum, within-category phoneme differences are preserved in the brain in some form. However, it is unclear when during processing phonological categories play a role, and whether such representations coexist with cue-level representations. Do listeners represent speech in terms of acoustic cues early in perception? Or are sounds immediately encoded as categories (either graded or discrete)? Answering these questions is critical not only for our understanding of the mechanisms underlying language

* Corresponding author at: Department of Psychological & Brain Sciences, Villanova University, 800 E Lancaster Ave, Villanova, PA 19085, United States.
E-mail address: joseph.toscano@villanova.edu (J.C. Toscano).

[1] VOT is defined as the time difference between the release of consonantal closure and the onset of laryngeal voicing for word-initial stop consonants (/b,d,g,p,t,k/).

https://doi.org/10.1016/j.bandl.2018.06.006
Received 28 September 2017; Received in revised form 3 April 2018; Accepted 12 June 2018
0093-934X/ © 2018 Elsevier Inc. All rights reserved.

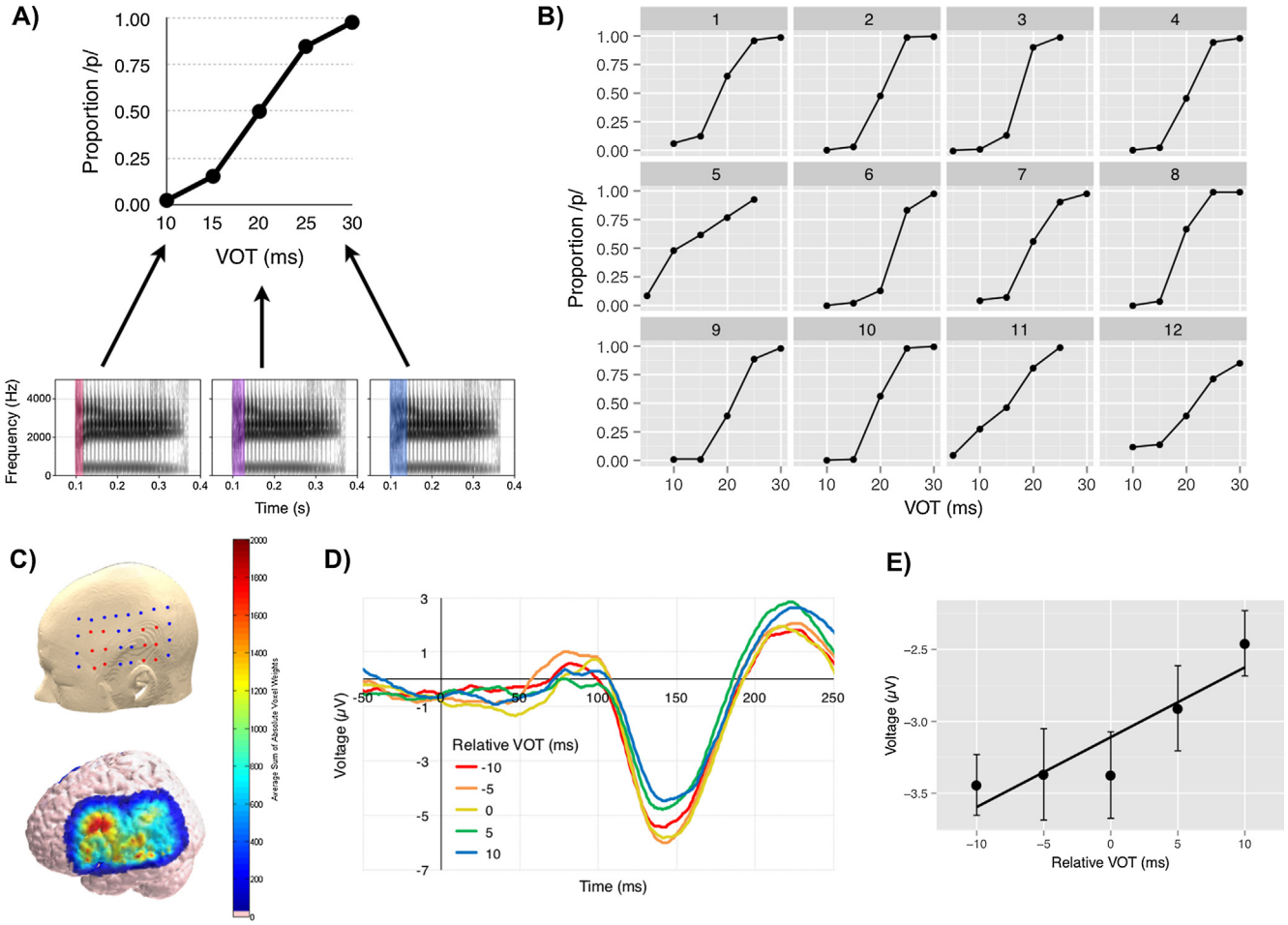

Fig. 1. (A) Hypothetical response function that would be obtained from a human listener when presenting sounds varying in VOT. As VOT values increase, listeners are more likely to report a voiceless (/p/) percept. Spectrograms show onsets of stimuli used in the experiment, with the period of aspiration at onset (which determines the VOT) highlighted. (B) The actual behavioral response functions obtained from listeners in the experiment. Note that, for each listener, a pre-test was conducted so that stimuli were centered on their category boundary (50% point) ± 10 ms VOT. (C) Locations of infrared sources (red) and detectors (blue) on the scalp (top), and heat map showing average cortical coverage obtained from the montage across the participants in the experiment (bottom). Hot colors represent areas for which the optical montage provided greatest sensitivity (i.e., the greatest number of overlapping paths from source-detector pairs). The montage provided good coverage over the ROIs used in the study. (D) Grand average ERP waveforms at Fz as a function of VOT relative to each listeners' category boundary (relative VOT). (E) Mean N1 amplitude (mean voltage at Fz from 100 to 150 ms post-stimulus, chosen to capture the N1 and minimize overlap with the P2; cf. Toscano et al., 2010) as a function of relative VOT. As VOT increased, N1 amplitude decreased. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

comprehension, but also for the development of computer-based speech recognition systems that mimic processes used by humans (Scharenborg, 2007) and for evaluating neurobiological models of spoken language processing (Hickok & Poeppel, 2007; Scott & Johnsrude, 2003).

2. Cue- vs. category-based representations

Behavioral experiments have been unable to determine whether the brain encodes acoustic cues independently of phonemes, since behavioral responses reflect both early perceptual processing and later categorization stages. More importantly, they provide only indirect information about the time-course of speech perception. If the brain encodes continuous cues, how and when are these representations used in identifying phonemes? Are both types of representations maintained in different cortical areas simultaneously? These questions relate to a larger debate about whether language processing is inherently a serial or a parallel process (Trueswell, Tanenhaus, & Garnsey, 1994), and distinguishing these models requires us to examine how speech sounds are represented *during* perception.

Recently, speech researchers have turned to neurophysiological measures, primarily using functional MRI (fMRI) and event-related potential (ERP) techniques, and several neurobiological models of speech perception have been based on such data. However, both techniques have inherent limitations that prevent us from simultaneously measuring the early time-course of speech perception and localizing the brain regions involved. fMRI can identify brain areas that are engaged when processing lexical and phonological information, as well as sub-phonemic differences (Myers, Blumstein, Walsh, & Eliassen, 2009; Blumstein, Myers, & Rissman, 2005), but because it relies on hemodynamic responses, this approach has limited temporal resolution, making it impossible to distinguish early perceptual encoding from later-occurring processes.

In contrast, the ERP technique has excellent temporal resolution, allowing us to demonstrate that early sensory components such as the auditory N1 vary with continuous acoustic cues but not phonological categories (Toscano et al., 2010). However, it is unclear where these responses are generated in the brain (Picton, Hillyard, Krausz, & Galambos, 1974; Giard et al., 1994). It is also difficult to separate contributions of multiple sources (which may encode different types of information) in scalp-recorded EEG, and therefore, to determine whether different types of representations may simultaneously co-exist in the brain.

Download English Version:

https://daneshyari.com/en/article/7283324

Download Persian Version:

https://daneshyari.com/article/7283324

Daneshyari.com