BRAINSTEM ENCODING OF VOICE ONSET TIME: PRELIMINARY FINDINGS

Jinghua Ou & Alan C. L. Yu

University of Chicago
jou@uchicago.edu, aclyu@uchicago.edu

ABSTRACT

Stop voicing contrasts are very common in the world's languages and the perception of acoustic cues underlying these contrasts has been examined thoroughly with voice onset time (VOT) as the primary cue. While previous neurophysiological studies have revealed the cortical correlates of VOT, much less is known about how this critical voicing feature is encoded at the subcortical level. This study presents preliminary findings of auditory evoked potentials measuring activity from the brainstem of four neurotypical subjects to a VOT continuum (/ba/-/pa/). Spectrotemporal information corresponding to stop release and the onset of voicing was identified in the brainstem responses, and responses were compared across stimuli with respect to the duration between these two acoustic landmarks. Results show systematic changes in brainstem functions in response to the VOT alterations in the stimuli suggesting a high degree of neural synchrony to the timing of acoustic features at the brainstem level.

Keywords: Auditory brainstem response, VOT, Stop voicing, EEG.

1. INTRODUCTION

Consonant voicing contrasts are very common in the world's languages [1] and the perception of acoustic cues underlying these contrasts has been examined thoroughly with VOT as the primary cue [2]. Given its significant function in nearly all languages, it is important to clarify the neural encoding of this acoustic cue. VOT is defined as the duration between the release of the oral constriction and the onset of vocal-fold vibration [2]. Previous electrophysiological studies have demonstrated the neural representation of VOT using the cortical auditory evoked potentials [3,4,5]. The N1 component is shown to reflect the basic encoding of acoustic information in the central auditory cortex [6], and it is sensitive to fine temporal events [3,7]. Changes in latency and/or morphology of the N1 have been shown to coincide with a change in perception from voiced to voiceless stop initials, with short VOTs eliciting a single-peak N1, but long VOTs eliciting a double-peak N1 [3]. However, other studies have reported a single peak for N1 for both short and long VOTs, with its amplitude varying as a function of the VOT values [4,5]. The longer the VOT, the larger the N1 amplitude [5]. Despite the mixed findings, these studies overall suggest that the cortical potentials N1 is able to encode VOT information, including the morphology/amplitude changes corresponding to different VOTs.

Much less is known about how this critical cue is encoded at the brainstem level, however. Unlike their cortical counterparts, the auditory brainstem responses (ABR) closely mimics the spectrotemporal properties of the original auditory stimuli [8], to the extent that listeners can recognize words from the neural responses that have been converted into sound stimuli. One of most extensively studied speech sounds in ABR research is the syllable /da/ [9,10]. ABR to /da/ includes both transient and sustained response, with the transient onset response corresponding to the stop burst associated with /d/, and the sustained responses reflecting phase-locked neural activity (usually referred to as frequency-following response) to the periodic features contained in the vowel segment /a/, i.e. fundamental frequency (f0) and its harmonics [8]. Previous studies have mainly focused on ABR to alterations in spectral cues [10,11], and how ABR changes in response to temporal cues e.g., VOT remains largely unexplored. To the best of our knowledge, this study is among the first to explore brainstem encoding of VOT variation.

2. METHOD

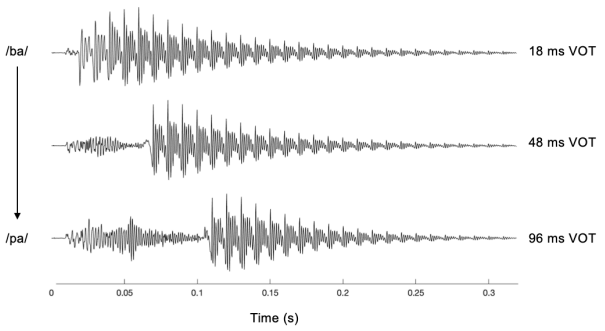
2.1. Participants

Four Cantonese-speaking young adults (age: 22.8 ± 2.1, males = 2) participated in the experiment. All reported normal hearing and, no previous history of neurological or psychiatric illness.

2.2. Stimuli

A /ba/-/pa/ continuum differing only in VOT was created using the Klatt-style synthesizer implemented in Praat 6.0.36 (Boersma and Weenik 2010). Tokens were designed to sound like a male talker. The script, based on a script developed by Jessamyn Schertz (https://goo.gl/27kDuT), created a /ba/-/pa/ continuum with the syllable duration held constant at 315 ms and with VOT increasing at 6 ms increments. We selected three steps from the continuum: VOT = 18 ms, VOT = 48 ms, and VOT = 96 ms, to elicit perception of canonical /ba/, ambiguous between /ba/ and /pa/ and canonical /pa/ (Figure 1).

Figure 1: Acoustic waveforms of each stimulus along the /ba/-/pa/ VOT continuum.



2.3. EEG recording procedures

During EEG recording, participants were encouraged to rest or fall asleep in a electromagnetically shielded booth. Stimuli were presented through insert earphones (ER-3a, Etymotic Research) to both ears at around 80 dB SPL, using Neuroscan Stim^2 (Compumedics). The order of the three stimuli were counterbalanced across participants. The inter-stimulus-interval jittered between 74 and 104 ms. Responses were collected using CURRY Scan 7 Neuroimaging Suite (Compumedics) with four Ag-AgCl scalp electrodes, differentially recorded from vertex (Cz, active) to bilateral linked mastoids (M1+ M2, references), with the forehead (Fpz) as ground. Contact impedance was less than 5 kΩ for all electrodes. For each stimulus, 2000 sweeps were collected at alternating polarity with a sampling rate of 20k Hz, lasting around 30 minutes.

Filtering, artifact rejection, and averaging were performed offline using CURRY 7. Responses were band-pass filtered from 80 to 1500 Hz, 12 dB/octave, and trials with activity greater than ± 35 μV rejected. Continuous EEG was segmented into 370 ms (50 ms pre-stimulus onset to 320 post stimulus onset).

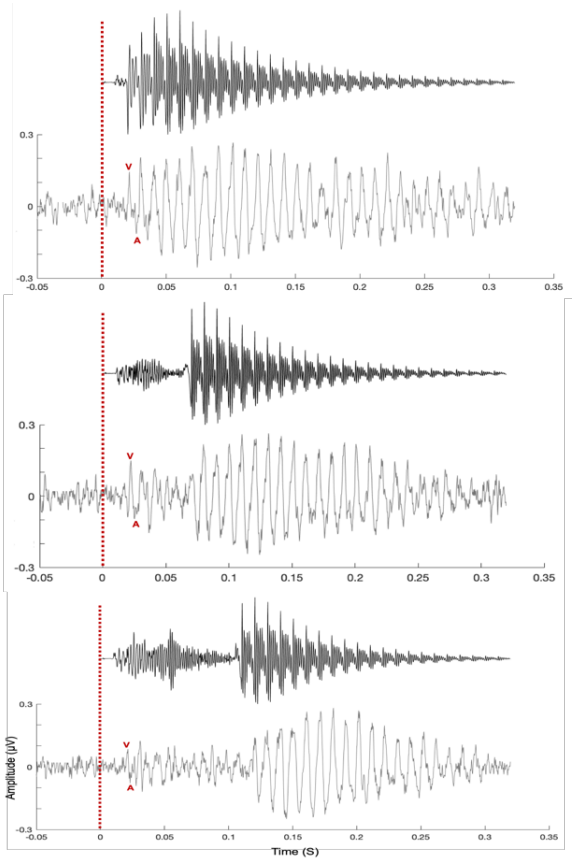
2.4. EEG analysis

Based on the definition of VOT - the duration between stop release and onset of voicing, we sought to identify peak of the onset response associated with stop burst, and the onset of the sustained responses associated with periodic voicing feature. The latency differences between these two onsets were then computed to reflect VOT encoding in ABR.

2.4.1. Transient onset response

Grand-average waveforms (averaging across sweeps and participants) for each stimulus were computed to identify the time lag between onset response and the corresponding stimulus landmark i.e. stop burst. Figure 2 illustrates the average ABR waveforms to each stimulus. Following the practice of Skoe and Kraus [8], the first sharp peak post stimulus onset is associated with the broadband stop burst, and identified as wave V-A. The V-A complex is analogous to the click-evoked wave V-Vn complex [8]. As can be observed in Fig. 2, peaks occur ~ 20 ms after the stop burst across three stimuli. Taking this time lag into account, we then performed visual inspection of the individual waveform from each subject to identify the latency of the onset response.

Figure 2: Waveforms of the grand-average ABRs of all subjects to each of the stimulus. V-A marks the peaks of onset responses. Stimulus onset (time = 0 ms) is indicated by the vertical dotted line.



2.4.2. Sustained response

We adopted two measures to detect the onset of the sustained responses in ABR, 1) cross-correlation between response and stimulus, 2) Spectral analysis - inter-trial-coherence (ITC) of the response.

Cross-correlation. Cross-correlation is a useful tool to compare the timing and morphology of two signals [12]. Signal 1 is shifted in time relative to signal 2 to find the shift that produces the strongest correlation. In the present case, the stimuli were resampled to match the sampling rate of the response prior to the analysis. The vowel segment of the stimulus was spliced out and used to correlate with the response. The vowel segment of the stimulus was

shifted in time to locate the point at which maximal correlation was reached. In this manner, the time point at which the response starts to follow the periodic segment of the stimulus i.e. vowel, can be determined.

Inter-trial-coherence (ITC). In addition to the temporal domain, we aimed to identify the sustained responses in the spectral domain. ITC reflects the consistency of spectral response across trials, i.e. responses that are phase-locked to the stimulus, with values ranging from 0 (no phase-locking) to 1 (perfect phase-locking) [13]. The single-trial EEG data were decomposed into time-frequency representations using ITC as implemented in EEGLAB [13]. This measure represents the relative change in EEG spectral power from baseline to stimulus period. For each time-frequency point in the stimulus window, the spectral power is divided by the average power in the baseline window and transformed into a dB value (10*log10 of the signal). Specifically, the single-trial data were convolved using Hanning window between 100 and 500 Hz (in steps of 5 Hz) from -50 to 320 ms, which resulted in a time by frequency matrix. Non-parametric bootstrap resampling was used to test for significant spectral change relative to stimulus onset. The onset of the significant ITC was then used to indicate the beginning of the sustained responses.

By adopting measures in both temporal and spectral domains, we hope to obtain converging results on the onset latency of the sustained responses.

3. RESULTS

3.1. Transient onset responses

A visual analysis of the first 50 ms post stimulus onset showed the peak latencies of the onset responses to each stimulus for each subject. Figure 3 demonstrates the ABR waveforms of each subject to each stimulus. As mentioned above, the time lag between stimulus and response onsets was ~ 20 ms, therefore onset peaks were identified by taking this lag into consideration. The first sharp peaks appeared 20 ms post stimulus onset were marked as V-A to indicate onset responses.

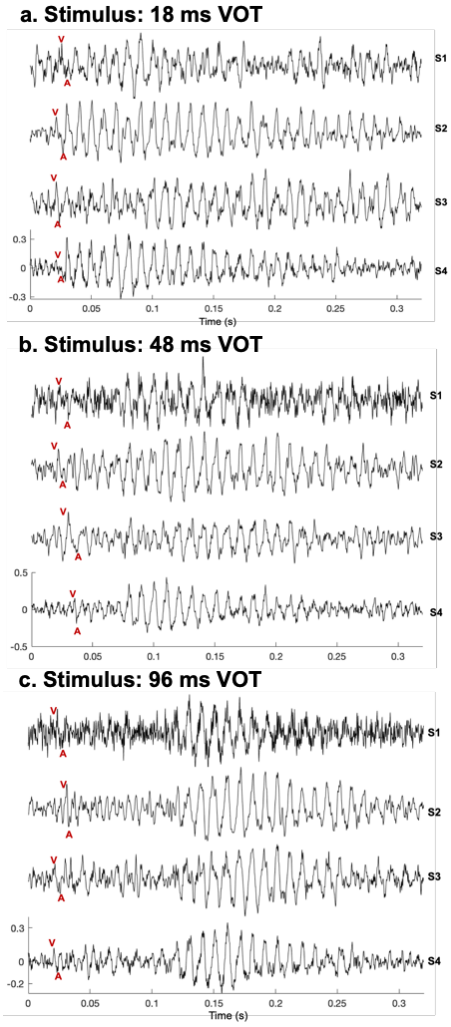


Figure 3: Waveforms of each subject to each stimulus condition a. 18 ms VOT, b. 48 ms VOT, and c. 96 ms VOT. V-A marks the peaks of onset responses.

3.2. Sustained responses

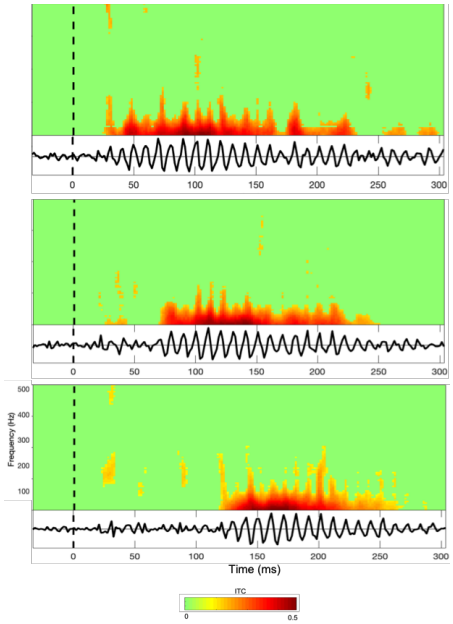
Results from cross-correlation between response and the vowel segment of the stimulus are summarised in Table 1. ITC analysis was also used to determine the onset of the sustained responses, and performed on ABR for each subject and for each stimulus condition. Significant ITC was tested with non-parametric permutation. Specifically, a surrogate data distribution was constructed by assigning each trial with spectral estimates randomly selected from the defined baseline (-50 to 0 ms), then averaged these spectral estimates across trials. This process was repeated 1000 times, and the resulting baseline ITC distribution was used to determine the statistical threshold ($p < .001$). By applying this significance threshold, ITC spectrograms can be presented with a statistical mask, by which non-significant spectral changes were set to 0 dB, as shown in Figure 4. Figure 4 illustrates the ITC spectrograms to grand-average ABR across all subjects to each stimulus. The onsets of the significant ITC clusters for each subject in each stimulus condition were summarized in Table 1.

Table 1. Onset of sustained responses (in ms, relative to stimulus onset) obtained from cross-correlation and ITC analysis.

		Cross-correlation	ITC
18 ms VOT	S1	40.8	51.6
	S2	21.6	24.2
	S3	96.4	88.5
	S4	26.4	25.8
	GA	36.2	36.0
48 ms VOT	S1	91.8	71.6
	S2	105.2	71.6
	S3	115.6	102.0
	S4	75.2	74.8
	GA	65.4	69.9
96 ms VOT	S1	126.4	118.9
	S2	152.0	122.3
	S3	156.0	145.9
	S4	131.6	113.8
	GA	115.0	118.0

Note: GA = grand-average ABR across four subjects.

Figure 4: ITC spectrograms of grand-average ABR across all subjects to stimulus of 18 ms VOT (upper), 48 ms VOT (middle), and 96 ms VOT (lower). Significant time-frequency regions are highlighted in red while non-significant spectral changes were set to 0 dB and masked in green. Dotted lines demarcate the onset of speech stimulus.



3.3. Brainstem encoding of VOT

To estimate how different VOTs were encoded in ABR, we computed the latency differences between the transient onset response (associated with stop burst) and the onset of the sustained response (associated with onset of voicing). Onsets of the sustained responses were obtained by averaging the latencies derived from cross-correlation and ITC measures (Table 1). The latency differences of grand-average ABR, and the individual ABR from each subject to each stimulus condition are summarized in Table 2.

Table 2. Summary of VOT encoding in ABR.

		Onset response (ms)	Onset of sustained response (ms)	Latency difference
18 ms VOT	S1	25.8	46.2	20.4
	S2	21.4	22.9	1.5
	S3	21.4	92.4	71
	S4	21.4	26.1	4.7
	GA	21.4	36.1	14.7
48 ms VOT	S1	23.2	81.7	58.5
	S2	22.0	88.4	66.4
	S3	22.6	108.8	86.2
	S4	22.2	75.2	52.9
	GA	22.4	67.6	45.2
96 ms VOT	S1	23.6	122.6	99.1
	S2	31.2	137.2	105.9
	S3	21.6	150.9	129.3
	S4	20.8	122.7	101.9
	GA	21.4	116.5	95.1

Note: GA = grand-average ABR across four subjects. Latency of each onset is relative to stimulus onset.

4. DISCUSSION

At the group level (GA in Table 2), the ABR showed great alignment to the acoustic landmarks that define VOT in the stimuli. That is, systematic changes were observed in brainstem functions in response to the VOT alterations in the stimuli, suggesting a high degree of neural synchrony to the timing of acoustic features at the brainstem level.

Despite this one-to-one correspondence at the group level, we do observe great individual variability in terms of how each subject encode different VOT values. Specifically, subject #4 seemed to be best in synchronizing brainstem responses to the VOT cues, whereas subject #3 appeared to have relatively poorer brainstem encoding. The brainstem encoding of the other two subjects were in between. Though the sample size is very small and quantitative comparison cannot be provided, the present findings nonetheless provide promising grounds for future study to further examine the brainstem encoding of VOT and its relationship with higher level processes. Brainstem encoding has been shown to associate with higher level functions such as speech perception [14], reading ability [15], and cognitive performance [16]. These findings suggest that brainstem encoding reflects intrinsic neurophysiology, and is an important marker for neurophysiological function. Future studies should report both group level as well as individual-level results to show how good encoders and poor encoders differ as a function of their higher perceptual operations.

5. REFERENCES

[1] Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages* (Vol. 1012). Oxford: Blackwell.

[2] Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422.

[3] Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America*, *106*(2), 1078-1083.

[4] Elangovan, S., & Stuart, A. (2011). A cross-linguistic examination of cortical auditory evoked potentials for a categorical voicing contrast. *Neuroscience letters*, *490*(2), 140-144.

[5] Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological science*, *21*(10), 1532-1540.

[6] Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375-425.

[7] Ostroff, J. M., Martin, B. A., & Boothroyd, A. (1998). Cortical evoked response to acoustic change within a syllable. *Ear and hearing*, *19*(4), 290-297.

[8] Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and hearing*, *31*(3), 302.

[9] Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical 'what'and 'where'pathways in the auditory system. *Trends in neurosciences*, *28*(4), 176-181.

[10] Johnson, K. L., Nicol, T., Zecker, S. G., Bradlow, A. R., Skoe, E., & Kraus, N. (2008). Brainstem encoding of voiced consonant–vowel stop syllables. *Clinical Neurophysiology*, *119*(11), 2623-2635.

[11] Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage*, *79*, 201-212.

[12] Akhoun, I., Gallégo, S., Moulin, A., Ménard, M., Veuillet, E., Berger-Vachon, C., ... & Thai-Van, H. (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme/ba/in normal-hearing adults. *Clinical Neurophysiology*, *119*(4), 922-933.

[13] Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

14] Hornickel, J., Chandrasekaran, B., Zecker, S., & Kraus, N. (2011). Auditory brainstem measures predict reading and speech-in-noise perception in school-aged children. *Behavioural brain research*, *216*(2), 597-605.

[15] Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proceedings of the National Academy of Sciences*, *106*(31), 13022-13027.

[16] Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences*, *109*(20), 7877-7881.