Clinical Neurophysiology 124 (2013) 1204–1215

Contents lists available at SciVerse ScienceDirect

Clinical Neurophysiology

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

Auditory cortical activity in normal hearing subjects to consonant vowels presented in quiet and in noise

Andrew Dimitrijevic [a,*], Hillel Pratt [c], Arnold Starr [b]

[a] Communication Sciences Research Center, Cincinnati Children's Hospital Medical Center, Department of Otolaryngology, Head and Neck Surgery, University of Cincinnati, USA
[b] Department of Neurology, University of California, Irvine, USA
[c] Evoked Potentials Laboratory, Technion – Israel Institute of Technology, Haifa, Israel

ARTICLE INFO

Article history:
Accepted 28 November 2012
Available online 29 December 2012

Keywords:
Evoked potentials
Speech perception
Voice onset time
Temporal processing
CLARA (Classical LORETA recursively applied)
Source analysis

HIGHLIGHTS

- Cortical evoked potentials (N100) to speech can be recorded as "onset responses" when the consonant-vowel is in first position and as "change responses" when it follows a vowel.
- N100 latency was prolonged in consonant-vowels with long voice onset times only in change responses whereas only N100 amplitude was affected by voice onset time in onset responses.
- Source estimations indicated similar regions of cortical activation during N100 of change and onset responses to speech, whereas the sources of the subsequent slow cortical potential changed across time.

ABSTRACT

Objective: Compare brain potentials to consonant vowels (CVs) as a function of both voice onset times (VOTs) and consonant position; initial (CV) versus second (VCV).
Methods: Auditory cortical potentials (N100, P200, N200, and a late slow negativity, (SN) were recorded from scalp electrodes in twelve normal hearing subjects to consonant vowels in initial position (CVs: /du/ and /tu/), in second position (VCVs: /udu/ and /utu/), and to vowels alone (V: /u/) and paired (VVs: /uu/) separated in time to simulate consonant voice onset times (VOTs).
Results: CVs evoked "acoustic onset" N100s of similar latency but larger amplitudes to /du/ than /tu/. CVs preceded by a vowel (VCVs) evoked "acoustic change" N100s with longer latencies to /utu/ than /udu/. Their absolute latency difference was less than the corresponding VOT difference. The SN following N100 to VCVs was larger to /utu/ than /udu/. Paired vowels (/uu/) separated by intervals corresponding to consonant VOTs evoked N100s with latency differences equal to the simulated VOT differences and SNs of similar amplitudes. Noise masking resulted in VCV N100 latency differences that were now equal to consonant VOT differences. Brain activations by CVs, VCVs, and VVs were maximal in right temporal lobe.
Conclusion: Auditory cortical activities to CVs are sensitive to: (1) position of the CV in the utterance; (2) VOTs of consonants; and (3) noise masking.
Significance: VOTs of stop consonants affect auditory cortical activities differently as a function of the position of the consonant in the utterance.

1. Introduction

Stop consonant-vowel (CV) utterances (e.g., /du/ and /tu/), are distinguished by acoustic differences of the time between onset of the consonant ("stop release burst") and onset of the following vowel ("voicing"). The difference in ms between the onsets of the consonant and the vowel is known as "voice onset time" or VOT (Lisker and Abramson, 1964). The consonant /d/ is identified when VOTs are relatively brief (<25 ms in English) whereas /t/ is identified when VOTs are relatively long (>45 ms in English). Consequently, discrimination between boundaries (i.e., /tu/ vs. /du/) is better than discrimination within boundaries ("short" /tu/ vs. "long" /tu"). VOTs that fall within these boundaries, are more variably classified as /t/ or /d/ when compared to short or long duration VOTs. The temporal boundaries described above are typical for English speakers (Lisker and Abramson, 1964) but can differ depending on the native language of the subject (e.g., Williams, 1977 for Spanish; Laufer, 1998 and Horev et al., 2007 for Hebrew).

* Corresponding author. Tel.: +1 513 636 3469.
E-mail address: andrew.dimitrijevic@cchmc.org (A. Dimitrijevic).

1388-2457/$36.00 © 2012 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.clinph.2012.11.014

There has been considerable interest in characterizing the central auditory processes underlying categorical characterization of stop consonants. Sharma and Dorman (1999, 2002) showed that CV syllables with short VOTs evoked a single N100 response to vowel-onset while syllables with long VOTs evoked a double N100 response: the first peak coincided with the onset of the consonant and the second with the onset of the vowel (Sharma and Dorman, 1999, 2000). However, these N100 changes were not systematically related to the temporal boundaries of category specific perception of stop consonants (Sharma and Dorman, 2000; Sharma et al., 2000; Eggermont and Ponton, 2002). Using synthetic speech stimuli, longer VOTs have resulted in delayed N100s (Frye et al., 2007; Tremblay et al., 2003). N100 amplitudes have also been described as being larger with short VOTs compared to long VOTs (Toscano et al., 2010) without any accompanying changes of latencies (Tremblay et al., 2003b). Toscano's study demonstrated that subjects engaged in an active task of classifying words based on VOT differences of the initial stop consonant (e.g., "beach" or "peach") showed change in amplitude of both N100 (sensitive to "sensory" features of the stimulus) and P300 (sensitive to "perceptual" features of stimulus) with changes of VOTs of the target stop consonant. Interestingly, there were no abrupt changes of amplitudes of either N100 or P300 corresponding to category boundaries.

Understanding the normal neural processing of speech, and particularly the role of temporal cues such as VOT, would provide insight into potential rehabilitative strategies of clinical populations with temporal processing disorders such as in auditory neuropathy (AN). AN is a disorder of speech comprehension with impaired auditory nerve function (AN; Starr et al., 1996). The magnitude of their speech comprehension deficits is larger than expected from the patients' degree of audibility changes, and reflects impaired processing of auditory temporal cues (Starr et al., 1996; Zeng et al., 2005). Specific etiologies for the disorder affect both pre-synaptic sites (e.g., impaired function of ribbon synapses of inner hair cells; Varga et al., 2006; Marlin et al., 2010) and post-synaptic sites (e.g., impaired function of auditory nerve fibers; Starr et al., 2003). Detailed analyses of the word processing errors in a group of AN subjects with Freidreich's ataxia, a post-synaptic disorder of neural transmission, showed abnormal identification of stop consonants distinguished by VOT (/tu//du/, /ba/ /pa/, /ka/ /ga/) but normal classification of sibilant consonants containing high frequencies (e.g., /s/ vs. /f/). In contrast, patients with a sensory hearing loss typically have an opposite pattern of speech comprehension deficits with impaired discriminations of high frequency fricatives and normal discrimination of stop consonants (Rance et al., 2008).

In this paper we report results of scalp-recorded brain potentials from normal-hearing subjects to spoken stop consonant vowel combinations presented with the consonant at the initial position of the phoneme (CV: /du/ and /tu/), or at second position following an initial vowel (VCV: /udu/ and /utu/). Acoustic features located at the initial position of a stimulus sequence evoke "acoustic onset responses" (N100/P200) of large amplitude and short latency. The same acoustic features located at later positions of the stimulus sequence evoke "acoustic change responses" (N100) that are both reduced in amplitude and delayed in latency compared to these measures in initial position (Hari et al., 1988; Jones and Perez, 2001; Martin and Boothroyd, 1999, 2000). The present study examined the effect of consonant position in the utterance on auditory cortical potentials. We measured latencies and amplitudes of scalp potentials, their global field power (GFP), and estimated their brain sources as a function of: (1) consonant position in the phoneme; (2) VOTs; and (3) noise masking. We hypothesized that auditory cortical activities would show amplitude and latency differences when comparing responses to CVs with different VOTs, in initial and second positions in the utterance (onset and change responses). Furthermore, we expected noise masking to alter the relative contribution of vowel and consonant acoustic cues to the evoked brain activities.

2. Methods

2.1. Subjects

Twelve (4 males, 8 females) subjects (mean age: 20 years, all self-reported right-handed) with normal pure tone thresholds (500–8000 Hz) and no history of neurological illnesses participated in the study. All subjects were tested in quiet and 6 of these subjects were tested with noise masking as well. All subjects gave informed consent prior to testing. All except one subject were tested with left ear stimulus presentation.

2.2. Stimuli

Speech sounds were recorded by a native English speaking male. Speech utterances included /u/, /tu/, /du/, /utu/, /udu/ and /uu/ (see Fig. 1). The /uu/ included two stimulus conditions (/uu_{15ms}/, /uu_{110ms}/) with temporal separations between the vowels that corresponded to the VOT separations in /udu/ and /utu/. The /uu/ conditions allowed the examination of brain activity to two vowels without an intervening stop consonant. The comparison of /tu/-/du/ and /utu/-/udu/ was used to test effects of consonant position in the bisyllabic utterances by evoking "acoustic onset responses" (/tu/ and /du/) and "acoustic change responses" (/udu/ and /utu/). The stimuli were created by manipulating a real speech recording of /u/ and /tu/ in CoolEdit. In this talker, a naturally produced /tu/ had a 110 ms noise burst (aspiration) before the vowel (periodic voicing) began; this condition will be referred to as the 110 ms voice onset time (VOT) condition. Removing the first 85 ms of aspiration in the /tu/ changed the perception from /tu/ to /du/; in this condition (15 ms VOT), there were only 15 ms of aspiration left from the initial 110 ms before the vowel began. The total duration of the /u/ stimulus was 160 ms, the /tu/ stimulus was 380 ms (110 ms VOT) and 285 ms for the /du/ stimulus (15 ms VOT). The /utu/ and /udu/ stimuli were created by concatenating the /tu/ or /du/ stimulus at the offset of the /u/ with a 5 ms gap. This gap was chosen as it approximates a "natural" sounding /utu/ or /udu/. The /uu/ condition was created by concatenating a more intense (6 dB) and time shifted version of the initial /u/ with an RMS value for the second /u/ equivalent to that of the /u/ in /tu/. Their acoustic waveforms are shown in Fig. 1. All speech stimuli were presented at 90 dB SPL. The rise time of /tu/ and /du/ was less than 5 ms; and the RMS of the initial 50 ms of /tu/ and /du/ was .0076 and .0040, respectively (arbitrary units of the sound file). Speech-shaped-noise maskers were continuously presented for the entire duration of the recording at 85 dB SPL. Both noise and speech sounds were calibrated using a continuous peak SPL measured on F setting. All sound stimuli were presented through Etymotic © ER3 insert earphones.

2.3. Categorical perception of CVs and VCVs

The identification of the CVs and VCVs was performed using a 2-alternative forced choice identification paradigm. CVs consisted of a /du/-/tu/ continuum with VOTs of 15, 25, 45, and 65 ms. Similarly, for VCVs an /udu/-/utu/ continuum with VOTs of 15, 25, 45, and 65 ms was used. After an initial training period, 20 tokens at each of the VOTs were randomly presented at 90 dB SPL. CVs and VCVs and the effect of speech-shaped noise masker were evaluated in separate runs. Subjects were instructed to indicate via button