



# Cortical Auditory Event-Related Potentials and Categorical Perception of Voice Onset Time in Children With an Auditory Neuropathy Spectrum Disorder

Tyler C. McFayden<sup>1</sup>, Paola Baskin<sup>2</sup>, Joseph D. W. Stephens<sup>3</sup> and Shuman He<sup>4,5\*</sup>

<sup>1</sup>Department of Psychology, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, <sup>2</sup>Department of Anesthesiology, School of Medicine, University of California, San Diego, San Diego, CA, United States, <sup>3</sup>Department of Psychology, North Carolina Agricultural and Technical State University, Greensboro, NC, United States, <sup>4</sup>Department of Otolaryngology—Head and Neck Surgery, Wexner Medical Center, The Ohio State University, Columbus, OH, United States, <sup>5</sup>Department of Audiology, Nationwide Children's Hospital, Columbus, OH, United States

## OPEN ACCESS

### Edited by:

Xing Tian,  
New York University Shanghai, China

### Reviewed by:

Yan H. Yu,  
St. John's University  
Benjamin Joseph Schloss,  
Pennsylvania State University (PSU),  
United States

### \*Correspondence:

Shuman He  
shuman.he@osumc.edu

### Specialty section:

This article was submitted to Speech and Language, a section of the journal *Frontiers in Human Neuroscience*

**Received:** 17 January 2020

**Accepted:** 27 April 2020

**Published:** 25 May 2020

### Citation:

McFayden TC, Baskin P, Stephens JDW and He S (2020) Cortical Auditory Event-Related Potentials and Categorical Perception of Voice Onset Time in Children With an Auditory Neuropathy Spectrum Disorder. *Front. Hum. Neurosci.* 14:184. doi: 10.3389/fnhum.2020.00184

**Objective:** This study evaluated cortical encoding of voice onset time (VOT) in quiet and noise, and their potential associations with the behavioral categorical perception of VOT in children with auditory neuropathy spectrum disorder (ANSO).

**Design:** Subjects were 11 children with ANSO ranging in age between 6.4 and 16.2 years. The stimulus was an /aba/-/apa/ vowel-consonant-vowel continuum comprising eight tokens with VOTs ranging from 0 ms (voiced endpoint) to 88 ms (voiceless endpoint). For speech in noise, speech tokens were mixed with the speech-shaped noise from the Hearing In Noise Test at a signal-to-noise ratio (SNR) of +5 dB. Speech-evoked auditory event-related potentials (ERPs) and behavioral categorization perception of VOT were measured in quiet in all subjects, and at an SNR of +5 dB in seven subjects. The stimuli were presented at 35 dB SL (re: pure tone average) or 115 dB SPL if this limit was less than 35 dB SL. In addition to the onset response, the auditory change complex (ACC) elicited by VOT was recorded in eight subjects.

**Results:** Speech evoked ERPs recorded in all subjects consisted of a vertex positive peak (i.e., P1), followed by a trough occurring approximately 100 ms later (i.e., N2). For results measured in quiet, there was no significant difference in categorical boundaries estimated using ERP measures and behavioral procedures. Categorical boundaries estimated in quiet using both ERP and behavioral measures closely correlated with the most-recently measured Phonetically Balanced Kindergarten (PBK) scores. Adding a competing background noise did not affect categorical boundaries estimated using either behavioral or ERP procedures in three subjects. For the other four subjects, categorical boundaries estimated in noise using behavioral measures were prolonged. However, adding background noise only increased categorical boundaries measured using ERPs in three out of these four subjects.

**Conclusions:** VCV continuum can be used to evaluate behavioral identification and the neural encoding of VOT in children with ANSD. In quiet, categorical boundaries of VOT estimated using behavioral measures and ERP recordings are closely associated with speech recognition performance in children with ANSD. Underlying mechanisms for excessive speech perception deficits in noise may vary for individual patients with ANSD.

**Keywords:** auditory neuropathy spectrum disorders, voice onset time, auditory event-related response, speech perception, categorical perception

## INTRODUCTION

Auditory neuropathy spectrum disorder (ANSD) is a form of hearing impairment characterized by normal outer hair cell function, as indicated by the presence of cochlear microphonics (CMs) and/or otoacoustic emissions (OAE), in conjunction with an aberrant auditory neural system, as revealed by an absent or abnormal auditory brainstem response (ABR). ANSD is estimated to be present in 5–10% of newly identified hearing loss cases each year (Rance, 2005; Vlastarakos et al., 2008; Berlin et al., 2010; Bielecki et al., 2012). The presence of ANSD has been linked to several risk factors including premature birth, neonatal distress (e.g., hyperbilirubinemia, anoxia, artificial ventilation), infection (e.g., mumps, meningitis), neuropathic disorders (e.g., Charcot-Marie-Tooth syndrome, Friedreich's Ataxia), genetic factors (e.g., mutations in the *otoferrin* gene), and ototoxic drugs (e.g., carboplatin; Rance et al., 1999; Madden et al., 2002). Cases of ANSD with no apparent risk factors or medical comorbidities have also been reported (Berlin et al., 2010; Teagle et al., 2010; Roush et al., 2011; Bielecki et al., 2012; Pelosi et al., 2012). The exact lesions underlying the pathophysiology of ANSD have yet to be determined. Proposed sites of lesions include, but are not limited to, cochlear inner hair cells, synapses between inner hair cells and Type I auditory nerve fibers, and synapses between neurons in the auditory pathway.

Patients with ANSD typically exhibit poorer speech perception capability than would otherwise be expected based on the degree of hearing loss (Rance, 2005). These speech perception deficits can be partially accounted for by an impaired ability of the auditory system to detect changes in stimuli over time (i.e., temporal processing). Results of previous studies have shown that patients with ANSD have temporal processing deficits, and the severity of the deficits strongly correlates with their speech perception abilities (Starr et al., 1991; Zeng et al., 1999, 2001, 2005; Michalewski et al., 2005; Rance, 2005; He et al., 2015). Also, patients with ANSD typically experience excessive difficulty in understanding speech in the presence of competing for background noise (Kraus et al., 1984, 2000; Shallop, 2002; Zeng and Liu, 2006; Rance et al., 2007; Berlin et al., 2010). For example, Kraus et al. (2000) reported a case of an adult patient with ANSD who exhibited 100% speech recognition of monosyllabic words in quiet but achieved only 10% correct recognition of words at a signal-to-noise ratio (SNR) of +3 dB. In contrast, normal hearing subjects were able to retain an average score of 40% at the same SNR. This case was unique in that the patient had normal hearing thresholds,

which indicated that her perceptual difficulties did not stem from diminished audibility, but rather from ANSD-associated neural pathophysiology such as neural dyssynchrony. Similarly, Shallop (2002) presented an adult female subject who scored 100% on a speech perception sentence test in quiet but was unable to correctly identify any sentences in noise conditions of +12 dB SNR, despite hearing thresholds indicating only mild to moderate hearing loss. To date, the underlying mechanisms of this excessive difficulty in understanding speech with background noise remain poorly understood for patients with ANSD.

Cortical auditory event-related potentials (ERPs), including the onset response and the auditory change complex (ACC), are neural responses generated at the auditory cortex that can be recorded from surface electrodes placed on the scalp. The onset of ERP is elicited by the onset of sound, and its presence indicates sound detection. The ACC is elicited by a stimulus change that occurs within an ongoing, long-duration signal, and its presence provides evidence of auditory discrimination capacity at the level of the auditory cortex (Martin et al., 2008). Previous animal studies have demonstrated that auditory cortex is an important contributor to signal-in-noise encoding (Phillips, 1985, 1990; Phillips and Hall, 1986; Phillips and Kelly, 1992). Robust ERP responses are recorded only if the listener possesses accurate neural synchronization in response to sound. Maintaining synchronized neural responses that are time-locked to the speech stimuli is critical for successful speech perception in quiet and noise (Kraus and Nicol, 2003). Therefore, using ERPs to examine the neurophysiological representation of speech sounds in quiet and competing background noise at the level of the auditory cortex will be extremely beneficial for better understanding and characterizing speech perception difficulties in patients with ANSD.

In natural speech, the voice onset time (VOT) is a temporal cue that is crucial for differentiating voiced and voiceless English stop consonants. VOT refers to the interval between the release of a stop consonant (the burst) and the beginning of vocal fold vibration or voicing onset (Lisker and Abramson, 1964). Voiced stop consonants (e.g., /ba/, /da/, /ga/) have a relatively short VOT (generally 0–20 ms) and voiceless stop consonants (e.g., /pa/, /ta/, and /ka/) have a longer VOT. As the VOT increases, the perception rapidly changes from a voiced stop consonant to a voiceless consonant at 20–40 ms. The abrupt change in consonant identification is a classic example of categorical speech perception. Acuity for VOT identification is highly dependent on the synchronized neural response evoked

by the onset of voicing (Sinex and McDonald, 1989; Sinex et al., 1991; Sinex and Narayan, 1994).

ERP measures have previously been used to objectively evaluate neural encoding of VOT at the level of the auditory cortex in normal-hearing (NH) listeners and cochlear implant users (Sharma and Dorman, 1999, 2000; Sharma et al., 2000; Steinschneider et al., 2003, 2005, 2013; Roman et al., 2004; Frye et al., 2007; Horev et al., 2007; King et al., 2008; Elangovan and Stuart, 2011; Dimitrijevic et al., 2012; Apeksha and Kumar, 2019). For example, King et al. (2008) used ERPs to investigate the underlying neural mechanisms of categorical speech perception in NH children. The results of this study showed prolonged P1 latency recorded at long VOTs. Morphological characteristics of ERPs measured at different VOTs were the same. However, using the same stimuli, Sharma and Dorman (1999) observed an extra peak in ERP responses at long VOTs in normal hearing (NH) adults and this extra peak is believed to be elicited by the onset of voicing (Steinschneider et al., 2003, 2005, 2013). It should be pointed out that speech tokens used in these two studies were synthetic stop consonant-vowel (CV) syllables with a relatively short duration: 200 ms. Studies have shown that neural generators of ERPs have longer recovery periods in children than in adults (Ceponiene et al., 1998; Gilley et al., 2005). Therefore, the lack of the extra peak evoked by VOT in children could be due to insufficient separation between the stimulus onset and VOT such that responses evoked by the syllable onset and VOT overlap, resulting in a single broad peak in children as observed in King et al. (2008). To address this potential issue, relatively long vowel-consonant-vowel (VCV) stimuli that contain VOTs were used in this study so that neurons would have sufficient recovery time after responding to the syllable onset.

In summary, patients with ANSD are known to have neural dyssynchrony. Theoretically, the neurophysiological encoding of VOT in these patients should be compromised, which should account for their impaired categorical perception of VOT (Rance et al., 2008). However, this predictive framework has not been systematically evaluated in the pediatric ANSD population. Furthermore, the neural encoding of speech stimuli in noise at the level of the auditory cortex and the association between neural encoding and behavioral categorical perception of VOT in quiet and noise in subjects with ANSD remains largely

unknown. To address these needs, this study investigated the behavioral categorical perception of VOT and ERPs evoked by VOT in children with ANSD in quiet and speech-shaped background noise. We hypothesized that: (1) the precision of neural encoding of VOT would affect behavioral categorical perception performance in children with ANSD, and (2) ERPs evoked by the VOT in children with ANSD would be adversely affected by competing for noise.

## MATERIALS AND METHODS

### Subjects

Study participants included 11 children with ANSD (S1–S11) ranging in age between 6.4 and 16.2 years (mean: 10.1 years, SD: 3.0 years,  $n = 5$  females). All of the subjects were clinically diagnosed with ANSD based on the gross inconsistency between cochlear and neural functions. Specifically, subjects with ANSD were diagnosed based on the presence of CM ( $\pm$ OAE) with absent or abnormal ABRs. Results of Magnetic Resonance Imaging (MRI) revealed no evidence of dysplasia of the inner ear or internal auditory canal in any of these subjects. None of the subjects had any known cognitive impairments or developmental delays that might affect the results of this study. All subjects were placed in mainstream classrooms. For all except for one subject (S8), English was the only language used in their families. S8 was learning English as her primary language in school and used a combination of English and Hebrew at home.

The ear with better pure-tone hearing thresholds was selected as the test ear in this study. The degree of hearing loss of the test ear ranged from normal to severe, with an average pure tone audiometric threshold of 54.2 dB HL (calculated as the average of thresholds at 0.5, 1.0, and 2.0 kHz for each patient). All subjects except for S6 were fitted with hearing aids in the ears tested in this study. S6 did not use any amplification at the time of testing. Detailed demographic and audiology information for these subjects are listed in **Table 1**.

All subjects were recruited from the Ear & Hearing Center within the Department of Otolaryngology/Head and Neck Surgery at the University of North Carolina at Chapel Hill (UNC-CH). This study was approved by the Institutional Review Board (IRB) at UNC-CH. Written consent for the study procedures was provided by legal guardians of all subjects. Written assent

**TABLE 1** | Demographic information of all subjects with auditory neuropathy spectrum disorder (ANSD) who participated in this study.

Subject number	Gender	Risk factor	Ear tested	Age at testing (years)	3-Frequency PTA (dB HL)	PBK word scores (%)
1	F	Prematurity, hyperbilirubinemia	R	7.6	53.33	64.00
2	M	None	R	9.4	61.7	76.00
3	M	None	R	12.5	85	48.00
4	M	Prematurity	R	13.3	43.3	96.00
5	F	None	R	16.2	70	68.00
6	M	Prematurity	L	8.5	26.7	76.00
7	F	Prematurity, hyperbilirubinemia	L	10.5	71.7	40.00
8	F	Prematurity	R	6.4	66.7	28.00
9	M	Prematurity, hypoxia	L	7.8	56.7	16.00
10	M	Prematurity	L	6.6	55	36.00
11	F	None	R	11.12	60	80.00

PTA, Pure-Tone Average; PBK, Phonetically Balanced Kindergarten.

for the study procedures was obtained from all subjects except for subjects 8 and 10. Oral assent for participating in this study was provided by these two subjects. Monetary compensation was provided to all subjects for participating in the study.

## Stimuli

The stimuli were a VCV continuum with the consonants ranging from a voiced /b/ to a voiceless /p/, in an /a/ context. For this study, the goal of the stimulus creation process was to produce a natural-sounding series in which VOT was manipulated while holding other acoustic/phonetic properties constant to the extent that such control was possible. In other studies that use voicing continua based on natural speech, stimuli are often created *via* cross-splicing between recordings of endpoint tokens (e.g., Ganong, 1980). However, VOT is only one of several acoustic/phonetic properties that may differ between natural tokens of voiced and voiceless stop consonants in English (Lisker, 1986). Perceptual judgments of voiced and voiceless consonants can be influenced by many differences, for example, variations in the first formant (F1) onset shift perception of voicing in addition to VOT (Stevens and Klatt, 1974). Thus, the current methods attempted to isolate the VOT manipulation by using a re-synthesis procedure instead of cross-splicing.

Stimuli were created based on materials and methods used in a previous study (Stephens and Holt, 2011) that manipulated VCV utterances by using linear predictive coding (LPC; Atal and Hanauer, 1971) to derive source and filter properties of natural utterances and resynthesize intermediate utterances based on modifications to the source and/or filter properties. For the current stimuli, Praat software (Boersma, 2001) was used to derive LPC filter coefficients from a natural token of /aba/. A separate natural token of /ada/, temporally aligned with the /aba/ utterance, was inverse filtered by its LPC coefficients to produce a voicing source, which was then edited to create eight different voicing sources with varying VOT. The series of source waveforms were created by deleting pitch periods from the original source and inserting equivalent lengths of Gaussian noise at the onset of the consonant to simulate aspiration. All other aspects of the voicing sources were held constant.

The LPC filter for /aba/ was then applied to each of the eight source signals to create sounds with identical formant structure to /aba/, but with varying VOTs. The resulting VOTs in the resynthesized stimuli were 0 ms (voiced endpoint), 11 ms, 24 ms, 29 ms, 36 ms, 49 ms, 62 ms, 75 ms, and 88 ms (voiceless endpoint). The slight variation in step size along the series resulted from the procedure of lengthening VOT by deleting pitch periods from the voicing source. On average, pitch periods in the signal were roughly 10 ms, corresponding to a fundamental frequency of approximately 100 Hz; however, there was some natural variation in the length of individual pitch periods. All sounds were sampled at 11,025 Hz and were matched in RMS amplitude after the re-synthesis procedure. The overall length of each stimulus was approximately 720 ms. Spectrograms of each of the eight stimuli are displayed in **Figure 1**. As shown in the figure, the onset of periodicity gradually shifted from one end of the series to the other, while the formant structure

remained relatively constant. In the longer-VOT stimuli, the formant frequencies (including F1) were excited by the aspiration noise that was inserted into the voicing sources (i.e., F1 was not cut back). It should be noted that the F1 and F2 frequencies *at the onset* of voicing were slightly different at different VOTs. In perception, these spectral differences immediately following voice onset may also contribute to the perception of consonant voicing, but systematic experiments with synthetic stimuli have found formant frequency at voice onset to be a relatively weaker cue than VOT (Lisker, 1975). Furthermore, formant frequency differences have been found to influence consonant voicing distinctions to a lesser extent in children (the population of interest in the current study), than in adults (Morrongiello et al., 1984). More importantly, altering formant frequency has been found to shift ERP response patterns and behavioral categorical perception of VOT in a parallel manner (Steinschneider et al., 2005). Nevertheless, stimuli used in this study did not solely contain timing cues, which needs to be taken into consideration when interpreting the results of this study.

For VOT identification in noise, speech tokens were mixed with the speech-shaped noise taken from the Hearing In Noise Test for Children (HINT-C; Nilsson et al., 1996) at an SNR of +5 dB using Audacity (v.1.3 Beta).

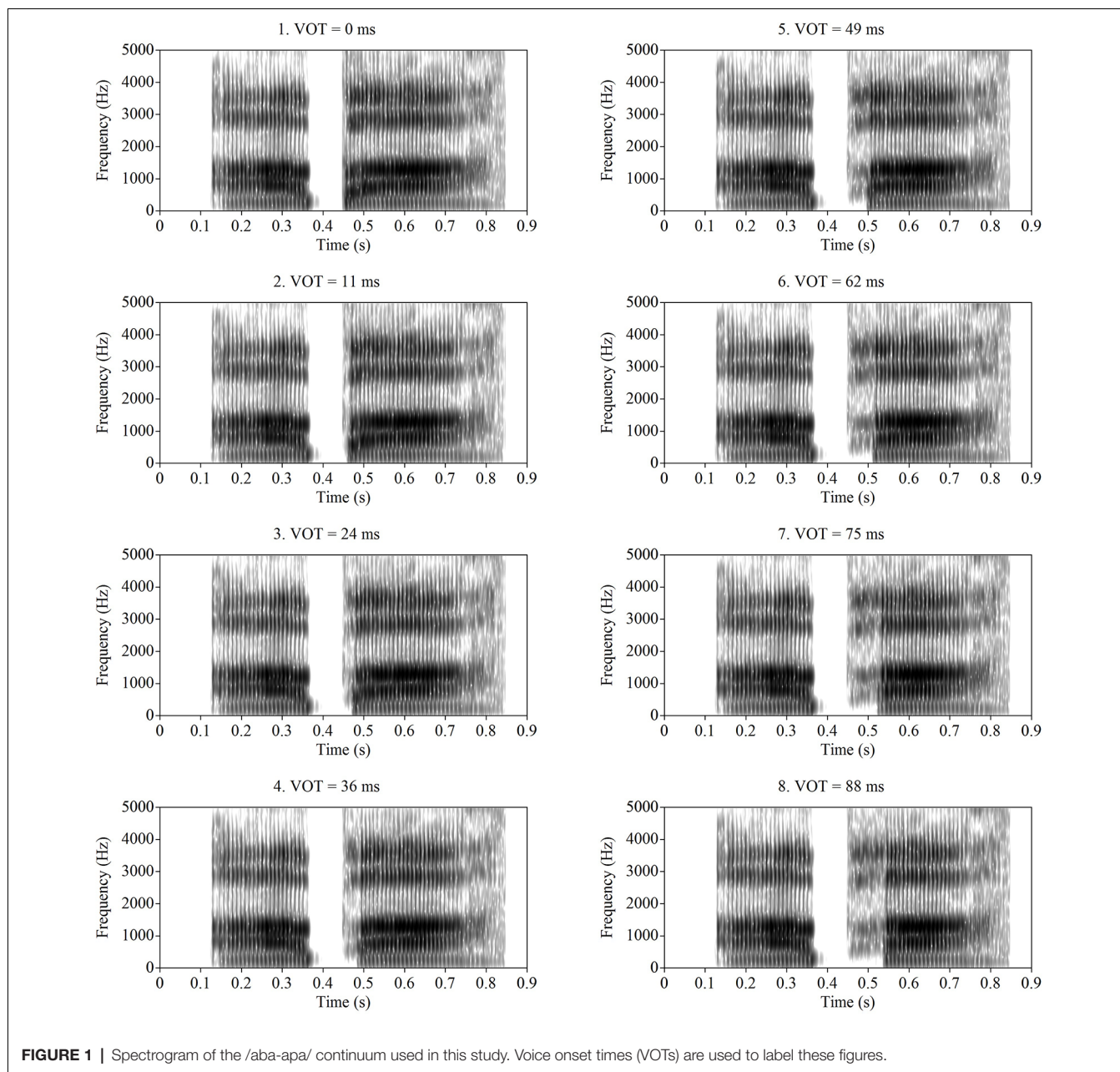
## General Procedures

Each subject completed both behavioral measures of categorical perception of VOT and electrophysiological recordings of ERPs for all stimulation conditions in which they were tested. Both procedures were undertaken *without* using subjects' hearing aids. These two measures were completed in different sessions scheduled on the same day. In general, it took about three and a half hours to complete both procedures for each test condition (i.e., quiet or noise).

Categorical perception of VOT and electrophysiological measures of the speech-evoked ERP were evaluated in quiet in all subjects. Besides, these two procedures were also undertaken at an SNR of +5 dB in seven subjects (S1–S5, S8, and S11). ERPs evoked by stimuli presented in noise with VOTs of 75 or 88 ms were not recorded in S11 due to time constraints. All stimuli were presented using the Neuroscan Stim2 (Compumedics, Charlotte, NC) at 35 dB SL (re: pure tone average) or 115 dB SPL (maximum output level of the stimulation system without any distortion) if this limit was less than 35 dB SL. Stimulus level was calibrated using a Larson-Davis 824 sound level meter, a 6-cc coupler for supra-aural earphone, and a 2-cc coupler for the ER-3A insert earphone.

## Categorical Perception of VOT

The stimulus was delivered through a Sennheiser supra-aural earphone (HD8 DJ). A two-interval, two-alternative forced-choice procedure was used. Listening intervals were visually indicated using computer graphics with /aba/ and /apa/ shown in yellow and blue, respectively. An initial practice session using tokens with VOTs of 0 and 88 ms was provided to each subject before data collection. Ten presentations of each of eight tokens (i.e., 80 in total) were presented to the test ear. The sequence



of these presentations was pseudo-randomized on a trial-by-trial basis. For each presentation, subjects were asked to indicate whether they heard /aba/ or /apa/ by pointing or selecting the interval with the associated graphic or color. No feedback was provided. The percentage of /aba/ responses was calculated. Subjects needed approximately 10 min to complete this task for each stimulation condition.

### Electrophysiological Recordings

For each subject, electrophysiological recordings were completed in up to four test sessions and each session lasted approximately 2 h. Subjects were tested in a single-walled sound booth. They were seated in a comfortable chair watching a silent movie with

closed captioning. Breaks were provided as necessary. All stimuli were presented through an ER-3A insert earphone. The inter-stimulus interval was 1,200 ms.

Electroencephalographic (EEG) activity was recorded using the Neuroscan SCAN 4.4 software and a SynAmpRT amplifier (Compumedics, Charlotte, NC, USA) with a sampling rate of 1,000 Hz. Disposable, sterile Ag-AgCl surface recording electrodes were used to record the EEG. In nine subjects, responses were recorded differentially from five electrodes (Fz, FCz, Cz, C3, and C4) to contralateral mastoid ( $A_{1/2}$ , reference) relative to body ground at the low forehead ( $F_{pz}$ ). However, S9 pulled off three electrodes placed on the scalp (i.e., FCz, Cz, and C4) before data collection was

completed. Replacing these scalp electrodes was rejected by this subject. Responses were only differentially recorded from Fz to contralateral mastoid in S8 and S10 due to the lack of sufficient subject compliance for the prolonged testing time required by this study. Therefore, ERP responses were only recorded from Fz for all stimulation conditions in these three subjects. Eyeblink activity was monitored using surface electrodes placed superiorly and inferiorly to one eye. Responses exceeding  $\pm 100 \mu\text{V}$  were rejected from averaging. Electrode impedances were maintained below  $5 \text{ k}\Omega$  for all subjects. The order in which the EEG data were collected was randomized across VOTs to minimize the potential effect of attention or fatigue on EEG results. The EEG was epoched and baseline corrected online using a window of 2,000 ms, including a 100-ms pre-stimulus baseline and a 1,900-ms peri/post-stimulus time. Auditory evoked responses were amplified and analog band-pass filtered online between 0.1 and 100 Hz (12 dB/octave roll-off). After artifact rejection, the remaining (at least 100) artifact-free sweeps were averaged to yield one replication. For each stimulation condition for each subject, three replications with 100 artifact-free sweeps/replication were recorded at all recording electrode locations. These recordings were digitally filtered between 1 and 30 Hz (12 dB/octave roll-off) offline before response identification and amplitude measurements.

## Data Analysis

For the results of the behavioral categorical perception test, the percentage of /apa/ responses were calculated and plotted as a function of VOT. These results were fitted using a least-squares procedure for each subject with a logistic function of the form:

$$P(x) = a / (1 + e^{-b(x-x_0)})$$

where  $P$  is the percentage of trials the token was perceived as /apa/ (0–1),  $a$  is the upper limit of the performance,  $x_0$  is the midpoint of the function, and  $b$  is the slope, where larger values represent steep functions.

The point on the psychometric function that corresponds to chance performance (the token was perceived to be /apa/ 50% of the time) was determined for each subject. The VOT that yielded this chance performance was defined as the behavioral categorical boundary of the /aba-apa/ continuum. This criterion has been used in previously published studies (e.g., Sharma and Dorman, 1999; Elangovan and Stuart, 2008).

Grand mean averages of ERPs recorded from all subjects were computed for each stimulating condition and used to determine the latency ranges for which the onset and the ACC response were measured. The windows for the onset and the ACC response were from 25 to 240 ms and from 390 to 640 ms, relative to the stimulus onset, respectively. For each subject, replications evoked by the same speech tokens were averaged for each stimulation condition. As a result, eight averaged responses were yielded for each subject in each stimulation condition except for S11 tested at the SNR of 5 dB. The averaged responses were used for peak identification, as well as peak amplitude and latency measures.

For the eight subjects from whom ERP responses were recorded at five recording electrode locations (i.e., S1–S7, S11), responses were examined across these electrode sites to help identify ACC responses. ERP responses recorded in these subjects were independently assessed by two researchers (authors PB and SH). The presence of the ACC response was determined based on two criteria: (1) a repeatable neural response within the expected time window for the ACC based on mutual agreement between the two researchers; and (2) an ACC response recorded in all five electrode sites. For three subjects whose ERPs were only recorded from Fz, their responses were evaluated by the third researcher (author TCM). The presence of the ACC response in these three cases was determined based on mutual agreement among all three researchers that a repeatable neural response could be identified within the expected time window for the ACC. The objective categorical boundary was defined as the shortest VOT that could reliably evoke the ACC response in this study. For the ACC response identification, the inter-judge agreement among three researchers was 91% and between authors, PL and SH were 93%. In cases where judges initially differed in peak identification, the differences were mutually resolved following consultation and discussions.

Despite a wide range of ages at the time of testing, both the onset response and the ACC recorded in all subjects consisted of a vertex positive peak (P1) followed by a negative trough (N2). For the onset and the ACC, the P1 was identified as a positive peak occurring within a time window between 40 and 150 ms and a time window between 410 and 520 ms after stimulus onset, respectively. The N2 was identified as the negative trough following the P1 occurring approximately 100 ms later. Latencies and amplitudes of the P1 and N2 peaks were measured using a custom-designed MATLAB (Mathworks) software at the maximum positivity or negativity in the estimated latency window of both peaks. The peak-to-peak amplitude was measured as the difference in voltage between the P1 and N2 peaks.

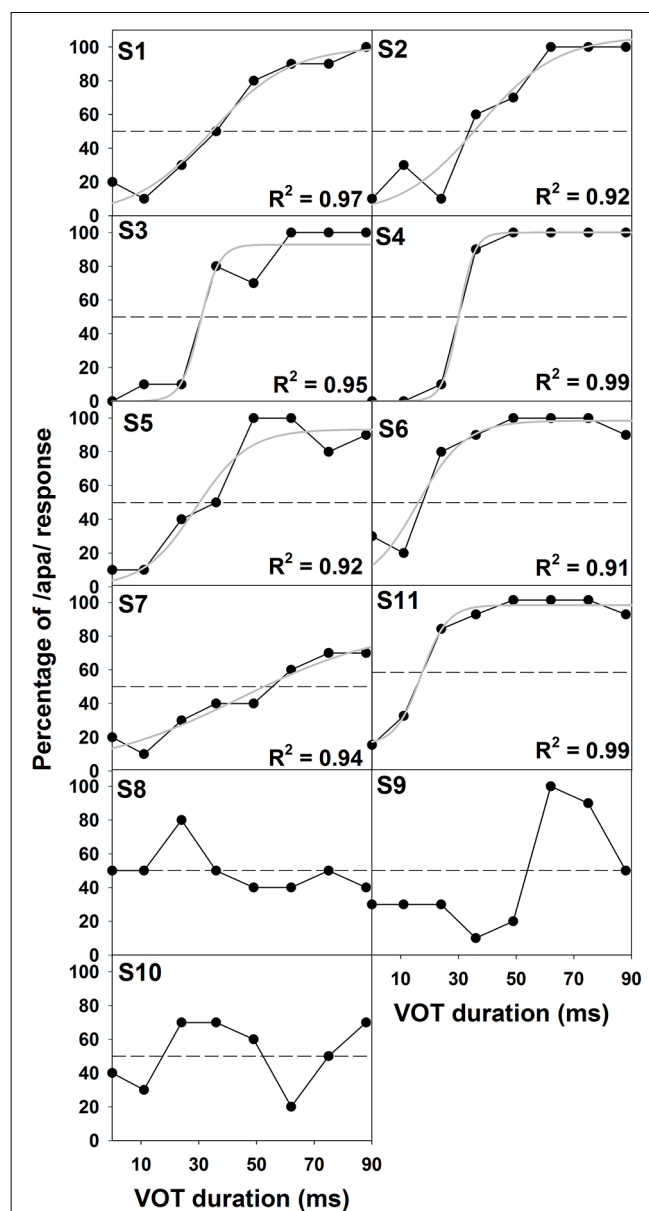
Dependent variables measured for ERP results included the objective categorical boundary, latencies of the P1 and the N2 peak, and the peak-to-peak amplitude of the onset and the ACC responses. The Friedman test was used to evaluate the effects of recording electrode location on ERP responses in a subgroup of seven subjects. The related-sample Wilcoxon Signed Rank test was used to compare: (1) behavioral and objective categorical boundaries; and (2) effects of competing for noise on amplitude and latency of P1 and N2 peaks of the onset and the ACC responses for a subgroup of seven subjects. The one-tailed Spearman Rank correlation test was used to evaluate the association between behavioral and objective categorical boundaries. Also, potential associations between categorical boundaries and the most recently measured aided Phonetically Balanced Kindergarten (PBK) word scores were evaluated using a one-tailed Spearman Rank Correlation test for these subjects. The PBK word scores were measured approximately 1 month before the study for all subjects.

## RESULTS

Overall, behavioral categorical perception of VOT and speech evoked ERPs were measured from subjects in both quiet and noise conditions.

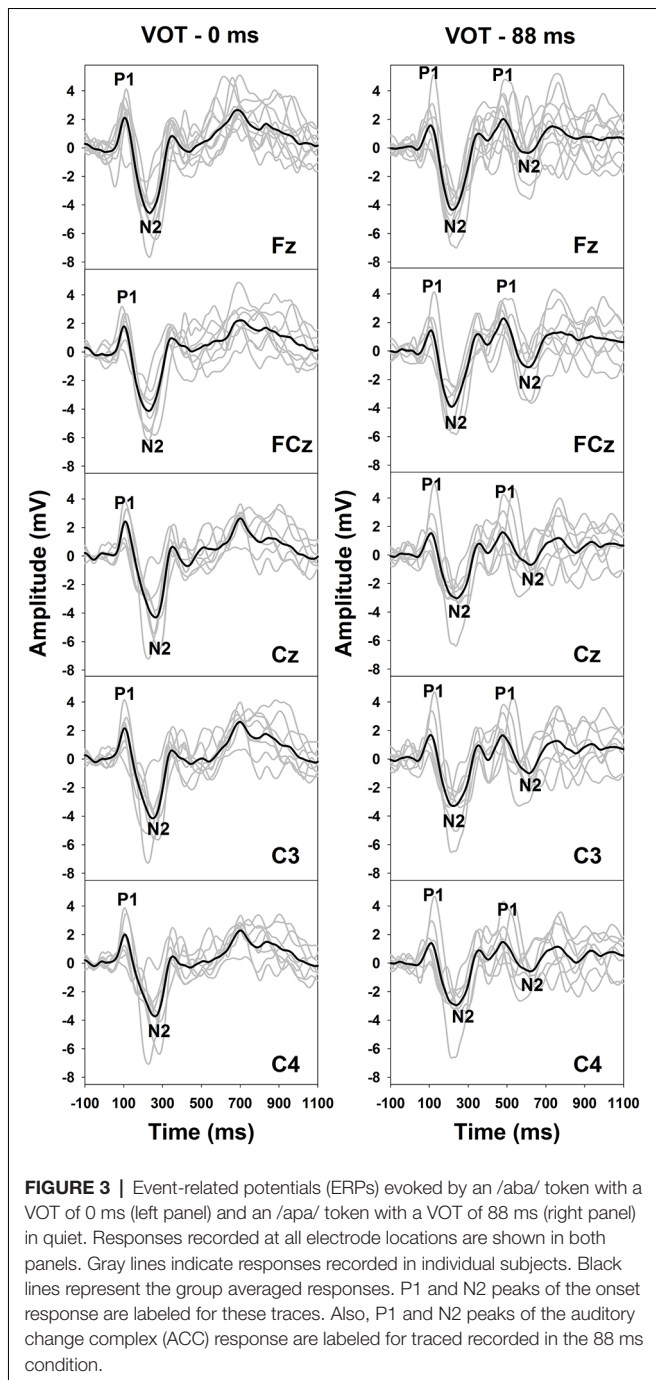
### Results Measured in Quiet

Figure 2 shows the results of behavioral categorical perception of VOT measured in all subjects. Each panel shows the percentage of trials that the speech token was perceived as /apa/ plotted as a function of VOT duration measured in one subject. In general, psychometric function fits for data recorded in S1–S7 and S11 were good, accounting for 91–99% of the variance in these data. For these eight subjects, auditory perception of these speech tokens changed from /aba/ to /apa/ as VOT duration increased. However, these results showed large individual variability. While some subjects demonstrated relatively short behavioral categorical boundaries of the /aba-apa/ continuum (e.g., 17.4 ms in S6), other subjects required relatively long VOTs for the speech token to be perceived as the /apa/ (e.g., 56 ms in S7). Also, the steepness of the fitted psychometric functions as indicated by the slope varied considerably across subjects. The slope of the psychometric function ranged from 2.7 to 16.4 percent of trial/ms with a mean of 9.2 percent of trial/ms ( $SD = 5.0$ ). Results of a Spearman Rank Correlation test showed no significant correlation between the slope of the psychometric function and behavioral categorical boundaries in these subjects  $\rho = -0.15$ ,  $p = 0.36$ ). Results measured in three subjects (S8–10) showed evidence of auditory confusion in VOT perception. Specifically, auditory perception of these speech tokens did not necessarily change from /aba/ to /apa/ as the VOT increased, which indicated that the VOT cannot be accurately perceived by these three subjects. These functions do not fit a sigmoidal distribution. Therefore, the behavioral categorical boundary of the VOT could not be defined for these three subjects. For data analysis, a conservative estimate of 89 ms was used as their behavioral categorical boundaries of the VOT in this study. Figure 3 shows ERPs recorded from individual subjects (gray lines) and grand averages (black lines) recorded at five electrode sites in quiet for VOTs of 0 ms (left panel) and 88 ms (right panel). It should be noted that data shown at Fz included ERP responses recorded from all subjects and data shown at other electrode locations only include results recorded in a subgroup of eight subjects. Robust onset responses could be easily identified in these results. For responses evoked by the VCV syllable with a VOT of 88 ms, the ACC response could also be identified in addition to the onset response. Both the onset and the ACC response consist of a P1 peak followed by an N2 peak despite individual variability in amplitudes and latencies of both responses. Figure 4 shows ERP responses recorded in three subjects (S2, S3, and S9). These three subjects were selected because their results extended the entire range of objective categorical boundaries measured in this study. ERP responses were only recorded at electrode sites Fz and C3 in S9 due to insufficient subject cooperation for prolonged testing time. Therefore, only ERPs recorded at the mid-line electrode site (Fz) were shown for all three subjects. Robust onset responses were recorded in each subject at all VOT



**FIGURE 2 |** Results of behavioral categorical perception tests recorded in all subjects. In each panel, the abscissa shows VOT durations tested in this study. The ordinate indicates the percentage of /apa/ response at different VOT durations in these subjects. The subject number is indicated in the upper left corner. Also shown is the fitted psychometric function (gray lines) for results measured in S1–S7 and S11. The percentage variance that can be explained by the psychometric function (i.e.,  $R^2$ ) is indicated in the lower right corner for each of these eight subjects. Results of behavioral categorical perception tests recorded in S8–S10 could not be characterized by a logistic regression function. Therefore, no psychometric function is shown for these three subjects.

durations. Besides, ACC responses elicited by VOTs were also recorded in S2 and S3. The objective categorical boundary was 11 and 36 ms in S2 and S3, respectively. The ACC response cannot be identified in ERP responses recorded in S9. Therefore, the objective categorical boundary was determined to be 89 ms



as a conservative estimate for data analysis. Results of the behavioral categorical test also showed that this subject could not perceive VOT despite the good audibility of the stimuli (Figure 2).

Effects of VOT durations on ACC amplitudes and latencies of P1 and N2 peak were evaluated using a Spearman Correlation test for results recorded in a subgroup of seven subjects (S1–S6 and S11). Responses were analyzed from the recording electrode Cz to compare our results with the published literature. Results of correlation analyses showed that P1 latencies increased as

the VOT duration increased in all seven subjects ( $p < 0.05$ ). However, the effects of the VOT duration on ACC amplitudes and N2 latencies were less consistent across subjects.

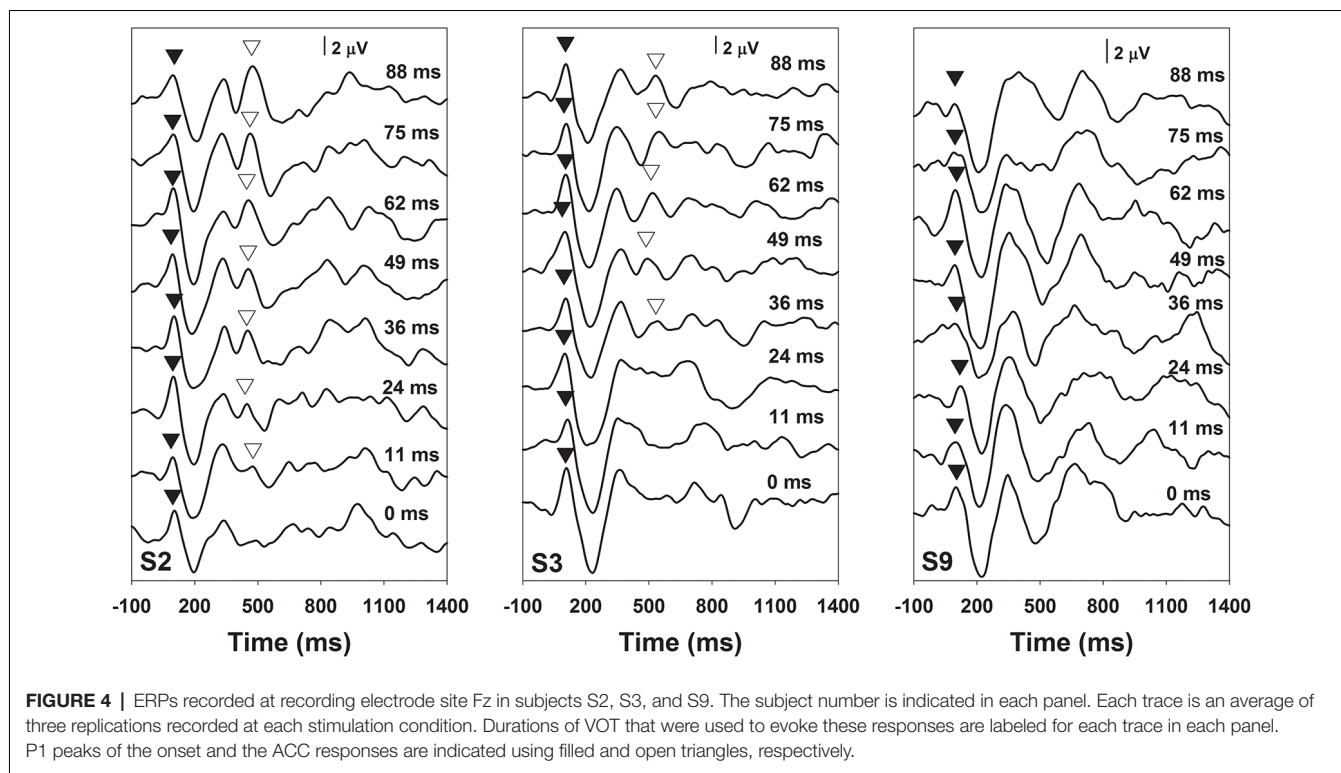
For ERPs recorded in the seven subjects who had results at all five recording locations, effects of recording locations on response amplitude, P1, and N2 latencies of the onset and the ACC responses evoked by stimuli with different VOT durations were evaluated using Friedman tests with the recording electrode location as the within-subject factor. Results showed that there were significant differences in amplitude of the onset response ( $\chi^2_{(4)} = 15.93, p < 0.05$ ) and the ACC ( $\chi^2_{(4)} = 11.35, p < 0.05$ ) recorded at different recording locations. For the onset response, results of related-sample Wilcoxon signed-rank tests showed significant differences in amplitude between results recorded at three electrode pairs [i.e., Fz vs. C4 ( $p = 0.012$ ), Fz vs. C3 ( $p = 0.018$ ), and Cz vs. C4 ( $p = 0.017$ )]. For the ACC, results of related-sample Wilcoxon signed-rank tests showed a significant difference in amplitude between results recorded at Fz and those measured at C4 ( $p = 0.036$ ). There was no significant difference in amplitudes of the onset of the ACC recorded between any other two recording electrode locations ( $p > 0.05$ ). Recording locations did not show significant effects on P1 latency (the onset:  $\chi^2_{(4)} = 0.84, p = 0.93$ ; the ACC:  $\chi^2_{(4)} = 8.08, p = 0.09$ ) or N2 latency (the onset:  $\chi^2_{(4)} = 6.27, p = 0.18$ ; the ACC:  $\chi^2_{(4)} = 1.47, p = 0.78$ ) of the onset of the ACC response.

VOT boundaries measured in quiet using behavioral and ERP measures are listed in Table 2. Results of the related-sample Wilcoxon Signed Rank test showed no difference between VOT boundaries measured in quiet using behavioral or ERP measures ( $p = 0.09$ ). Excluding the results of S8–S10 did not change this statistical result. Figure 5A shows the associations between VOT boundaries measured in quiet in all subjects using these two procedures. Also shown is the result of linear regression. Results of the one-tailed Spearman Rank Correlation test showed a robust correlation between VOT boundaries estimated using these two procedures  $\rho = 0.78, p < 0.05$ ). However, a careful inspection of the Figure 5A suggests that this significant correlation might have been driven by the results of S8–S10. Figure 5B shows the associations between VOT boundaries measured in quiet using these two procedures in S1–S7 and S11. The association between these two measures is less strong than that shown in Figure 5A. This observation is consistent with the non-significant correlation revealed by the results of the one-tailed Spearman Rank Correlation test  $\rho = 0.41, p = 0.16$ .

Figure 6 shows the association between PBK word scores, behavioral (left panel), and objective categorical boundaries of VOT (right panel) measured in quiet. Results of one-tailed Spearman Rank Correlation tests showed robust negative correlations between PBK word scores and categorical boundaries of the VOT estimated using both behavioral  $\rho = -0.90, p < 0.05$ ) and objective  $\rho = -0.79, p < 0.05$ ) measures.

In summary, for results measured in quiet, subjects tested in this study showed a wide range of behavioral categorical boundaries of VOT. While some subjects could not accurately perceive VOTs, other subjects required relatively short VOTs





**TABLE 2 |** Objective and behavioral categorical boundaries of voice onset time (VOT) in milliseconds measured in quiet and noise.

Subject number	Behavioral categorical boundary in quiet	Objective categorical boundary in quiet	Behavioral categorical boundary in noise	Objective categorical boundary in noise
1	35	11	89	24
2	35.6	11	49.2	24
3	36	36	36	36
4	30	24	41	24
5	30	49	31	49
6	17.4	11		
7	56	36		
8	89	89	89	89
9	89	89		
10	89	89		
11	17.4	11	36.6	36

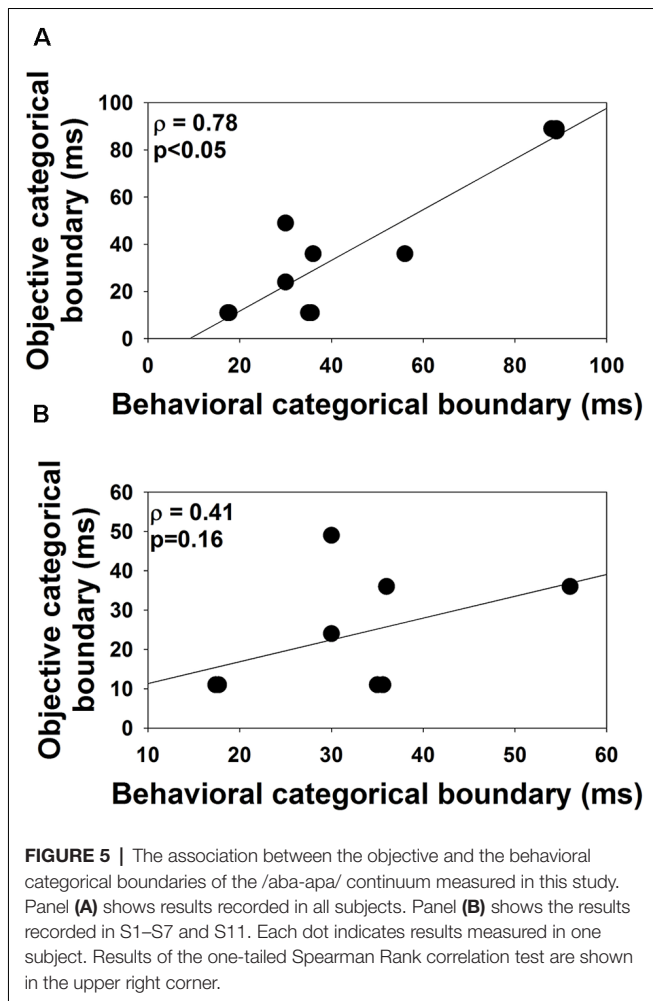
for the speech token to be perceived as the /apa/. The onset response of the speech evoked ERP was recorded in all subjects tested in this study. In contrast, not all subjects showed the ACC response. The objective categorical boundary ranged from 11 to 89 ms. Categorical boundaries of VOT measured in quiet using behavioral or ERP measures are negatively correlated with PBK word scores.

### Effect of Adding Competing Background Noise

VOT boundaries measured at an SNR of 5 dB using behavioral and ERP measures are listed in **Table 2**. Results of a related-sample Wilcoxon Signed Rank test showed a non-significant difference between behavioral and objective categorical boundaries of VOT measured in noise ( $p = 0.23$ ). The correlation between categorical boundaries of VOT estimated

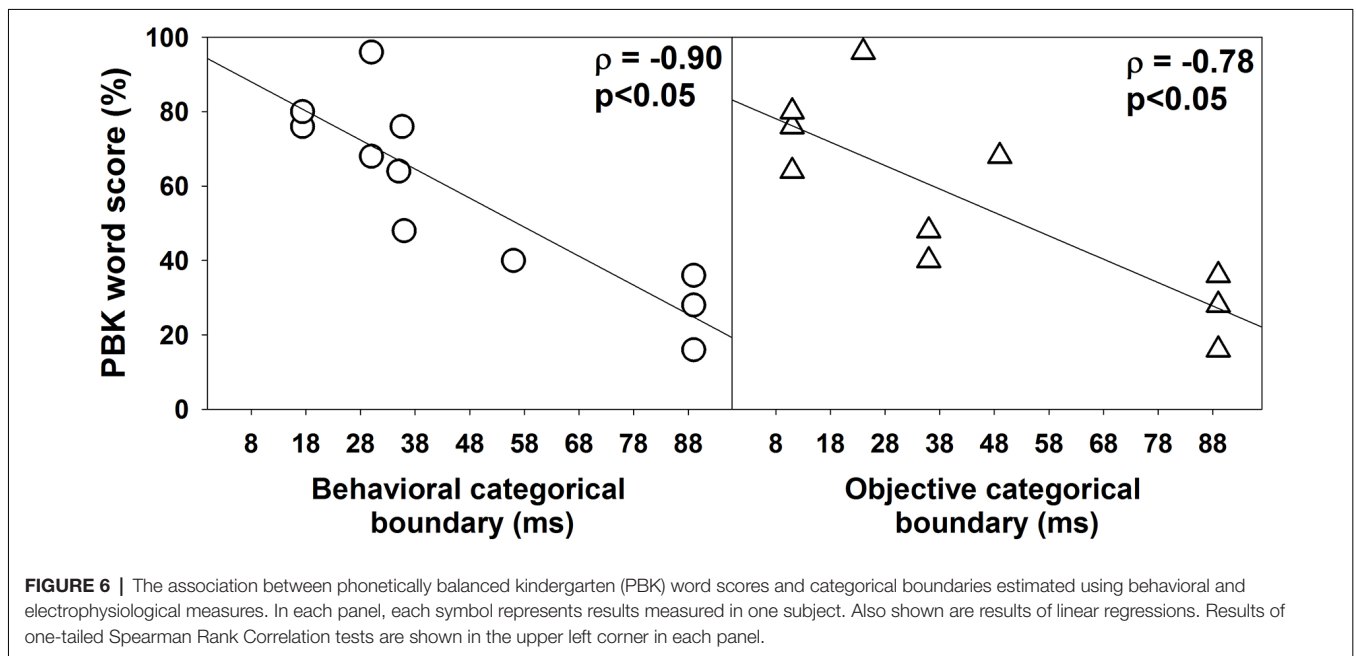
using these two procedures was not statistically significant, as revealed by the results of a one-tailed Spearman Rank Correlation test  $\rho = -0.26$ ,  $p = 0.28$ .

Inspection of **Table 2** suggests that adding a competing background noise has different effects on behavioral and objective categorical boundaries of VOT. For results measured using behavioral procedures, results of a related-sample Wilcoxon Signed Rank test showed that behavioral categorical boundaries of VOT measured in noise were significantly longer than those measured in quiet ( $p < 0.05$ ). However, there was substantial inter-subject variability. While some subjects showed significant increases in behavioral categorical boundaries (e.g., S1 and S11), other subjects demonstrated negligible changes in their behavioral categorical boundaries (e.g., S4 and S5). For results measured using objective procedures, results of a related-sample Wilcoxon Signed Rank test showed that adding



a competing background noise did not significantly increase the categorical boundaries of VOT ( $p = 0.10$ ). Similar to results measured using behavioral procedures, adding competing background noise showed mixed results across these seven subjects who were tested in both quiet and noise. While some subjects showed prolonged objective categorical boundaries (i.e., S1, S2, and S11), other subjects showed the same objective categorical boundaries regardless of the presence/absence of the background noise (i.e., S3–S5).

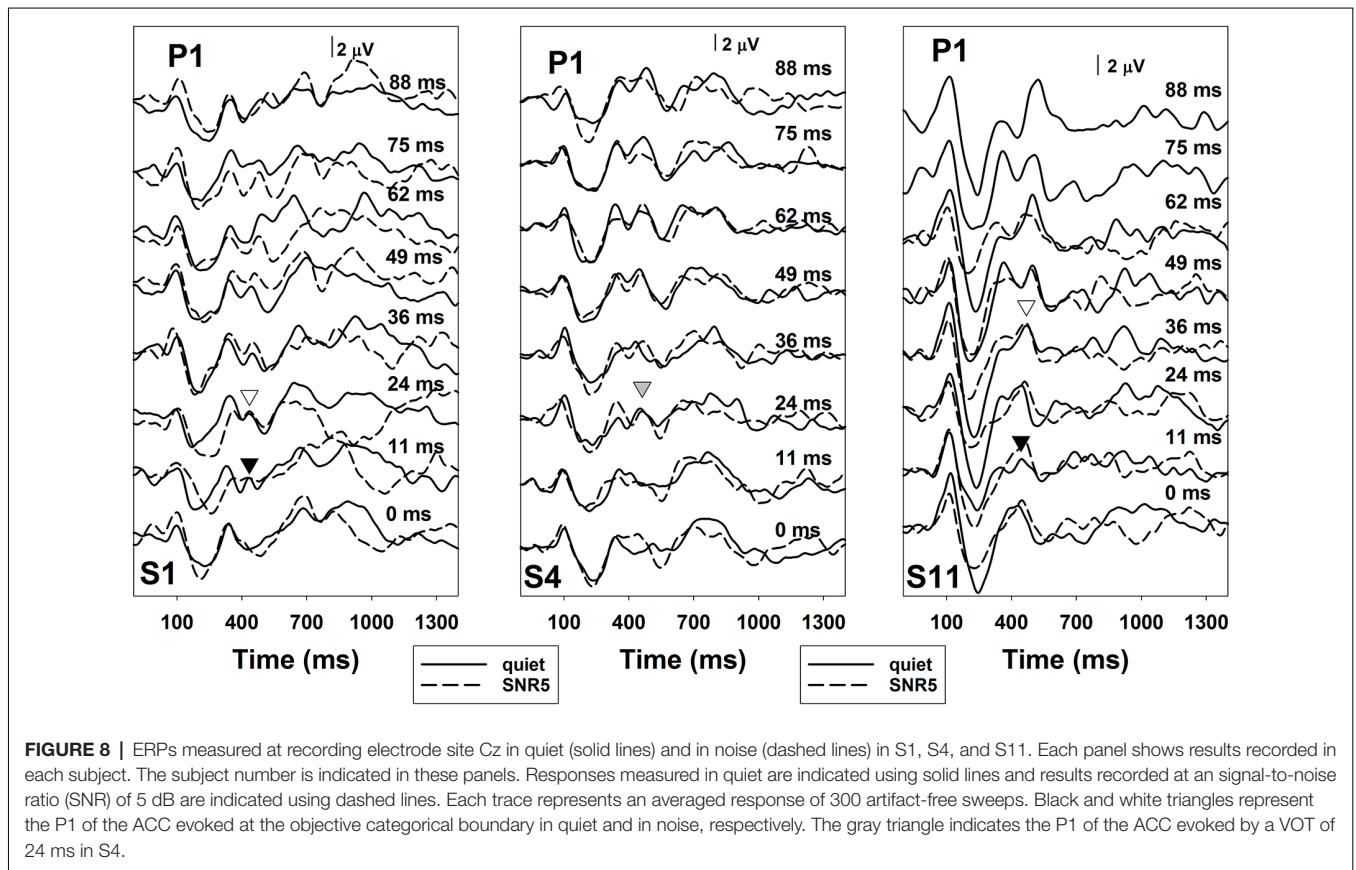
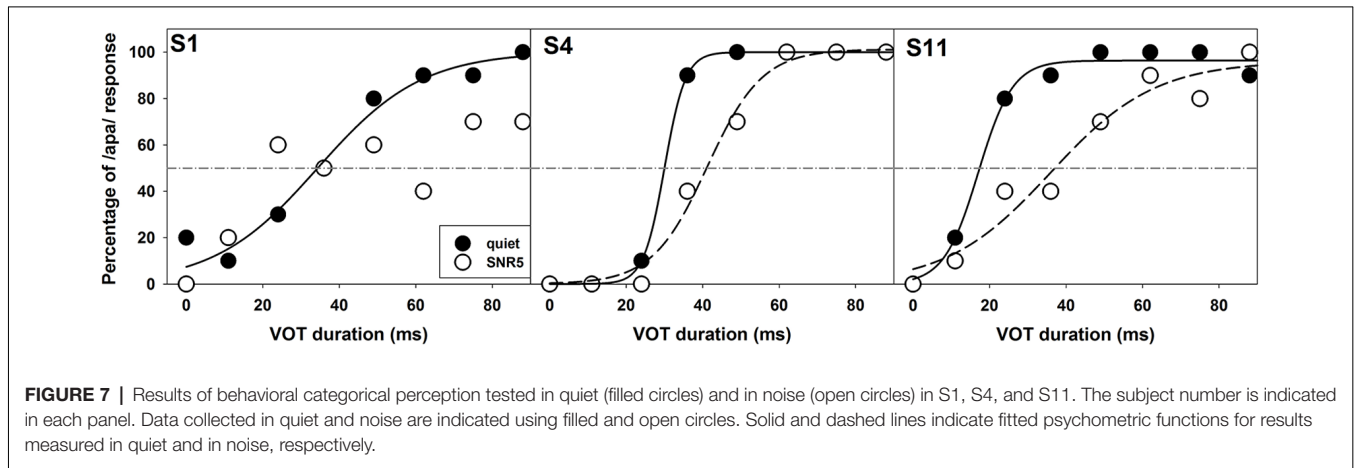
The comparison of noise effects on results of ERP measures and the behavioral categorical test showed some interesting findings. While masking noise affected both neural encoding and auditory perception of VOT in some subjects (S1, S2, S4, and S11), it did not affect results recorded in other subjects (S3 and S5). Also, for these four subjects whose results were affected by masking noise, the degree of these effects was different for ERPs and behavioral auditory perception of VOT. While some subjects showed masking noise affecting both measures in a similar degree (S2 and S11), other subjects demonstrated much larger noise effects on their auditory perception than the neural encoding of VOT (S1 and S4). **Figures 7, 8** show results recorded in three subjects with different effects of masking noise on these two measures. **Figure 7** shows the results of behavioral categorical perception. Left, middle and right panels show results measured in S1, S4, and S11, respectively. Adding a competing noise decreased the slopes of psychometric functions in S4 and S11. Also, objective categorical boundaries increased by 11 ms in S4 and 19.2 ms in S11 with the presence of noise. Results measured in S1 clearly showed that this subject could not perceive VOT with a competing background noise despite a good categorical perception in quiet. **Figure 8** shows ERPs recorded at Cz in S1, S4, and S11. Robust ERPs were recorded in all three subjects in both quiet and noise. Adding a competing



background noise had different effects on the neurophysiological representation of VOT in these three subjects. For S1 and S11, filled and open triangles represent the P1 of the ACC evoked at the objective categorical boundary in quiet and in noise, respectively. Adding a competing background noise at an SNR of 5 dB in S1 and S11 increased the objective categorical boundary by 13 and 24 ms, respectively. For S4, ERP measures in quiet and noise yielded the same categorical boundary (i.e., 24 ms). The P1 of the ACC evoked by a VOT of 24 ms is indicated using a gray triangle. Comparisons

between results shown in **Figures 7, 8** indicate that S1 cannot perceive VOT despite a relatively good neural encoding of this timing cue at an SNR of 5 dB. For S4, adding a competing noise did not affect ERPs evoked by the VOT, but did show a small effect in auditory perception. For S11, ERPs and auditory perception of VOT were affected by a competing noise in parallel.

For each study participant, amplitudes and latencies of the onset response recorded in all VOT conditions were averaged. The calculation was conducted separately for data recorded in



quiet and noise. Differences in these two variables between results measured in these two conditions were compared using related-sample Wilcoxon Signed Rank tests. Similar comparisons were conducted for the ACC results, except that the comparison only included data recorded in VOT conditions, where the ACC was recorded in both quiet and noise conditions. For the onset response, there was no significant difference in amplitude ( $p = 0.11$ ) or N2 latency ( $p = 0.71$ ) measured in quiet and noise. However, P1 latencies measured in noise were significantly longer than those measured in quiet ( $p < 0.05$ ). Adding a competing noise did not show a significant effect on amplitude ( $p = 0.06$ ), P1 latency ( $p = 0.21$ ), or N2 latency ( $p = 0.88$ ) of the ACC response.

In summary, adding a speech-shaped noise at an SNR of +5 dB only significantly increased the categorical boundaries of VOT estimated using behavioral procedures. The effects of competing for background noise on ERPs and auditory perception of VOT were not always in parallel and varied across subjects.

## DISCUSSION

This study investigated the categorical perception and neural encoding of the VOT in children with ANSD in both quiet and noise conditions. Relatively long VCV stimuli that contain VOTs were used in this study as novel stimuli to better separate responses evoked by the VOT from those evoked by the syllable onset. Speech evoked ERPs were recorded in all subjects tested in this study. Onset responses recorded in all subjects were robust, which indicates the good audibility of the stimuli. In addition to the onset response, the ACC response was recorded in eight subjects at VOTs of 11 ms or longer. The P1 latency of the ACC response increased as the VOT duration increased, which is generally consistent with the existing literature (Steinschneider et al., 1999; Sharma and Dorman, 2000; Sharma et al., 2000; Tremblay et al., 2003; Frye et al., 2007; King et al., 2008; Elangovan and Stuart, 2011; Dimitrijevic et al., 2012). In three subjects (S8–S10), the ACC was not recorded at any VOTs. These three subjects also could not perceive VOT when tested using behavioral procedures. Overall, these data established the feasibility of using a relatively long VCV continuum to evaluate cortical neural encoding of VOT cues. More importantly, these results support the idea that electrophysiological measures of the ERP can be used to evaluate the neural encoding of VOT in children with ANSD.

### Results Measured in Quiet

The first hypothesis tested in this study was that the precision of the neural encoding of VOT would affect the auditory perception of VOT in children with ANSD. When all subject data is included, our results measured in quiet showed a robust correlation between categorical boundaries estimated using electrophysiological measures of ERPs and behavioral procedures in children with ANSD, which is consistent with our first hypothesis. When the results of three subjects (S8–S10) who showed the longest VOT were excluded, the trend showing the potential association between categorical boundaries of VOT estimated using these two procedures

was not statistically significant. The non-significant correlation between results measured using behavioral and objective measures did not support our first hypothesis but is consistent with the literature. Specifically, in NH adults, ERPs evoked by VOTs show morphological changes in the N1 peak as VOT increases. However, the results of several studies show that these morphological changes appear to be only dependent on the acoustic properties of stimuli and independent of categorical perceptions of VOTs in NH listeners (Sharma and Dorman, 1999, 2000; Sharma et al., 2000; Elangovan and Stuart, 2011). These studies showed that the categorical pattern of the ERP response can be reliably evoked by VOTs of 40 ms or longer regardless of the subject's categorical perception.

For subjects tested in this study, categorical boundaries measured in quiet using both electrophysiological and behavioral procedures strongly correlated with their most recently measured PBK word scores (measured 1 month prior). Subjects with longer categorical boundaries of VOT had worse speech perception performance. These data are consistent with those reported in He et al. (2015) showing the negative correlation between gap detection threshold and speech perception scores in children with ANSD. Overall, these results suggest that the ability of children with ANSD to perceive timing cues is critical for speech understanding, which is consistent with previous studies (e.g., Zeng et al., 1999; Rance et al., 2008; Starr and Rance, 2015).

### Effects of Competing for Background Noise

Several studies have shown that adding a competing background noise can have a significant negative effect on speech perception in patients with ANSD (Kraus et al., 1984, 2000; Zeng and Liu, 2006; Rance et al., 2007; Berlin et al., 2010). The underlying mechanisms for this observation are largely unknown. The second hypothesis tested in this study was that cortical encoding of VOT as assessed using ERPs in children with ANSD would be adversely affected by a competing background noise, which might account for the excessive difficulty in speech understanding with the presence of a competing noise in these patients. However, our results showed that adding a competing background noise at an SNR of 5 dB only prolonged objective categorical boundaries in three subjects (S1, S2, and S11). Neurophysiological representation of VOT was largely unaffected by the competing noise in S3, S4, and S5. More interestingly, comparisons between behavioral and objective categorical boundaries of VOT measured in noise revealed some discrepancies. For example, S4 needed longer VOT for voiceless consonant perception in noise even though neural encoding of VOT is largely unaffected by noise. Also, S1 apparently could not perceive VOT at all with the presence of background noise even though VOTs of 24 ms or longer can be accurately encoded at the central auditory system. These results suggest that underlying mechanisms for excessive speech understanding in noise may vary among patients with ANSD, perhaps in relation to varying etiologies. While disruption of neural encoding of acoustic cues can account for the difficulty in some patients, the involvement of higher-order brain functions may exist in other patients.

Results of previous studies have suggested that the parallel maturation of cognitive and linguistic skills are important for speech recognition in adverse listening conditions for children (Sullivan et al., 2015; McCreery et al., 2017; MacCutcheon et al., 2019). Specifically, greater working memory capacities are associated with better speech perception in children (Stiles et al., 2012; Sullivan et al., 2015; McCreery et al., 2017; MacCutcheon et al., 2019). Possibly some children with ANSD (e.g., S1 and S4) might have declined working memory capacities, which reduced their capabilities of perceiving VOT cues. Unfortunately, due to limited testing paradigms and recording electrode sites and the lack of cognitive function evaluation, we cannot delineate underlying mechanisms for individual subjects. Further studies on these underlying mechanisms are warranted.

Our results showed that adding a competing noise prolonged P1 peak of the onset ERP response, which is generally consistent with the published literature (Whiting et al., 1998; Kaplan-Neeman et al., 2006; Billings et al., 2009, 2011, 2013; McCullagh et al., 2012; Baltzell and Billings, 2013; Kuruvilla-Mathew et al., 2015; Papesh et al., 2015). The effects of masking noise on ERP amplitudes have not been consistently reported across studies. Whiting et al. (1998) showed that masking noise did not significantly affect ERP amplitudes for responses evoked at SNRs of 0 dB or better. In our study, the presence of competing for background noise at an SNR of 5 dB did not show significant effects on ERP amplitudes, which is consistent with the results of Whiting et al. (1998). An insignificant effect of masking noise on ERP amplitudes has also been reported in other studies (e.g., Kuruvilla-Mathew et al., 2015). However, changes in ERP amplitudes with the presence of masking noises have been observed and reported (e.g., Alain et al., 2009; Parbery-Clark et al., 2011; Billings et al., 2013; Papesh et al., 2015). The discrepancy among results reported in these studies is not entirely clear. Papesh et al. (2015) recently showed that ERP amplitudes are affected by stimulus presentation factors and response component of interest. For example, adding a low-level background noise could enhance the N1 peak at fast presentation rates, but decrease the P1 and the P2 peak regardless of the presentation rate. The amount of enhancement in N1 amplitude depends on the noise level, with higher noise levels resulting in smaller enhancement. Slower presentation rates (e.g., 2 Hz) lead to reductions in N1 amplitudes in noise. Using binaural presentation increases the amplitude of the P1 and the P2 peak, but does not significantly affect the N1 peak. Therefore, differences in stimulus presentation factors and methods used to quantify ERP amplitudes among these studies might, at least partially, account for the discrepancy in reported study results. The effects of masking noise on ACC responses have not been systematically evaluated in human listeners. Further studies are needed to confirm our findings.

## Study Limitations

This study has several potential limitations. First, due to time constraints, there were only a limited number of VOTs tested in this study. In three study participants (S8, S9, and S10), the ACC was not recorded at the longest VOT tested

(i.e., 88 ms). For these participants, the objective category boundary was conservatively estimated to be 89 ms and used in data analysis. This conservative approach should not have changed the overall direction of the correlation between the objective and behavioral categorical boundaries of the VOT or between categorical boundaries of the VOT and PBK word scores. However, it could have reduced the magnitude of these correlations because the actual objective categorical boundary was longer than 89 ms in these participants. Second, due to the challenges of recruiting children with ANSD and normal cognitive function who could complete multiple testing sessions scheduled on the same day, only 11 subjects were tested in this study, which might have limited the statistical power of this study. Third, even though maximum efforts were implemented to hold other acoustic/phonetic properties constant when VOT was manipulated in this study, the stimulus might still include spectral cues (Figure 1). Therefore, the results of this study might not only reflect temporal processing capacities in children with ANSD. Finally, whereas ERPs evoked by the VOT evaluate passive neural encoding of VOT cues, the behavioral categorical perception test requires active subject participation and relies on the subject's cognitive function. Therefore, the possibility that these two measures may assess different mechanisms underlying speech perception cannot be excluded. Due to these study limitations, the results of this study need to be interpreted with caution.

## Contributions to the Literature

Despite these study limitations, the results of this study made three contributions to the literature. First, the results of this study established the feasibility of using a relatively long VCV continuum to evaluate cortical neural encoding of VOT cues. This is a novel stimulus for ERP measures. Second, these results supported the idea that electrophysiological measures of the ERP can be used to evaluate the neural encoding of VOT in children with ANSD, which addressed a knowledge gap in the field. Finally, data from this study indicated that the underlying mechanisms for excessive speech understanding in noise for patients with ANSD are heterogeneous. Unfolding these mechanisms will require assessing a large group of patients with ANSD using a comprehensive testing battery, including neurocognitive assessments, electrophysiological, and behavioral testing procedures.

## CONCLUSION

VCV continua can be used to evaluate behavioral identification and the neural encoding of VOT in children with ANSD. Categorical boundaries of VOT estimated using behavioral measures and ERP recordings are closely associated in quiet. Therefore, electrophysiological measures of the ERP can be used to evaluate the neural encoding of VOT in children with ANSD. The results of these measures also show a relationship with speech perception scores in subjects tested in this study. Underlying mechanisms for excessive speech perception deficits in noise may vary for individual patients with ANSD.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Institutional Review Board (IRB) at The University of North Carolina at Chapel Hill. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

TM and PB participated in data collection, data analysis, results interpretation, and are accountable for all aspects of this study. They also provided critical revision and approved the final version of this article. JS designed the speech stimuli used in

this study, provided critical comments, and approved the final version of this article. SH designed and is accountable for all aspects of this study, as well as drafted and approved the final version of this article.

## FUNDING

This work was supported by the Junior Faculty Career Development Award from School of Medicine, The University of North Carolina at Chapel Hill, and the R01 grant from National Institute on Deafness and Other Communication Disorders (NIDCD; R01DC017846 for supporting SH's effort on this project and for the publication cost of this article).

## ACKNOWLEDGMENTS

Portions of this article were presented at the 8<sup>th</sup> International Symposium on Objective Measures in Auditory Implants, Toronto, CA, USA in 2014.

## REFERENCES

- Alain, C., Quan, J., McDonald, K., and Van Roon, P. (2009). Noise-induced increase in human auditory evoked neuromagnetic fields. *Eur. J. Neurosci.* 30, 132–142. doi: 10.1111/j.1460-9568.2009.06792.x
- Apeksha, K., and Kumar, U. A. (2019). Effect of acoustic features on discrimination ability in individuals with auditory neuropathy spectrum disorder: an electrophysiological and behavioral study. *Eur. Arch. Otorhinolaryngol.* 276, 1633–1641. doi: 10.1007/s00405-019-05405-9
- Atal, B. S., and Hanauer, S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *J. Acoust. Soc. Am.* 50, 637–655. doi: 10.1121/1.1912679
- Baltzell, L. S., and Billings, C. J. (2013). Sensitivity of offset and onset cortical auditory evoked potentials to signals in noise. *Clin. Neurophysiol.* 125, 370–380. doi: 10.1016/j.clinph.2013.08.003
- Berlin, C. I., Hood, L. J., Morlet, T., Wilensky, D., Li, L., Mattingly, K. R., et al. (2010). Multi-site diagnosis and management of 260 patients with auditory neuropathy/dys-synchrony (auditory neuropathy spectrum disorder). *Int. J. Audiol.* 49, 30–43. doi: 10.3109/14992020903160892
- Bielecki, I., Horbulewicz, A., and Wolan, T. (2012). Prevalence and risk factors for auditory neuropathy spectrum disorder in a screened newborn population at risk for hearing loss. *Int. J. Pediatr. Otorhinolaryngol.* 76, 1668–1670. doi: 10.1016/j.ijporl.2012.08.001
- Billings, C. J., Bennett, K. O., Molis, M. R., and Leek, M. R. (2011). Cortical encoding of signals in noise: effects of stimulus type and recording paradigm. *Ear Hear.* 32, 53–60. doi: 10.1097/AUD.0b013e3181ec5c46
- Billings, C. J., McMillan, G., Penman, T., and Gille, S. M. (2013). Predicting perception in noise using cortical auditory evoked potentials. *J. Otolaryngol.* 14, 891–903. doi: 10.1007/s10162-013-0415-y
- Billings, C. J., Tremblay, K. L., Stecker, G. C., and Tolin, W. M. (2009). Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hear. Res.* 254, 15–24. doi: 10.1016/j.heares.2009.04.002
- Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glott International.* 5, 341–345.
- Ceponiene, R., Cheour, M., and Näätänen, R. (1998). Interstimulus interval and auditory event-related potentials in children: evidence for multiple generators. *Electroencephalogr. Clin. Neurophysiol.* 108, 345–354. doi: 10.1016/s0168-5597(97)00081-6
- Dimitrijevic, A., Pratt, H., and Starr, A. (2012). Auditory cortical activity in normal hearing subjects to consonant vowels presented in quiet and in noise. *Clin. Neurophysiol.* 124, 1204–1215. doi: 10.1016/j.clinph.2012.11.014
- Elangovan, S., and Stuart, A. (2008). Natural boundaries in gap detection are related to categorical perception of stop consonants. *Ear Hear.* 29, 761–774. doi: 10.1097/aud.0b013e318185ddd2
- Elangovan, S., and Stuart, A. (2011). A cross-linguistic examination of cortical auditory evoked potentials for a categorical voicing contrast. *Neurosci. Lett.* 490, 140–144. doi: 10.1016/j.neulet.2010.12.044
- Frye, R. E., Fisher, J. M., Coty, A., and Zarella, M. (2007). Linear coding of voice onset time. *J. Cogn. Neurosci.* 19, 1476–1487. doi: 10.1162/jocn.2007.19.9.1476
- Ganong, W. F. III. (1980). Phonetic categorization in auditory word perception. *J. Exp. Psychol. Hum. Percept. Perform.* 6, 110–125. doi: 10.1037/0096-1523.6.1.110
- Gilley, P. M., Sharma, A., Dorman, M., and Martin, K. (2005). Developmental changes in refractoriness of the cortical auditory evoked potential. *Clin. Neurophysiol.* 116, 648–657. doi: 10.1016/j.clinph.2004.09.009
- He, S., Grose, J. H., Teagle, H. F., Woodard, J., Park, L. R., Hatch, D. R., et al. (2015). Acoustically evoked auditory change complex in children with auditory neuropathy spectrum disorder: a potential objective tool for identifying cochlear implant candidates. *Ear Hear.* 36, 289–301. doi: 10.1097/aud.000000000000119
- Horev, N., Most, T., and Pratt, H. (2007). Categorical perception of speech (VOT) and analogous non-speech (FOT) signals: behavioral and electrophysiological correlates. *Ear Hear.* 28, 111–128. doi: 10.1097/01.aud.0000250021.69163.96
- Kaplan-Neeman, R., Kishon-Rabin, L., Henkin, Y., and Muchnik, C. (2006). Identification of syllables in noise: electrophysiological and behavioral correlates. *J. Acoust. Soc. Am.* 120, 926–933. doi: 10.1121/1.2217567
- King, K. A., Campbell, J., Sharma, A., Martin, K., Dorman, M., and Langran, J. (2008). The representation of voice onset time in the cortical auditory evoked potentials of young children. *Clin. Neurophysiol.* 119, 2855–2861. doi: 10.1016/j.clinph.2008.09.015
- Kraus, N., Bradlow, A. R., Cheatham, J., Cunningham, J., King, C. D., Koch, D. B., et al. (2000). Consequences of neural asynchrony: a case of auditory neuropathy. *J. Assoc. Res. Otolaryngol.* 1, 33–45. doi: 10.1007/s10162010004
- Kraus, N., and Nicol, T. (2003). Aggregate neural responses to speech sounds in the central auditory system. *Speech Commun.* 41, 35–47. doi: 10.1016/s0167-6393(02)00091-2
- Kraus, N., Özdamar, O., Stein, L., and Reed, N. (1984). Absent auditory brainstem response: peripheral hearing loss or brainstem dysfunction? *Laryngoscope* 94, 400–406. doi: 10.1288/00005537-198403000-00019

- Kuruville-Mathew, A., Purdy, S. C., and Welch, D. (2015). Cortical encoding of speech acoustics: effects of noise and amplification. *Int. J. Audiol.* 54, 852–864. doi: 10.3109/14992027.2015.1055838
- Lisker, L. (1975). Is it VOT or a first-formant transition detector? *J. Acoust. Soc. Am.* 57, 1547–1551. doi: 10.1121/1.380602
- Lisker, L. (1986). “Voicing” in English: a catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Lang. Speech* 29, 3–11. doi: 10.1177/002383098602900102
- Lisker, L., and Abramson, A. S. (1964). A cross language study of voicing in initial stops: acoustical measurements. *Word* 20, 384–422. doi: 10.1080/00437956.1964.11659830
- MacCutcheon, D., Pausch, F., Füllgrabe, C., Eccles, R., van der Linde, J., Panebianco, C., et al. (2019). The contribution of individual differences in memory span and ability to spatial release from masking in young children. *J. Speech Hear. Res.* 62, 3741–3751. doi: 10.1044/2019\_jslhr-s-19-0012
- Madden, C., Rutter, M., Hilbert, L., Greinwald, J. H. Jr., and Choo, D. I. (2002). Clinical and audiological features in auditory neuropathy. *Arch. Otolaryngol. Head Neck Surg.* 128, 1026–1030. doi: 10.1001/archotol.128.9.1026
- Martin, B. A., Tremblay, K. L., and Korczak, P. (2008). Speech evoked potentials: from the laboratory to the clinic. *Ear Hear.* 29, 285–313. doi: 10.1097/aud.0b013e3181662c0e
- McCreery, R., Spratford, M., Kirby, B., and Brennan, M. (2017). Individual differences in language and working memory affect children’s speech recognition in noise. *Int. J. Audiol.* 56, 306–315. doi: 10.1080/14992027.2016.1266703
- McCullagh, J., Musiek, F. E., and Shinn, J. B. (2012). Auditory cortical processing in noise in normal-hearing young adults. *Audiol. Med.* 10, 114–121. doi: 10.3109/1651386x.2012.707354
- Michalewski, H. J., Starr, A., Nguyen, T. T., Kong, Y. Y., and Zeng, F. G. (2005). Auditory temporal processes in normal-hearing individuals and in patients with auditory neuropathy. *Clin. Neurophysiol.* 116, 669–680. doi: 10.1016/j.clinph.2004.09.027
- Morrongiello, B. A., Robson, R. C., Best, C., and Clifton, R. K. (1984). Trading relations in the perception of speech by 5-year-old children. *J. Exp. Child Psychol.* 37, 231–250. doi: 10.1016/0022-0965(84)90002-x
- Nilsson, J. M., Soli, S. D., and Gelnett, D. J. (1996). *Development of the Hearing in Noise Test for Children (HINT-C)*. Los Angeles, CA: House Ear Institute.
- Papesh, M. A., Billings, C. J., and Baltzell, L. S. (2015). Background noise can enhance cortical auditory evoked potentials under certain conditions. *Clin. Neurophysiol.* 126, 1319–1330. doi: 10.1016/j.clinph.2014.10.017
- Parbery-Clark, A., Marmar, F., Bair, J., and Kraus, N. (2011). What subcortical-cortical relationships tell us about processing speech in noise. *Eur. J. Neurosci.* 33, 549–557. doi: 10.1111/j.1460-9568.2010.07546.x
- Pelosi, S., Rivas, A., Haynes, D. S., Bennett, M. L., Labadie, R. F., Hedley-Williams, A., et al. (2012). Stimulation rate reduction and auditory development in poorly performing cochlear implant users with auditory neuropathy. *Otol. Neurotol.* 33, 1502–1506. doi: 10.1097/mao.0b013e31826bec1e
- Phillips, D. P. (1985). Temporal response features of cat auditory cortex neurons contributing to sensitivity to tones delivered in the presence of continuous noise. *Hear. Res.* 19, 253–268. doi: 10.1016/0378-5955(85)90145-5
- Phillips, D. P. (1990). Neural representation of sound amplitude in the auditory cortex: effects of noise masking. *Behav. Brain Res.* 37, 197–214. doi: 10.1016/0166-4328(90)90132-x
- Phillips, D. P., and Hall, S. E. (1986). Spike-rate intensity functions of cat cortical neurons studied with combined tone-noise stimuli. *J. Acoust. Soc. Am.* 80, 177–187. doi: 10.1121/1.394178
- Phillips, D. P., and Kelly, J. B. (1992). Effects of continuous noise maskers on tone-evoked potentials in cat primary auditory cortex. *Cereb. Cortex* 2, 134–140. doi: 10.1093/cercor/2.2.134
- Rance, G. (2005). Auditory neuropathy/dys-synchrony and its perceptual consequences. *Trends Amplif.* 9, 1–43. doi: 10.1177/108471380500900102
- Rance, G., Barker, E. J., Mok, M., Dowell, R., Rincon, A., and Garratt, R. (2007). Speech perception in noise for children with auditory neuropathy/dys-synchrony type hearing loss. *Ear Hear.* 28, 351–360. doi: 10.1097/aud.0b013e3180479404
- Rance, G., Beer, D. E., Cone-Wesson, B., Shepherd, R. K., Dowell, R. C., King, A. M., et al. (1999). Clinical findings for a group of infants and young children with auditory neuropathy. *Ear Hear.* 20, 238–252. doi: 10.1097/00003446-199906000-00006
- Rance, G., Fava, R., Baldock, H., Chong, A., Barker, E., Corben, L., et al. (2008). Speech perception ability in individuals with Friedreich ataxia. *Brain* 131, 2002–2012. doi: 10.1093/brain/awn104
- Roman, S., Canevet, G., Lorenzi, C., Triglia, J. M., and Liégeois-Chauvel, C. (2004). Voice onset time encoding in patients with left and right cochlear implant. *Neuroreport* 15, 601–605. doi: 10.1097/00001756-200403220-00006
- Roush, P., Frymark, T., Venediktov, R., and Wang, B. (2011). Audiologic management of auditory neuropathy spectrum disorder in children: a systematic review of the literature. *Am. J. Audiol.* 20, 159–170. doi: 10.1044/1059-0889(2011/10-0032)
- Shalloo, J. K. (2002). Auditory neuropathy spectrum disorder in adults and children. *Semin. Hear.* 23, 215–223. doi: 10.1055/s-2002-34474
- Sharma, A., and Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J. Acoust. Soc. Am.* 106, 1078–1083. doi: 10.1121/1.428048
- Sharma, A., and Dorman, M. F. (2000). Neuropsychologic correlates of cross-language phonetic perception. *J. Acoust. Soc. Am.* 107, 2697–2703. doi: 10.1121/1.428655
- Sharma, A., Marsh, C. M., and Dorman, M. F. (2000). Relationship between N1 evoked potential morphology and the perception of voicing. *J. Acoust. Soc. Am.* 108, 3030–3035. doi: 10.1121/1.1320474
- Sinex, D., and McDonald, L. (1989). Synchronized discharge rate representation of voice-onset time in the chinchilla auditory nerve. *J. Acoust. Soc. Am.* 85, 1995–2004. doi: 10.1121/1.397852
- Sinex, D., McDonald, L., and Mott, J. B. (1991). Neural correlates of nonmonotonic temporal acuity for voice-onset time. *J. Acoust. Soc. Am.* 90, 2441–2449. doi: 10.1121/1.402048
- Sinex, D., and Narayan, S. (1994). Auditory-nerve fiber representation of temporal cues to voicing in word-medial stop consonants. *J. Acoust. Soc. Am.* 95, 897–903. doi: 10.1121/1.408400
- Starr, A., McPherson, D., Patterson, J., Don, M., Luxford, W., Shannon, R., et al. (1991). Absence of both auditory evoked potentials and auditory percepts dependent on timing cues. *Brain* 111, 1157–1180. doi: 10.1093/brain/114.3.1157
- Starr, A., and Rance, G. (2015). “Auditory neuropathy,” in *Handbook of Clinical Neurology*, eds G. G. Celesia and G. Hickok (Australia: Elsevier), 495–508.
- Steinschneider, M., Fishman, Y. I., and Arezzo, J. C. (2003). Representation of the voice onset time (VOT) speech parameter in population responses within primary cortex of the awake monkey. *J. Acoust. Soc. Am.* 114, 307–321. doi: 10.1121/1.1582449
- Steinschneider, M., Nourski, K. V., and Fishman, Y. I. (2013). Representation of speech in human auditory cortex: is it special? *Hear. Res.* 305, 57–73. doi: 10.1016/j.heares.2013.05.013
- Steinschneider, M., Volkov, I. O., Fishman, Y. I., Oya, H., Arezzo, J. C., and Howard, M. A. III. (2005). Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cereb. Cortex* 15, 170–186. doi: 10.1093/cercor/bhh120
- Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard, M. A. III. (1999). Temporal encoding of the voice onset time phonetic parameter by field potential recorded directly from human auditory cortex. *J. Neurophysiol.* 82, 2346–2357. doi: 10.1152/jn.1999.82.5.2346
- Stephens, J. D. W., and Holt, L. L. (2011). A standard set of American-English voiced stop-consonant stimuli from morphed natural speech. *Speech Commun.* 53, 877–888. doi: 10.1016/j.specom.2011.02.007
- Stevens, K. N., and Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *J. Acoust. Soc. Am.* 55, 653–659. doi: 10.1121/1.1914578

- Stiles, D., Bentler, R., and McGregor, K. (2012). The speech intelligibility index and the pure-tone average as predictors of lexical ability in children fit with hearing aids. *J. Speech Lang. Hear. Res.* 55, 764–778. doi: 10.1044/1092-4388(2011/10-0264)
- Sullivan, J. R., Osman, H., and Schafer, E. C. (2015). The effect of noise on the relationship between auditory working memory and comprehension in school-age children. *J. Speech Lang. Hear. Res.* 58, 1043–1051. doi: 10.1044/2015\_jslhr-h-14-0204
- Teagle, H. F. B., Roush, P. A., Woodard, J. S., Hatch, D. R., Zdanski, C. J., Buss, E., et al. (2010). Cochlear implantation in children with auditory neuropathy spectrum disorder. *Ear Hear.* 31, 325–335. doi: 10.1097/AUD.0b013e3181ce693b
- Tremblay, K. L., Friesen, L., Martin, B. A., and Wright, R. (2003). Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear Hear.* 24, 225–232. doi: 10.1097/01.aud.0000069229.84883.03
- Vlastarakos, P., Mikopoulous, T., Tavoulari, E., Papacharalambous, G., and Korres, S. (2008). Auditory neuropathy: endocochlear lesion or temporal processing impairment? Implications for diagnosis and management. *Int. J. Ped. Otorhinolaryngol.* 72, 1135–1150. doi: 10.1016/j.ijporl.2008.04.004
- Whiting, K. A., Martin, B. A., and Stapells, D. R. (1998). The effects of broadband noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Ear Hear.* 19, 218–231. doi: 10.1097/00003446-199806000-00005
- Zeng, F. G., Kong, Y. Y., Michalewski, H. J., and Starr, A. (2005). Perceptual consequences of disrupted auditory nerve activity. *J. Neurophysiol.* 93, 3050–3063. doi: 10.1152/jn.00985.2004
- Zeng, F. G., and Liu, S. (2006). Speech perception in individuals with auditory neuropathy. *J. Speech Lang. Hear. Res.* 49, 367–380. doi: 10.1044/1092-4388(2006/029)
- Zeng, F. G., Oba, S., Garde, S., Sininger, Y., and Starr, A. (1999). Temporal and speech processing deficits in auditory neuropathy. *Neuroreport* 10, 3429–3435. doi: 10.1097/00001756-199911080-00031
- Zeng, F. G., Oba, S., Garde, S., Sininger, Y., and Starr, A. (2001). “Psychoacoustic and speech perception in auditory neuropathy,” in *Auditory Neuropathy: A New Perspective on Hearing Disorder*, eds Y. Sininger and A. Starr (San Diego, CA: Singular Press), 141–164.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 McFayden, Baskin, Stephens and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.